# A Modern Introduction to Linear Algebra

HENRY RICARDO

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

# A Modern Introduction to
# Linear Algebra

# A Modern Introduction to Linear Algebra

**HENRY RICARDO**

CRC Press
Taylor & Francis Group
Boca Raton London New York

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

MATLAB® is a trademark of The MathWorks, Inc. and is used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. This book's use or discussion of MATLAB® software or related products does not constitute endorsement or sponsorship by The MathWorks of a particular pedagogical approach or particular use of the MATLAB® software.

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


Version Date: 20131121

International Standard Book Number-13: 978-1-4398-9461-3 (eBook - PDF)



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

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

*For my wife, Catherine—"Age cannot wither her,*
*nor custom stale her infinite variety."—*
*and for*

*Henry and Marta Ricardo*
*Tomás and Nicholas Ricardo*

*Cathy and Mike Corcoran*
*Christopher Corcoran*

*Christine and Greg Gritmon*

# Contents

An asterisk * marks optional sections, those discussions not needed for later work.

# Author

**Henry Ricardo** received his BS from Fordham College and his MA and PhD from Yeshiva University.

Dr. Ricardo has taught at Manhattan College and has worked at IBM in various technical and business positions. He is currently professor of mathematics at Medgar Evers College of the City University of New York, where he was presented with the 2008 Distinguished Service Award by the School of Science, Health and Technology. In 2009, he was given the Distinguished Service Award by the Metropolitan New York Section of the MAA. The second edition of his previous book, *A Modern Introduction to Differential Equations*, was published by Academic Press in 2009. The first edition, published by Houghton Mifflin, has been translated into Spanish.

Professor Ricardo is a member of Phi Beta Kappa and Sigma Xi, as well as of the AMS, MAA, NCTM, SIAM, and the International Linear Algebra Society. He is currently the governor of the Metropolitan New York Section of the MAA.

# Introduction

## I.1 Rationale

Linear algebra has always been an important foundation course for higher mathematics. Traditionally, this has been the course that introduces students to the concept of **mathematical proof**, although in recent years there have been alternative paths to this enlightenment. More recently, especially with the availability of computers and powerful handheld calculators, linear algebra has emerged as an essential prerequisite for many areas of application—even penetrating more deeply than calculus.

This book is for a one-semester or a two-semester course at the sophomore–junior level intended for a wide variety of students: mathematics majors, engineering students, and those in other scientific areas such as biology, chemistry, and computer science. It stresses **proofs**, is **matrix-oriented**, and has examples, applications, and exercises reflecting some of the many disciplines that use linear algebra. Although it is assumed that the student has had at least two semesters of calculus, those remarks, examples, and exercises invoking calculus can easily be skipped.

If the instructor chooses to deemphasize proofs and avoid the more theoretical topics, he or she can even use this book for students in business, economics, and the social sciences. To echo the last page of a recent book[*] in another area of mathematics, the subject is LINEAR ALGEBRA and NOT Calculus, Numerical Analysis, Graph Theory, Physics, Computer Science, Economics, or Sociology. I provide indications of some of the applications of linear algebra in many disciplines, but I delegate the burden of more extensive expositions to the instructors in these areas. While providing some applied examples, as well as discussions of numerical aspects, **I have tried to focus on those beautiful and useful concepts that are at the heart of linear algebra**. I hope that this book will be read as a textbook and then kept as a reference for linear algebra concepts and results.

## I.2 Pedagogy

In teaching linear algebra, I have been guided by my experiences as a student and teacher and (more recently) by the efforts and recommendations of the Linear Algebra Curriculum Study Group (LACSG), the ATLAST[†] project, and the Linear Algebra Modules Project (LAMP). In 1997 I was a participant in an ATLAST workshop at the University of Wisconsin (Madison). Based on their collective research, teaching experience, and consultations with users of linear algebra in industry, these groups are unanimous in advocating that the first course in linear algebra be taught from a matrix-oriented point of view and that software be used to enhance teaching and learning.[‡] The content of my text reflects the recommendations cited; and although I have been using software (primarily *Maple*®) in my courses for several years, I have written my text in a "pencil and paper" manner, but with opportunities for calculator or computer usage.

[*] D.A. Sánchez. *Ordinary Differential Equations: A Brief Eclectic Tour* (Washington, DC: MAA, 2002).

[†] An acronym for Augmenting the Teaching of Linear Algebra through the use of Software Tools.

[‡] See, for example, the *Special Issue on Linear Algebra*, *Coll. Math. J.* **24**(1), (January 1993), *Resources for Teaching Linear Algebra* (MAA Notes, No. 42, 1997), and *Linear Algebra Gems: Assets for Undergraduate Mathematics* (MAA Notes, No. 59, 2002).

*My writing style is informal, even conversational at points*. I provide **proofs** for virtually all results, leaving some others (or parts of others) as exercises. I have generally provided more details and more "between the lines" explanation than is customary in mathematical proofs, even in textbooks. I have tried to find the "best" proofs from the point of view of the reader, usually accompanying them with motivating examples and illuminating discussions.

Student access to computers and calculators makes the analysis of theorems and algorithms an important part of the learning process, but I do not impose technology on the reader. Virtually all linear algebra calculations can be done routinely using graphing calculators and computer algebra systems. The individual instructor should determine the extent to which his or her students make use of technology.

I believe in teaching linear algebra computationally (numerically, algorithmically), algebraically, and geometrically, with an emphasis on the algebraic aspects. Concrete, easy-to-understand examples motivate the theory. My typical pedagogical cycle is **motivating concrete example(s) → analysis → general principle (theorem and proof) → additional concrete example(s)**. I also rely on a spiral approach, by which I introduce an idea in a very simple situation and then revisit it often, building through successive sections and chapters to the full glory of the concept. For example, the concept of *linear transformation* is introduced briefly in the context of matrix–vector multiplication (Section 3.2), discussed more generally in Example 5.2.4 (the range as a subspace of a vector space), and developed fully in Chapter 6.

Each chapter begins with an **introductory text** that motivates the topics to come. These opening remarks link the material that has been treated previously with the content of the present chapter. There are **Exercises** at the end of each section, divided into A, B, and C problems. I have found such a triage useful in the past. In this book, there are over **1200 numbered exercises**, many with multiple parts. Generally, "A" exercises are drill problems, providing practice with routine calculations and symbol manipulations. "B" exercises are typically more theoretical, asking for simple proofs (sometimes the completion of proofs in the text proper) or demanding more elaborate calculations. Among "B" exercises are requests for examples or counterexamples and exercises encouraging the recognition of patterns and generalizations. "C" exercises are meant to challenge student understanding of the concepts and calculations taught in the text. These problems may require more complex calculations, manipulations, and/or proofs. A typical "C" exercise will be broken up into several parts and may have more generous hints given for its solution. Overall, many of the exercises involve concepts and applications that do not appear in the book, thus enriching the student's view of linear algebra. These exercises also introduce some concepts that will be developed further in later sections.

*I have included problems for those who have access to graphing calculators and computers, but these exercises are not platform-specific*. I have labeled certain exercises with the symbol to suggest that technology be used; but, in fact, an instructor may choose to allow students to use technology for many exercises involving calculation—perhaps asking for a hand computation followed by a check via calculator or computer.

There are many **figures**, both algebraic and geometric, to illustrate various results.

I end each chapter with a **Summary** of key definitions and results.

Finally, the **Appendices** review or introduce certain basic mathematical tools. In particular, there is an exposition of mathematical induction, which is used throughout the text.

There will be an **Instructor's Solutions Manual**, containing worked-out solutions to all exercises.

## I.3 Content

The content of my book reflects the recommendations of the groups mentioned in the last section: LACSG, ATLAST, and LAMP. An asterisk (*) marks optional sections, those topics not essential for later work.

Although most of the drama takes place in the context of finite-dimensional real vector spaces, there are roles for complex spaces and cameo appearances by infinite-dimensional spaces.

**Chapter 1** introduces the reader to vectors, the fundamental units of linear algebra. The space $\mathbb{R}^n$ is introduced naturally and its algebraic and geometric properties are explored. Using many examples in $\mathbb{R}^2$ and $\mathbb{R}^3$, the concepts of linear independence, spans, bases, and subspaces are discussed. In particular, the calculations required to determine the linear independence/dependence of sets of vectors serve as motivation for the work in Chapter 2.

Systems of linear equations are discussed thoroughly in **Chapter 2**, from geometric, algebraic, and numerical points of view. Gaussian elimination and reduced row echelon forms are at the heart of this chapter. The connection between the rank and the nullity of a matrix is established and is shown to have important consequences in the analysis of systems of $m$ linear equations in $n$ unknowns.

**Chapter 3** provides a basic minicourse in the theory of matrices. Introduced as notational devices in Chapter 2, matrices now develop a personality of their own. The explanation of matrix–vector multiplication provides a glimpse of the matrix as a dynamic object, operating on vectors to transform them linearly. Inverses of matrices are handled via elementary matrices and Gaussian elimination. The important LU and PA = LU factorizations are explained and illustrated. In the next chapter, ***the determinant of a square matrix is defined in terms of these decompositions***.

Even though the role of determinants in linear algebra has diminished, there is still some need for this concept. In a unique way, **Chapter 4** *defines* the determinant of a matrix $A$ in terms of the PA = LU factorization. Fundamental properties are then proved easily and traditional methods for hand calculation of determinants are introduced. There follows a treatment of similarity and diagonalization of matrices. The Cayley–Hamilton Theorem is proved for diagonalizable matrices and the *minimal polynomial* is introduced.

**Chapter 5** introduces the abstract idea of a *real vector space*. The connection between the concepts introduced in Chapter 1 in the context of $\mathbb{R}^n$ and the more abstract ideas is stressed. In particular, this chapter emphasizes that *every n-dimensional real vector space is algebraically identical to* $\mathbb{R}^n$. There is some discussion of the concept of an *infinite-dimensional vector space*.

**Chapter 6** provides a comprehensive coverage of *linear transformations*, including details of the algebra of linear transformations and the matrix representation of linear transformations with respect to bases. Invertibility, isomorphisms, and similarity are also treated.

**Chapter 7** introduces complex vector spaces and expands the treatment of the dot product defined in Chapter 1. General inner product spaces over $\mathbb{R}$ and $\mathbb{C}$ are discussed. Orthonormal bases and the Gram–Schmidt process are explained and illustrated. The chapter also treats unitary and orthogonal matrices, the Schur factorization, the Cayley–Hamilton Theorem for general square matrices, the QR factorization, orthogonal complements, and projections.

**Chapter 8** introduces the concept of a linear functional and the adjoint of an operator. Discussions of Hermitian and normal matrices (including spectral theorems) follow, with a digression on quadratic forms. The book concludes with the Singular Value Decomposition and the Polar Decomposition.

**Appendices A through D** provide the basics of **set theory**, an explanation of **summation and product notation**, an exposition of **mathematical induction**, and a quick survey of **complex numbers**.

## I.4 Using This Book

Depending on the requirements of the course and the sophistication of the students, this book contains enough material for a one- or two-semester course. Chapters 1 through 4 form the core of a one-semester course for an average class. This material is the heart of linear algebra and prepares a student for future abstract courses as well as applied courses having linear algebra as a prerequisite. A class with a strong background may want to investigate parts of Chapters 5 and 6.

Alternatively, if the instructor has the luxury of a two-semester linear algebra course, he or she may want to cover the relatively concrete Chapters 1 through 4 in the first semester and explore the more abstract Chapters 5 through 7 in the second semester. Additional material from Chapter 8 may be used to complete the student's introduction to linear algebraic concepts.

Overall, an instructor may skip or deemphasize Chapter 4 and go directly to Chapters 5 through 8 from the first three chapters. The second half of the book, more abstract than the first, requires only a nodding acquaintance with determinants.

MATLAB® is a registered trademark of The MathWorks, Inc. For product information, please contact:

The MathWorks, Inc.
3 Apple Hill Drive
Natick, MA 01760-2098 USA
Tel: 508 647 7000
Fax: 508-647-7001
E-mail: info@mathworks.com
Web: www.mathworks.com

## Acknowledgments

Dr. Michael Aissen was my first linear algebra teacher and was responsible for introducing me to other beautiful and interesting areas of mathematics, as my classroom instructor and as my mentor in undergraduate research projects. I thank him for all this.

Students of my generation were inspired by the classic linear algebra texts of Paul Halmos and Hoffman/Kunze. (Halmos also inspired my later PhD work in operator theory.) In writing this book, I find myself indebted to many fine contemporary researchers and expositors, notably Carl Meyer and Gilbert Strang. My online review for the MAA of *Handbook of Linear Algebra*, edited by Leslie Hogben (Chapman & Hall/CRC, Boca Raton, FL, 2007), did not do justice to the richness of information in this authoritative tome.

At Medgar Evers College, it has been my pleasure to have Darius Movasseghi as my supportive Chair and to have inspiring colleagues, among them Tatyana Flesher, Mahendra Kawatra, Umesh Nagarkatte, Mahmoud Sayrafiezadeh, and Raymond Thomas. Several generations of students have endured my efforts to demonstrate the beauty of linear algebra, among them Patrice Williams, Reginald Dorcely, Jonathan Maitre, Frantz Voltaire, Yeku Oladapo, Andre Robinson, and Barry Gregory.

I am grateful to Lauren Schultz Yuhasz for suggesting that I write a book on linear algebra. I thank David Grubbs, my editor at CRC Press, for his encouragement and his enthusiastic support of my project. Amy Blalock, the project coordinator, deserves my appreciation for guiding me through the production process, as does Joette Lynch, the project editor.

I thank my reviewers for their insightful comments and helpful suggestions, especially Hugo J. Woerdeman (Drexel University) and Matthias K. Gobbert (University of Maryland).

Above all, I owe my wife a great deal for her support of my work on this book.

Of course, any errors that remain are my own responsibility. I look forward to your comments, suggestions, and questions.

**Henry Ricardo**
henry@mec.cuny.edu

1

# Vectors

We begin our investigation of linear algebra by studying *vectors*, the basic components of much of the theory and applications of linear algebra. Historically, vectors were not the first objects in linear algebra to be studied intensively. The solution of systems of linear equations provided much of the early motivation for the subject, with determinants arising as an early topic of interest. Anyone reading this book has probably worked with linear systems and may even have encountered determinants in this context. We will investigate these topics fully in Chapters 2 and 4.

Now that the subject has matured to some extent—although both the theory and applications continue to be developed—we can begin our study logically with vectors, the fundamental mathematical objects which form the foundation of this magnificent structure known as linear algebra. Vectors are not only important in their own right, but they link linear algebra and geometry in a vital way.

## 1.1 Vectors in $\mathbb{R}^n$

At some time in our earlier mathematical studies, we learned the concept of *ordered pairs* $(x, y)$ of real numbers, then *ordered triplets* $(x, y, z)$, and perhaps *ordered n-tuples* $(x_1, x_2, \ldots, x_n)$. This idea and these notations are simple and natural. In addition to being more important in theoretical applications (in defining a *function*, for example), *these n-tuples serve to hold data and to carry information*.* For example, the ordered 4-tuple (123456789, 35, 70, 145) might describe a person with Social Security number 123-45-6789 who is 35 years old, 5′10″ tall, and weighs 145 lb. In interpreting such data, it is important that we understand the *order* in which the information is given.

Despite the generality of the idea we have just been discussing, we will consider only real numerical data for most of our study of linear algebra. Eventually, beginning in Chapter 5, we will illustrate the flexibility and wider applicability of the vector concept by allowing complex numbers and functions as data elements.

* The word *vector* is the Latin word for "carrier." In pathology and epidemiology, this word refers to an organism (tick, mosquito, or infected human) that carries a disease from one host to another.

### Definition 1.1.1

A **vector** is an ordered finite list of real numbers. A **row vector**, denoted by $[x_1 \quad x_2 \quad \cdots \quad x_n]$, is an ordered finite list of numbers written horizontally. A **column vector**, denoted by $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, is an ordered finite list of real numbers written vertically. Each number $x_i$ making up a vector is called a **component** of the vector.

*From now on, we will use lowercase* ***bold*** *letters to denote vectors and all other lowercase letters (usually italicized) to represent real numbers.* For example,

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

denotes a (column) vector **y** whose components (real numbers) are $y_1, y_2,$ and $y_3$. (An alternative, especially prevalent in applied courses, is to use arrows to denote vectors: $\vec{v}$. This form is also used in classroom presentations because it is difficult to indicate boldface with chalk or markers.)

It is useful to introduce the concept of a *transpose* here. If we have a column vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, then its **transpose**, denoted by $\mathbf{x}^{\mathrm{T}}$, is the row vector $[x_1 \quad x_2 \quad \ldots \quad x_n]$. Similarly, the **transpose** of a row vector, $\mathbf{w} = [w_1 \quad w_2 \quad \ldots \quad w_n]$, denoted by $\mathbf{w}^{\mathrm{T}}$, is the column vector $\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$. The form of vector we use will depend on the context of a problem and will change to suit various theoretical and applied considerations—sometimes just for typographical convenience.

A simple data table can illustrate the use of both row and column vectors.

## Example 1.1.1: Column Vectors and Row Vectors

Consider the following grade sheet for an advanced math class.

| Student | Test 1 | Test 2 | Test 3 |
|---|---|---|---|
| Albarez, Javier | 78 | 84 | 87 |
| DeLuca, Matthew | 85 | 76 | 89 |
| Farley, Rosemary | 93 | 88 | 94 |
| Nguyen, Bao | 89 | 94 | 95 |
| VanSlambrouck, Jennifer | 62 | 79 | 87 |

We can form various vectors from this table. For example, the column vector $\begin{bmatrix} 84 \\ 76 \\ 88 \\ 94 \\ 79 \end{bmatrix}$, formed from the second column of the table, gives all the grades on Test 2. The position (row) of a number associates the grade with a particular student. Similarly, the row vector $[\,89 \quad 94 \quad 95\,]$ represents all Bao Nguyen's test marks, the position (column) of a number indicating the number of the test. Note that in this illustration, the vectors can have different shapes and numbers of components depending on the meaning of the vectors.

---

In working with vectors, it is important that the components be *ordered*, that is, each component represents a specific piece of information and any rearrangement of the data represents a different vector. A more precise way of saying this is that *two vectors*, say $\mathbf{x} = [\,x_1 \quad x_2 \quad \ldots \quad x_n\,]$ *and* $\mathbf{y} = [\,y_1 \quad y_2 \quad \ldots \quad y_n\,]$, *are* **equal** *if and only if* $x_1 = y_1$, $x_2 = y_2, \ldots, x_n = y_n$. Of course, two column vectors are said to be equal if the same component-by-component equalities hold.

## 1.1.1 Euclidean *n*-Space

From this point on, for reasons that will be made clear in Chapters 2 and 3, we show a preference for *column* vectors, meaning that unless otherwise indicated, the term *vector* will refer to a column vector.

**Definition 1.1.2**

For a positive integer $n$, a set of all vectors $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ with components consisting of real numbers is called **Euclidean *n*-space**, and is denoted by $\mathbb{R}^n$:

$$\mathbb{R}^n = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} : x_i \in \mathbb{R} \text{ for } i = 1, 2, \ldots, n \right\} .$$
$$= \left\{ [\,x_1 \quad x_2 \quad \cdots \quad x_n\,]^{\mathrm{T}} : x_i \in \mathbb{R} \text{ for } i = 1, 2, \ldots, n \right\}$$

(If necessary, see Appendix A for a discussion of set notation.) The positive integer $n$ is called the **dimension** of the space. For instance, the column vectors in Example 1.1.1 are elements of Euclidean 5-space, or $\mathbb{R}^5$, whereas the row vectors derived from the data table are the transposes of elements of $\mathbb{R}^3$. We will develop a deeper understanding of "dimension" in Section 1.5 and in Chapter 5.

Of course, $\mathbb{R}^1$, or just $\mathbb{R}$, is the set of all real numbers, whereas $\mathbb{R}^2$ and $\mathbb{R}^3$ are familiar geometrical entities: the sets of all points in two-dimensional and three-dimensional "space," respectively. For values of $n > 3$, we lose the geometric interpretation, but we can still deal with $\mathbb{R}^n$ algebraically.

## 1.1.2 Vector Addition/Subtraction

The next example suggests a way of combining two vectors and motivates the general definition that follows it.

### Example 1.1.2: Addition of Vectors

Suppose the components of vector $\mathbf{v}_1 = \begin{bmatrix} 322 \\ 283 \\ 304 \\ 292 \end{bmatrix}$ in $\mathbb{R}^4$ represent the revenue (in dollars) from sales of a certain item in a store during weeks 1, 2, 3, and 4, respectively, and vector $\mathbf{v}_2 = \begin{bmatrix} 187 \\ 203 \\ 194 \\ 207 \end{bmatrix}$ represents the revenue from sales of a different item during the same time period. If we combine ("add") the two vectors of data in a component-by-component way, we get the total revenue generated by the two items in the same 4 week period:

$$\mathbf{v}_1 + \mathbf{v}_2 = \begin{bmatrix} 322 \\ 283 \\ 304 \\ 292 \end{bmatrix} + \begin{bmatrix} 187 \\ 203 \\ 194 \\ 207 \end{bmatrix} = \begin{bmatrix} 322+187 \\ 283+203 \\ 304+194 \\ 292+207 \end{bmatrix} = \begin{bmatrix} 509 \\ 486 \\ 498 \\ 499 \end{bmatrix}.$$

**Definition 1.1.3**

If $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ are two vectors in $\mathbb{R}^n$, then the **sum** of $\mathbf{x}$ and $\mathbf{y}$ is the vector in $\mathbb{R}^n$, defined by

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix},$$

and the **difference** of $\mathbf{x}$ and $\mathbf{y}$ is the vector in $\mathbb{R}^n$, defined by

$$\mathbf{x} - \mathbf{y} = \begin{bmatrix} x_1 - y_1 \\ x_2 - y_2 \\ \vdots \\ x_n - y_n \end{bmatrix}.$$

Clearly, we can combine vectors only if they have the same number of components, that is, only if both vectors come from $\mathbb{R}^n$ for the same value of $n$. Because the sum (or difference) of two vectors in $\mathbb{R}^n$ is again a vector in $\mathbb{R}^n$, we say that $\mathbb{R}^n$ is **closed under addition** (or **closed under subtraction**). In other words, we do not find ourselves outside the space $\mathbb{R}^n$ when we add or subtract two vectors in $\mathbb{R}^n$.

## Example 1.1.3: Addition and Subtraction of Vectors

In $\mathbb{R}^5$,

$$\begin{bmatrix} -5 \\ 7 \\ 0 \\ -4 \\ 11 \end{bmatrix} + \begin{bmatrix} 3 \\ -2 \\ 5 \\ 4 \\ 6 \end{bmatrix} = \begin{bmatrix} -5+3 \\ 7+(-2) \\ 0+5 \\ -4+4 \\ 11+6 \end{bmatrix} = \begin{bmatrix} -2 \\ 5 \\ 5 \\ 0 \\ 17 \end{bmatrix}$$

and

$$\begin{bmatrix} -5 \\ 7 \\ 0 \\ -4 \\ 11 \end{bmatrix} - \begin{bmatrix} 3 \\ -2 \\ 5 \\ 4 \\ 6 \end{bmatrix} = \begin{bmatrix} -5-3 \\ 7-(-2) \\ 0-5 \\ -4-4 \\ 11-6 \end{bmatrix} = \begin{bmatrix} -8 \\ 9 \\ -5 \\ -8 \\ 5 \end{bmatrix}.$$

## 1.1.3 Scalar Multiplication

Another operation we need is that of *multiplying a vector by a number*. We can motivate the process in a natural way.

### Example 1.1.4: Multiplication of a Vector by a Number

Suppose the vector $\begin{bmatrix} 18.50 \\ 24 \\ 15.95 \end{bmatrix}$ in $\mathbb{R}^3$ represents the prices of three items in a store.

If the store charges $7\frac{1}{4}\%$ sales tax, then we can compute the tax on the three items as follows, producing a ''sales tax vector'':

$$\text{Sales tax} = 0.0725 \begin{bmatrix} 18.50 \\ 24 \\ 15.95 \end{bmatrix} = \begin{bmatrix} 0.0725(18.50) \\ 0.0725(24) \\ 0.0725(15.95) \end{bmatrix} = \begin{bmatrix} 1.34 \\ 1.74 \\ 1.16 \end{bmatrix}.$$

Now we can find the ''total cost'' vector:

$$\text{Total cost} = \text{Price} + \text{Sales tax} = \begin{bmatrix} 18.50 \\ 24 \\ 15.95 \end{bmatrix} + 0.0725 \begin{bmatrix} 18.50 \\ 24 \\ 15.95 \end{bmatrix}$$

$$= \begin{bmatrix} 18.50 \\ 24 \\ 15.95 \end{bmatrix} + \begin{bmatrix} 1.34 \\ 1.74 \\ 1.16 \end{bmatrix} = \begin{bmatrix} 19.84 \\ 25.74 \\ 17.11 \end{bmatrix}.$$

We can generalize this multiplication in an obvious way.

**Definition 1.1.4**

If $k$ is a real number and $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, then **scalar multiplication** of the vector $\mathbf{x}$ by the number $k$ is defined by $k\mathbf{x} = \begin{bmatrix} kx_1 \\ kx_2 \\ \vdots \\ kx_n \end{bmatrix}$. The number $k$ in this case is called a **scalar** (or **scalar quantity**) to distinguish it from a vector.

Because a scalar multiple of a vector in $\mathbb{R}^n$ is again a vector in $\mathbb{R}^n$, we say that $\mathbb{R}^n$ is **closed under scalar multiplication**.

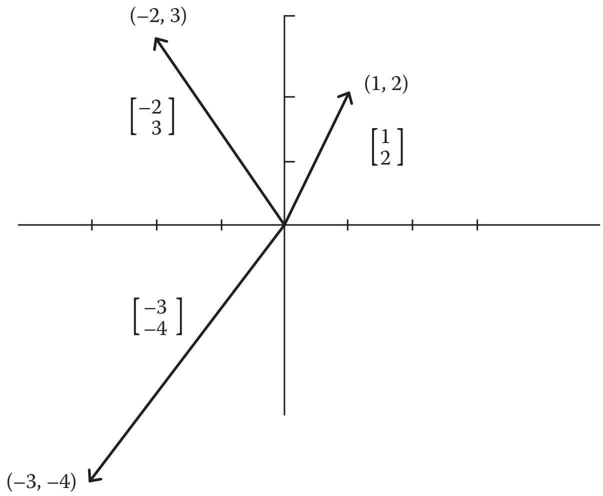



**Figure 1.1** Vectors in $\mathbb{R}^2$.

### 1.1.4 Geometric Vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$

Even though the concept of a vector is simple, vectors are very important in scientific applications. To a physicist or engineer, a vector is a quantity that has both *magnitude* (size) and *direction*, for example, velocity, acceleration, and other forces. In $\mathbb{R}^2$, we can view vectors themselves, vector addition, and scalar multiplication in a nice geometric way that is consistent with the physicist's view. A vector $\begin{bmatrix}x\\y\end{bmatrix}$ is interpreted as a directed line segment, or arrow, from the origin to the point $(x, y)$ in the usual Cartesian coordinate plane (Figure 1.1):*

The *magnitude* of a vector, a nonnegative quantity, is indicated by the length of the arrow. The *direction* of such a geometric vector is determined by the angle $\theta$ which the arrow makes with the positive $x$-axis (measured in a counterclockwise direction).

The addition of vectors is carried out according to the **Parallelogram Law**, and the sum of two vectors is usually referred to as their **resultant** (vector) (Figure 1.2).

Multiplication of a vector by a positive scalar $k$ does not change the direction of the vector, but affects its magnitude by a factor of $k$. Multiplication of a vector by a negative scalar *reverses* the vector's direction and affects its magnitude by a factor of $|k|$ (Figure 1.3).

Similarly, in $\mathbb{R}^3$, a vector $\begin{bmatrix}x\\y\\z\end{bmatrix}$ can be interpreted as an arrow connecting the origin $(0, 0, 0)$ to a point $(x, y, z)$ in the usual $x$–$y$–$z$ plane. The

* In vector analysis courses, it is usual to consider vectors that emanate from points other than the origin. See, for example, H. Davis and A. Snider, *Introduction to Vector Analysis*, 7th edn. (Dubuque, IA: Wm. C. Brown, 1995). However, we will use our definition of vectors to illustrate key linear algebra concepts, which will be generalized in this chapter.

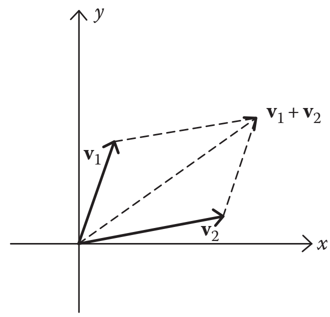



**Figure 1.2** Parallelogram law.

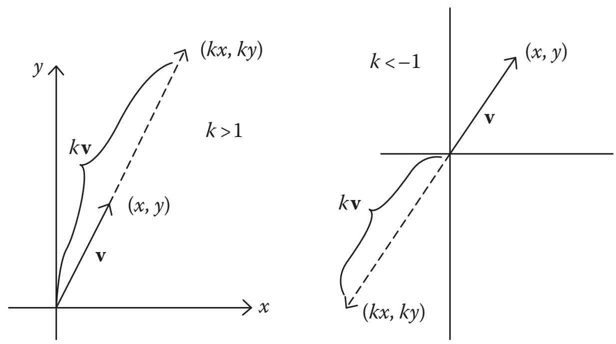



**Figure 1.3** Scalar multiplication.

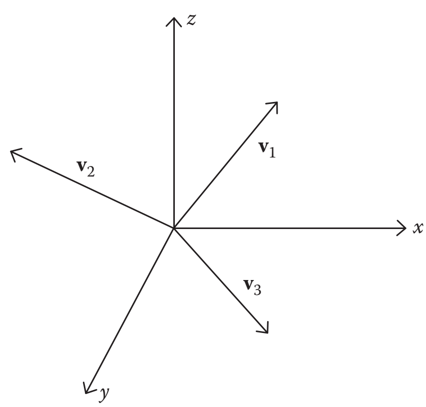



**Figure 1.4** Vectors in $\mathbb{R}^3$.

operations of addition and scalar multiplication have the same geometric meaning as in $\mathbb{R}^2$ (Figure 1.4). In Chapters 4, 5, and 8, we will return to these geometrical interpretations for motivation.

### 1.1.5 Algebraic Properties

Because the components of vectors are real numbers, we should expect the vector operations we have defined to follow the usual rules of algebra.

***Theorem 1.1.1: Properties of Vector Operations***

If $\mathbf{x}, \mathbf{y}$, and $\mathbf{z}$ are any elements of $\mathbb{R}^n$ and $k$, $k_1$, and $k_2$ are real numbers, then

(1) $\mathbf{x}+\mathbf{y}=\mathbf{y}+\mathbf{x}$. [*Commutativity of Addition*]
(2) $\mathbf{x}+(\mathbf{y}+\mathbf{z})=(\mathbf{x}+\mathbf{y})+\mathbf{z}$. [*Associativity of Addition*]
(3) There is a **zero vector**, denoted by $\mathbf{0}$, such that $\mathbf{x}+\mathbf{0}=\mathbf{x}=\mathbf{0}+\mathbf{x}$ for every $\mathbf{x}\in\mathbb{R}^n$. [*Additive Identity*]
(4) For any vector $\mathbf{x}$, there exists a vector denoted by $-\mathbf{x}$ such that $\mathbf{x}+(-\mathbf{x})=\mathbf{0}$, where $\mathbf{0}$ denotes the zero vector. [*Additive Inverse of* $\mathbf{x}$]
(5) $(k_1k_2)\mathbf{x}=k_1(k_2\mathbf{x})$. [*Associative Property of Scalar Multiplication*]
(6) $k(\mathbf{x}+\mathbf{y})=k\mathbf{x}+k\mathbf{y}$. [*Distributivity of Scalar Multiplication over Vector Addition*]
(7) $(k_1+k_2)\mathbf{x}=k_1\mathbf{x}+k_2\mathbf{x}$. [*Distributivity of Scalar Multiplication over Scalar Addition*]
(8) $1\cdot\mathbf{x}=\mathbf{x}$. [*Identity Element for Scalar Multiplication*]

We will prove some of these properties and leave the rest as exercises.

*Proof of* (1) Suppose that $\mathbf{x}=\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}$ and $\mathbf{y}=\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}$. Then

$$\mathbf{x}+\mathbf{y}=\begin{bmatrix}x_1+y_1\\x_2+y_2\\\vdots\\x_n+y_n\end{bmatrix}=\begin{bmatrix}y_1+x_1\\y_2+x_2\\\vdots\\y_n+x_n\end{bmatrix}\quad\text{[because all components are real numbers and real numbers commute]}$$

$$=\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}+\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=\mathbf{y}+\mathbf{x}.$$

*Proof of* (3) In $\mathbb{R}^n$, define $\mathbf{0}=\left.\begin{bmatrix}0\\0\\\vdots\\0\end{bmatrix}\right\}n$ zeros. Then, for any $\mathbf{x}=\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}\in\mathbb{R}^n$,

$$\mathbf{x}+\mathbf{0}=\begin{bmatrix}x_1+0\\x_2+0\\\vdots\\x_n+0\end{bmatrix}=\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}=\mathbf{x}.$$

*Proof of* (4) Given any vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, define $-\mathbf{x}$ to be the vector $\begin{bmatrix} -x_1 \\ -x_2 \\ \vdots \\ -x_n \end{bmatrix}$. It follows that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$, the zero vector. This property gives us another way to view vector subtraction: $\mathbf{x} - \mathbf{y} = \mathbf{x} + (-\mathbf{y}) = \mathbf{x} + (-1)\mathbf{y}$. In other words, subtraction can be interpreted as the addition of an additive inverse.

Note that in the proof of property (3), there are two distinct uses of the symbol zero: the scalar 0 and the zero vector $\mathbf{0}$.

## Exercises 1.1

**A.**

1. If $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \\ 8 \\ 6 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 3 \\ -5 \\ 2 \\ 0 \end{bmatrix}$, and $\mathbf{w} = \begin{bmatrix} 6 \\ 6 \\ -6 \\ -6 \end{bmatrix}$, compute each of the following vectors.
   a. $\mathbf{u} + \mathbf{w}$ b. $5\mathbf{w}$ c. $\mathbf{v} - \mathbf{u}$ d. $3\mathbf{u} + 7\mathbf{v} - 2\mathbf{w}$ e. $\frac{1}{2}\mathbf{w} + \frac{3}{4}\mathbf{u}$
   f. $\mathbf{u} - \mathbf{w} - \mathbf{v}$ g. $-2\mathbf{u} + 3\mathbf{v} - 100\mathbf{w}$

2. If $\mathbf{v}$ and $\mathbf{w}$ are vectors such that $2\mathbf{v} - \mathbf{w} = \mathbf{0}$, what is the relationship between the components of $\mathbf{v}$ and those of $\mathbf{w}$?

3. Find three vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in $\mathbb{R}^3$ such that $\mathbf{w} = 3\mathbf{u}$, $\mathbf{v} = 2\mathbf{u}$, and $2\mathbf{u} + 3\mathbf{v} + 4\mathbf{w} = \begin{bmatrix} 20 \\ 10 \\ -25 \end{bmatrix}$.

4. a. Show that the vector equation $x\begin{bmatrix} 5 \\ 7 \end{bmatrix} + y\begin{bmatrix} 3 \\ -10 \end{bmatrix} = \begin{bmatrix} -16 \\ 77 \end{bmatrix}$ represents two simultaneous linear equations in the two variables $x$ and $y$.
   b. Solve the simultaneous equations from part (a) for $x$ and $y$ and substitute into the vector equation above to check your answer.

5. Suppose that we associate with each person a vector in $\mathbb{R}^3$ having the following components: age, height, and weight. Would it make sense to add the vectors associated with two different persons? Would it make sense to multiply one of these vectors by a scalar?

6. Each day, an investment analyst records the high and low values of the price of Microsoft stock. The analyst stores the data for a given week in two vectors in $\mathbb{R}^5$ : $\mathbf{v}_{\mathrm{H}}$, giving the daily high values, and $\mathbf{v}_{\mathrm{L}}$, giving the daily low values. Find a vector expression that provides the average daily values of the price of Microsoft stock for the entire 5 day week.

## B.

1. Find nonzero real numbers $a$, $b$, and $c$ such that

$$a\mathbf{u} + b(\mathbf{u} - \mathbf{v}) + c(\mathbf{u} + \mathbf{v}) = \mathbf{0}$$

for every pair of vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^2$.

2. Prove property (2) of Theorem 1.1.1.

3. Prove property (6) of Theorem 1.1.1.

4. Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. Define $\mathbf{x} \geq \mathbf{0}$ to mean $x_1 \geq 0$ and $x_2 \geq 0$. Define $\mathbf{x} \leq \mathbf{0}$ analogously. If $(x_1 + x_2)\mathbf{x} \geq \mathbf{0}$, what must be true of $\mathbf{x}$?

5. Using the definition in the previous exercise, define $\mathbf{u} \geq \mathbf{v}$ to mean $\mathbf{u} - \mathbf{v} \geq \mathbf{0}$, where $\mathbf{u}$ and $\mathbf{v}$ are vectors having the same number of components. Consider the following vectors: $\mathbf{x} = \begin{bmatrix} 3 \\ 5 \\ -1 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 6 \\ 5 \\ 6 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 0 \\ 0 \\ -2 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix}$.
   a. Show that $\mathbf{x} \geq \mathbf{u}$.
   b. Show that $\mathbf{v} \geq \mathbf{u}$.
   c. Is there any relationship between $\mathbf{x}$ and $\mathbf{v}$?
   d. Show that $\mathbf{y} \geq \mathbf{x}, \mathbf{y} \geq \mathbf{u}$, and $\mathbf{y} \geq \mathbf{v}$.

6. Extending the definitions in Exercises B4 and B5 in the obvious way, prove the following results for vectors $\mathbf{x}, \mathbf{y}$, $\mathbf{z}$, and $\mathbf{w}$ in $\mathbb{R}^n$:
   a. If $\mathbf{x} \leq \mathbf{y}$ and $\mathbf{y} \leq \mathbf{z}$, then $\mathbf{x} \leq \mathbf{z}$.
   b. If $\mathbf{x} \leq \mathbf{y}$ and $\mathbf{z} \leq \mathbf{w}$, then $\mathbf{x} + \mathbf{z} \leq \mathbf{y} + \mathbf{w}$.
   c. If $\mathbf{x} \leq \mathbf{y}$ and $\lambda$ is a nonnegative real number, then $\lambda\mathbf{x} \leq \lambda\mathbf{y}$. If $\mathbf{x} \leq \mathbf{y}$ and $\mu$ is a negative real number, then $\mu\mathbf{x} \geq \mu\mathbf{y}$.

*A set of vectors, $S \subseteq \mathbb{R}^n$, is a* ***convex*** *set if $\mathbf{x} \in S$ and $\mathbf{y} \in S$ imply $t\mathbf{x} + (1 - t)\mathbf{y} \in S$ for all real numbers $t$ such that $0 \leq t \leq 1$. Geometrically, a set in $\mathbb{R}^n$ is convex if, whenever it contains two points (vectors), it also contains the line segment joining them. The* ***convex hull*** *of a set of vectors, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$, is the set of all vectors of the form $a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k$, where $a_i \geq 0$, $i = 1, 2, \ldots, k$, and $a_1 + a_2 + \cdots + a_k = 1$. Use these definitions in Exercises C1 through C6.*

**C.**

1. Prove that the intersection of two convex sets $S_1$ and $S_2$ is either convex or empty.

2. Show by example that the union of two convex sets does not have to be convex.

3. Is the complement of a set $\{\mathbf{x}\}$ with a single vector a convex set? Explain your answer.

4. Does a convex set in $\mathbb{R}^3$ that contains the vectors $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\begin{bmatrix} -1 \\ -3 \\ 4 \end{bmatrix}$, and $\begin{bmatrix} 0 \\ 1 \\ -7 \end{bmatrix}$ necessarily contain the vector $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$? Justify your answer.

5. Suppose that $S$ is a finite subset of $\mathbb{R}^n$. Prove the following statements.
   a. $S$ is a subset of its convex hull.
   b. The convex hull of $S$ is the smallest convex set containing $S$. [*Hint*: Suppose that $\mathcal{H}$ is the convex hull of $S$ and $S \subseteq T \subseteq \mathcal{H} \subseteq \mathbb{R}^n$, where $T$ is a convex set. Then show $\mathcal{H} \subseteq T$.]

## 1.2 The Inner Product and Norm

In Example 1.1.4, we dealt with three quantities (prices), each of which had to be multiplied by the same number (a sales tax rate). There are situations in which each of several quantities must be multiplied by a different number. The next example shows how to handle such problems using vectors.

### Example 1.2.1: A Product of Vectors

Suppose that the vector $[\,1235 \quad 985 \quad 1050 \quad 3460\,]^{\mathrm{T}}$ holds four prices expressed in Euros, British pounds, Australian dollars, and Mexican pesos, respectively. On a particular day, we know that 1 Euro = \$1.46420, 1 British pound = \$1.83637, 1 Australian dollar = \$0.83580, and 1 Mexican peso = \$0.09294. How can we use vectors to find the total of the four prices in U.S. dollars?

Arithmetically, we can calculate the total U.S. dollars as follows:

$$\begin{aligned}\text{Total} &= 1235(1.46420) + 985(1.83637) + 1050(0.83580) \\ &\quad + 3460(0.09294) \\ &= \$4816.27 \text{ (rounding to the nearest cent at the end).}\end{aligned}$$

We can interpret this answer as the result of combining two vectors—one holding the original prices and the other carrying the currency conversion rates—in a way that multiplies the vectors' corresponding components and then adds the resulting products:

$$\overbrace{\begin{bmatrix} 1235 \\ 985 \\ 1050 \\ 3460 \end{bmatrix}}^{\text{Price vector}} \bullet \overbrace{\begin{bmatrix} 1.46420 \\ 1.83637 \\ 0.83580 \\ 0.09294 \end{bmatrix}}^{\text{Conversion rate vector}}$$

$$= 1235(1.46420) + 985(1.83637) + 1050(0.83580) + 3460(0.09294)$$

$$= \overbrace{\$4816.27}^{\text{Total price (a scalar)}}.$$

---

A problem such as the one given in the last example naturally leads to a new vector operation.

**Definition 1.2.1**

If $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, then the (**Euclidean**) **inner product** (or **dot product**) of $\mathbf{x}$ and $\mathbf{y}$, denoted $\mathbf{x} \bullet \mathbf{y}$, is defined as follows:

$$\mathbf{x} \bullet \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

The dot product of two vectors is a *scalar* quantity.

In more sophisticated mathematical terms, we can describe the dot product as a function from the set $\mathbb{R}^n \times \mathbb{R}^n$ into the set $\mathbb{R}$. (A set such as $\mathbb{R}^n \times \mathbb{R}^n$ is called a *Cartesian product*. See Appendix A.)

Dot product notation is valuable in ways that have nothing to do with numerical calculations. For example, $-3x + 4y = 5$, the equation of a straight line in $\mathbb{R}^2$, can be written using the dot product: $\begin{bmatrix} -3 \\ 4 \end{bmatrix} \bullet \begin{bmatrix} x \\ y \end{bmatrix} = 5$. More generally, any linear equation, $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$, can be represented as $\mathbf{a} \bullet \mathbf{x} = b$, where $\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \in \mathbb{R}^n$ is a **vector of coefficients**, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ is a **vector of variables** (or a **vector of unknowns**),

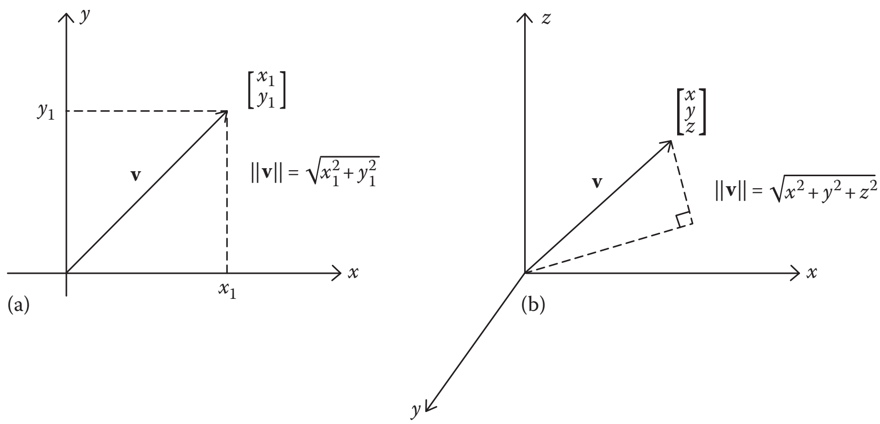



**Figure 1.5** The norm of a vector in (a) $\mathbb{R}^2$ and (b) $\mathbb{R}^3$.

and $b$ is a scalar. This representation sets the stage for important developments in Chapters 2 and 3.

Returning to the geometrical interpretation of vectors in Section 1.1, we see that if $\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix}$ is a vector in $\mathbb{R}^2$, then its **length**, often called the **norm** of $\mathbf{v}$ and written as $\|\mathbf{v}\|$, corresponds to the length of the hypotenuse of a right triangle and is given by the Pythagorean theorem as $\sqrt{x^2 + y^2}$ (Figure 1.5a).

For the same reason, the length of $\mathbf{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ in $\mathbb{R}^3$ is $\sqrt{x^2 + y^2 + z^2}$ (Figure 1.5b).

We see that these vector lengths are related to dot products:

$$\text{In } \mathbb{R}^2\text{: } \|\mathbf{v}\| = \left\| \begin{bmatrix} x \\ y \end{bmatrix} \right\| = \sqrt{x^2 + y^2} = \sqrt{\mathbf{v} \bullet \mathbf{v}}$$

$$\text{In } \mathbb{R}^3\text{: } \|\mathbf{v}\| = \left\| \begin{bmatrix} x \\ y \\ z \end{bmatrix} \right\| = \sqrt{x^2 + y^2 + z^2} = \sqrt{\mathbf{v} \bullet \mathbf{v}}$$

.

## Example 1.2.2: Norms of Vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$

If $\mathbf{v} = \begin{bmatrix} -2 \\ -3 \end{bmatrix}$, then $\| \mathbf{v} \| = \sqrt{(-2)^2 + (-3)^2} = \sqrt{\mathbf{v} \bullet \mathbf{v}} = \sqrt{13}$;

if $\mathbf{w} = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$, then $\| \mathbf{w} \| = \sqrt{1^2 + (-2)^2 + 3^2} = \sqrt{\mathbf{w} \bullet \mathbf{w}} = \sqrt{14}$.

This definition of vector length makes perfect geometric sense in $\mathbb{R}^2$ and in $\mathbb{R}^3$, and we can generalize this idea to vectors in any space $\mathbb{R}^n$.

**Definition 1.2.2**

If $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ is an element of $\mathbb{R}^n$, then we define the (**Euclidean**) **norm** (or **length**) of $\mathbf{x}$ as follows:

$$\| \mathbf{x} \| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{\mathbf{x} \bullet \mathbf{x}}.$$

The norm is a *nonnegative scalar quantity*.

The (Euclidean) norm of a vector is a consistent and useful measure of the magnitude of a vector. Furthermore, expressions such as $\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$ occur often in mathematics and its applications, and can be related to vector concepts. In the exercises following this section, we will see other ways to define a product of vectors and a norm of a vector, but Definitions 1.2.1 and 1.2.2 are the most commonly used because of their geometric interpretations in $\mathbb{R}^2$ and $\mathbb{R}^3$.

## Example 1.2.3: Norms of Vectors in $\mathbb{R}^4$ and $\mathbb{R}^5$

For $\mathbf{v} = \begin{bmatrix} 1 \\ -2 \\ 3 \\ 4 \end{bmatrix} \in \mathbb{R}^4$, we have $\|\mathbf{v}\| = \sqrt{1^2 + (-2)^2 + 3^2 + 4^2} = \sqrt{30}$;

if $\mathbf{w} = \begin{bmatrix} 5 \\ -1 \\ 2 \\ 0 \\ 4 \end{bmatrix}$ in $\mathbb{R}^5$, then $\|\mathbf{w}\| = \sqrt{5^2 + (-1)^2 + 2^2 + 0^2 + 4^2} = \sqrt{46}$.

The inner product and the norm have some important algebraic properties that we state in the following theorem.

***Theorem 1.2.1: Properties of the Inner Product and the Norm***

If $\mathbf{v}$, $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are any elements of $\mathbb{R}^n$ and $k$ is a real number, then

(a) $\mathbf{v}_1 \bullet \mathbf{v}_2 = \mathbf{v}_2 \bullet \mathbf{v}_1$
(b) $\mathbf{v}_1 \bullet (\mathbf{v}_2 + \mathbf{v}_3) = \mathbf{v}_1 \bullet \mathbf{v}_2 + \mathbf{v}_1 \bullet \mathbf{v}_3$ and $(\mathbf{v}_1 + \mathbf{v}_2) \bullet \mathbf{v}_3 = (\mathbf{v}_1 \bullet \mathbf{v}_3) + (\mathbf{v}_2 \bullet \mathbf{v}_3)$
(c) $(k\mathbf{v}_1) \bullet \mathbf{v}_2 = \mathbf{v}_1 \bullet (k\mathbf{v}_2) = k(\mathbf{v}_1 \bullet \mathbf{v}_2)$
(d) $\|k\mathbf{v}\| = |k|\|\mathbf{v}\|$

The proofs of these properties follow easily from the basic definitions and the algebraic properties of real numbers, and they appear as exercises at the end of this section.

## 1.2.1 The Angle between Vectors

Now let us get back to the geometry of $\mathbb{R}^2$, as outlined in Section 1.1. Given two vectors $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \neq \mathbf{0}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \neq \mathbf{0}$, we can focus on the angle $\theta$ between them (Figure 1.6), where the positive direction of measurement is counterclockwise.

First of all, $\theta = A - B$, so that a standard trigonometric formula gives us

$$\begin{aligned} \cos\theta &= \cos(A - B) \\ &= \cos A \cos B + \sin A \sin B \\ &= \frac{v_1}{\|\mathbf{v}\|} \times \frac{u_1}{\|\mathbf{u}\|} + \frac{v_2}{\|\mathbf{v}\|} \times \frac{u_2}{\|\mathbf{u}\|} \\ &= \frac{u_1 v_1 + u_2 v_2}{\|\mathbf{u}\|\|\mathbf{v}\|} = \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}. \end{aligned}$$

Thus, if $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ and $\theta$ is the angle between these vectors, we have $\cos\theta = \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}$. It follows that $-1 \leq \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} \leq 1$, or $|\mathbf{u} \bullet \mathbf{v}| \leq \|\mathbf{u}\|\,\|\mathbf{v}\|$.

This last inequality is a special case of the **Cauchy–Schwarz Inequality**,* whose validity for vectors in $\mathbb{R}^n$ for $n > 2$ will allow us to extend the concept of the angle between vectors.

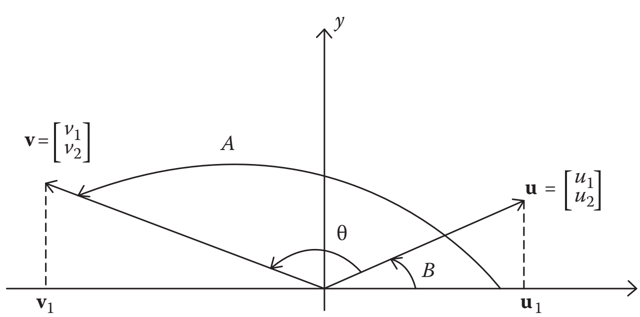



**Figure 1.6** The angle between vectors in $\mathbb{R}^2$.

* This inequality is named for the French mathematician Augustin-Louis Cauchy (1789–1857) and the German mathematician Hermann Amandus Schwarz (1843–1921). This result is sometimes called the Cauchy–Bunyakovsky–Schwarz (CBS) Inequality, acknowledging an extension of Cauchy's work by the Russian V.J. Bunyakovsky (1804–1889).

***Theorem 1.2.2: The Cauchy–Schwarz Inequality***

If $\mathbf{x} = [x_1 \ \ x_2 \ \ \ldots \ \ x_n]^{\mathrm{T}}$ and $\mathbf{y} = [y_1 \ \ y_2 \ \ \ldots \ \ y_n]^{\mathrm{T}}$ are vectors in $\mathbb{R}^n$, then

$$\boxed{|\mathbf{x} \bullet \mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\|}.$$

[In courses such as advanced analysis and statistics, the Cauchy–Schwarz inequality is often used in the form

$$|x_1y_1 + x_2y_2 + \cdots + x_ny_n| \leq (x_1^2 + x_2^2 + \cdots + x_n^2)^{1/2}(y_1^2 + y_2^2 + \cdots + y_n^2)^{1/2}.]$$

*Proof* We give a proof by induction on the dimension of $\mathbb{R}^n$ (see Appendix C). We have already established the inequality for $n = 2$. Now we assume the inequality is valid for some $n > 2$ and show that the result is true for $n + 1$.

Consider

$$\mathbf{x} = [x_1 \ \ x_2 \ \ \cdots \ \ x_n \ \ x_{n+1}]^{\mathrm{T}}, \mathbf{y} = [y_1 \ \ y_2 \ \ \cdots \ \ y_n \ \ y_{n+1}]^{\mathrm{T}} \in \mathbb{R}^{n+1}.$$

Then

$$\begin{aligned}|\mathbf{x} \bullet \mathbf{y}| &= |x_1y_1 + x_2y_2 + \cdots + x_ny_n + x_{n+1}y_{n+1}| \\ &= |(x_1y_1 + x_2y_2 + \cdots + x_ny_n) + x_{n+1}y_{n+1}| \\ &\leq |x_1y_1 + x_2y_2 + \cdots + x_ny_n| + |x_{n+1}y_{n+1}| \\ &\leq \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}\,\sqrt{y_1^2 + y_2^2 + \cdots + y_n^2} + |x_{n+1}y_{n+1}|,\end{aligned}$$

where we have used the inductive hypothesis that the inequality holds for vectors in $\mathbb{R}^n$.

If we apply the Cauchy–Schwarz inequality to the vectors $\mathbf{X} = \begin{bmatrix} \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \\ |x_{n+1}| \end{bmatrix}$ and $\mathbf{Y} = \begin{bmatrix} \sqrt{y_1^2 + y_2^2 + \cdots + y_n^2} \\ |y_{n+1}| \end{bmatrix}$ in $\mathbb{R}^2$, we find that $|\mathbf{X} \bullet \mathbf{Y}| \leq \|\mathbf{X}\|\|\mathbf{Y}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2 + x_{n+1}^2}\,\sqrt{y_1^2 + y_2^2 + \cdots + y_n^2 + y_{n+1}^2}$. Thus

$$\begin{aligned}|\mathbf{x} \bullet \mathbf{y}| &\leq \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}\sqrt{y_1^2 + y_2^2 + \cdots + y_n^2} + |x_{n+1}y_{n+1}| = |\mathbf{X} \bullet \mathbf{Y}| \\ &\leq \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2 + x_{n+1}^2}\sqrt{y_1^2 + y_2^2 + \cdots + y_n^2 + y_{n+1}^2} = \|\mathbf{x}\|\|\mathbf{y}\|.\end{aligned}$$

Thus the Cauchy–Schwarz inequality is valid in all Euclidean spaces $\mathbb{R}^n$ $(n \geq 2)$.

Equality holds in the Cauchy–Schwarz inequality if and only if one of the vectors is a scalar multiple of the other: $\mathbf{x} = c\mathbf{y}$ or $\mathbf{y} = k\mathbf{x}$ for scalars $c$ and $k$ (Exercise B11).

As a consequence of the Cauchy–Schwarz inequality, the formula for the cosine of the angle between two vectors can be extended in a natural way to $\mathbb{R}^3$ and generalized—without the geometric visualization—to the case of any two vectors in $\mathbb{R}^n$. Because the graph of $y = \cos\theta$ for

$0 \leq \theta \leq \pi$ shows that for any real number $r \in [-1, 1]$, there is a unique real number $\theta$ such that $\cos\theta = r$, we see that there is a unique real number $\theta$ such that $\cos\theta = \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|}$, $0 \leq \theta \leq \pi$, for any nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$.

**Definition 1.2.3**

If $\mathbf{u}$ and $\mathbf{v}$ are nonzero elements of $\mathbb{R}^n$, then we define the **angle** $\theta$ **between u and v** as the unique angle between 0 and $\pi$ rad inclusive, satisfying

$$\cos\theta = \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{u}\|\,\|\mathbf{v}\|}.$$

## Example 1.2.4: The Angle between Two Vectors in $\mathbb{R}^3$

Suppose we want to find the angle between the two vectors in $\mathbb{R}^3$

$$\mathbf{u} = \begin{bmatrix} 1 \\ \sqrt{3} \\ 2 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} 2\sqrt{3} \\ 2 \\ \sqrt{3} \end{bmatrix}.$$

We have

$$\cos\theta = \frac{\mathbf{u} \bullet \mathbf{v}}{\|\mathbf{u}\|\,\|\mathbf{v}\|} = \frac{\begin{bmatrix} 1 \\ \sqrt{3} \\ 2 \end{bmatrix} \bullet \begin{bmatrix} 2\sqrt{3} \\ 2 \\ \sqrt{3} \end{bmatrix}}{\left\|\begin{bmatrix} 1 \\ \sqrt{3} \\ 2 \end{bmatrix}\right\| \left\|\begin{bmatrix} 2\sqrt{3} \\ 2 \\ \sqrt{3} \end{bmatrix}\right\|}$$

$$= \frac{2\sqrt{3} + 2\sqrt{3} + 2\sqrt{3}}{\sqrt{1^2 + (\sqrt{3})^2 + 2^2}\sqrt{(2\sqrt{3})^2 + 2^2 + (\sqrt{3})^2}}$$

$$= \frac{6\sqrt{3}}{\sqrt{8} \cdot \sqrt{19}} = \frac{3\sqrt{3}}{\sqrt{38}} = \frac{3\sqrt{114}}{38},$$

and $\theta = \arccos\left(\frac{3\sqrt{114}}{38}\right) \approx 0.5681$ rad or $32.55^{\circ}$.

We can write the formula in Definition 1.2.3 as

$$\boxed{\mathbf{u} \bullet \mathbf{v} = \|\mathbf{u}\|\,\|\mathbf{v}\| \cos\theta}.$$

If vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$ are *perpendicular* to each other, then $\theta = \pi/2$ rad and $\cos\theta = 0$, so $\mathbf{u} \bullet \mathbf{v} = 0$. On the other hand, if $\mathbf{u}$ and $\mathbf{v}$ are nonzero, it is obvious that $\|\mathbf{u}\| > 0$ and $\|\mathbf{v}\| > 0$, so the only way for $\mathbf{u} \bullet \mathbf{v}$ to equal 0 is if $\cos\theta = 0$, that is, if $\theta = \pi/2$ and $\mathbf{u}$ and $\mathbf{v}$ are therefore perpendicular. (Definition 1.2.3 requires $0 \leq \theta \leq \pi$.) In later discussions, a generalization of the concept of perpendicularity will be important, so we make a formal definition.

**Definition 1.2.4**

Two vectors **u** and **v** in $\mathbb{R}^n$ are called **orthogonal*** (or **perpendicular** if $n = 2$, 3) if $\mathbf{u} \bullet \mathbf{v} = 0$. In this case, we write $\mathbf{u} \perp \mathbf{v}$ (pronounced "u perp v").

## Example 1.2.5: Orthogonal Vectors in $\mathbb{R}^5$

In $\mathbb{R}^5$, if $\mathbf{u} = \begin{bmatrix} 1 \\ -2 \\ 3 \\ -4 \\ 5 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 10 \\ -4 \\ 1 \\ -1 \\ -5 \end{bmatrix}$, then $\mathbf{u} \perp \mathbf{v}$ because

$$\begin{bmatrix} 1 \\ -2 \\ 3 \\ -4 \\ 5 \end{bmatrix} \bullet \begin{bmatrix} 10 \\ -4 \\ 1 \\ -1 \\ -5 \end{bmatrix} = 10 + 8 + 3 + 4 - 25 = 0.$$

Vector algebra is often used to prove theorems in geometry.

## Example 1.2.6: A Geometry Theorem

Let us prove that *the diagonals of a rhombus are perpendicular*. We recall that a rhombus is a parallelogram with all four sides having the same length (Figure 1.7).

We may assume that the rhombus is the figure OABC, positioned so that its lower left-hand vertex is at the origin and its

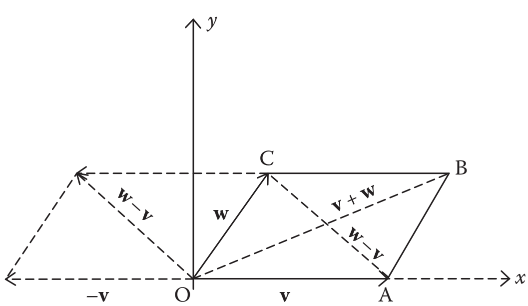



**Figure 1.7** A rhombus.

* The word *orthogonal* comes from Greek words meaning "straight" and "angle."

bottom side lies along the $x$-axis. Let $\mathbf{v}$ denote the vector from O to A and let $\mathbf{w}$ denote the vector from the origin to C.

Then the diagonal $\overline{OB}$ represents $\mathbf{v}+\mathbf{w}$ (by the Parallelogram Law). Because the side $\overline{AB}$ has the same direction and length as $\mathbf{w}$, this side can also represent the vector $\mathbf{w}$.

We can see that the diagonal $\overline{AC}$ has the same magnitude and direction as $\mathbf{w}-\mathbf{v}=\mathbf{w}+(-\mathbf{v})$ (again, by the Parallelogram Law). Then, using Definition 1.2.2 and Theorem 1.2.1,

$$\begin{aligned}(\mathbf{w}-\mathbf{v})\bullet(\mathbf{w}+\mathbf{v}) &= (\mathbf{w}-\mathbf{v})\bullet\mathbf{w}+(\mathbf{w}-\mathbf{v})\bullet\mathbf{v}\\ &= \mathbf{w}\bullet\mathbf{w}-\mathbf{v}\bullet\mathbf{w}+\mathbf{w}\bullet\mathbf{v}-\mathbf{v}\bullet\mathbf{v}\\ &= \|\mathbf{w}\|^2-\|\mathbf{v}\|^2.\end{aligned}$$

Because all sides of a rhombus are equal, it follows that $\|\mathbf{w}\|=\|\mathbf{v}\|$ and so $(\mathbf{w}-\mathbf{v})\bullet(\mathbf{w}+\mathbf{v})=0$. Thus vectors $\mathbf{w}-\mathbf{v}$ and $\mathbf{w}+\mathbf{v}$ are orthogonal, indicating that diagonals $\overline{OB}$ and $\overline{AC}$ must be perpendicular, which is what we wanted to prove.

## Exercises 1.2

### A.

1. Let $\mathbf{u}=\begin{bmatrix}2\\-1\\2\end{bmatrix}$, $\mathbf{v}=\begin{bmatrix}5\\0\\2\end{bmatrix}$, $\mathbf{x}=\begin{bmatrix}-7\\1\\2\end{bmatrix}$, and $\mathbf{y}=\begin{bmatrix}0\\8\\3\end{bmatrix}$.

   Compute the following.
   a. $(\mathbf{u}+\mathbf{v})\bullet(\mathbf{x}+\mathbf{y})$ b. $((3\mathbf{u}\bullet\mathbf{x})\mathbf{v})\bullet\mathbf{y}$ c. $\mathbf{u}\bullet\mathbf{x}-4\mathbf{v}\bullet\mathbf{y}$
   d. $\mathbf{u}\bullet\mathbf{x}+3\mathbf{u}\bullet\mathbf{y}-\mathbf{v}\bullet\mathbf{y}$ e. $(2(\mathbf{v}+\mathbf{u})\bullet\mathbf{y})-5\mathbf{u}\bullet\mathbf{y}$
   f. $4\mathbf{u}\bullet\mathbf{x}+6[\mathbf{v}\bullet(3\mathbf{x}-\mathbf{y})]$

2. If $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are in $\mathbb{R}^n$, does the "law"

$$\mathbf{v}_1\bullet(\mathbf{v}_2\bullet\mathbf{v}_3)=(\mathbf{v}_1\bullet\mathbf{v}_2)\bullet\mathbf{v}_3$$

   make sense? Explain.

3. Calculate $\|\mathbf{u}\|$ for each of the following vectors.

   a. $\mathbf{u}=\begin{bmatrix}-3\\5\end{bmatrix}$ b. $\mathbf{u}=\begin{bmatrix}1\\0\\-7\end{bmatrix}$ c. $\mathbf{u}=\begin{bmatrix}2\\2\\2\\2\end{bmatrix}$

   d. $\mathbf{u}=\begin{bmatrix}1\\-1\\2\\-2\\3\end{bmatrix}$ e. $\mathbf{u}=\begin{bmatrix}1\\2\\3\\4\\5\\6\end{bmatrix}$ f. $\mathbf{u}=\begin{bmatrix}-2\\3\\1\\\sqrt{10}\\4\\-3\end{bmatrix}$

4. What is the length of the vector $\begin{bmatrix} \cos a \cos b \\ \cos a \sin b \\ \sin a \end{bmatrix}$, where $a$ and $b$ are any real numbers? (Your answer should be independent of $a$ and $b$.)

5. For each pair of vectors **u**, **v**, calculate the angle between **u** and **v**. Express your answer in both radians and degrees.

a. $\mathbf{u} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 5 \\ 12 \end{bmatrix}$ b. $\mathbf{u} = \begin{bmatrix} 4 \\ 6 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} -3 \\ 2 \end{bmatrix}$

c. $\mathbf{u} = \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 5 \\ -4 \\ 2 \end{bmatrix}$ d. $\mathbf{u} = \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 5 \\ -3 \\ 2 \end{bmatrix}$

e. $\mathbf{u} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}$ f. $\mathbf{u} = \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 3 \\ 4 \\ -1 \end{bmatrix}$

g. $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} -2 \\ 1 \\ -3 \\ 4 \end{bmatrix}$ h. $\mathbf{u} = \begin{bmatrix} 2 \\ 0 \\ 1 \\ 0 \\ 3 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} -1 \\ 4 \\ -2 \\ 3 \\ 0 \end{bmatrix}$

6. The angle between the vectors $\begin{bmatrix} 1 \\ 7 \\ b \end{bmatrix}$ and $\begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix}$ is given by $\arccos(1/3)$. Find $b$.

7. Suppose that **u** and **v** are nonzero vectors in $\mathbb{R}^n$ and $\alpha$, $\beta$ are positive real numbers. Show that the angle between $\alpha\mathbf{u}$ and $\beta\mathbf{v}$ is the same as the angle between **u** and **v**.

8. Find the angles between the sides and a diagonal of a $1 \times 2$ rectangle.

9. Suppose that $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Find a real number $\alpha$ so that $\mathbf{u} - \alpha\mathbf{v}$ is orthogonal to **v**, given that $\|\mathbf{v}\| = 1$.

10. Given two vectors $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$ in $\mathbb{R}^3$, the **cross product** (also called the **vector product**), $\mathbf{u} \times \mathbf{v}$, is the vector defined as

$$\mathbf{u} \times \mathbf{v} = \begin{bmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{bmatrix}^{*}.$$

* The cross product is a useful concept in many physics and engineering applications—for example, in discussing *torque*. See, for example, D. Zill and M. Cullen, *Advanced Engineering Mathematics*, 7th edn. (Sudbury, MA: Jones and Bartlett, 2000). In contrast to the dot product, the cross product of two vectors in $\mathbb{R}^3$ results in a vector.

Calculate $\mathbf{u} \times \mathbf{v}$ in each of the following cases.

a. $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix}$

b. $\mathbf{u} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix}$

c. $\mathbf{u} = \begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

d. $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

11. Let $\mathbf{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\mathbf{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, and $\mathbf{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Using the definition of the *cross product* given in Exercise A10, fill in the entries of the following "multiplication table."

| $\times$ | **i** | **j** | **k** |
|---|---|---|---|
| **i** | | | |
| **j** | | | |
| **k** | | | |

## B.

1. Prove property (a) of Theorem 1.2.1 for vectors in $\mathbb{R}^3$.

2. Prove property (b) of Theorem 1.2.1 for vectors in $\mathbb{R}^3$.

3. Prove property (c) of Theorem 1.2.1 for vectors in $\mathbb{R}^3$.

4. Prove property (d) of Theorem 1.2.1 for vectors in $\mathbb{R}^3$.

5. For what values of $a$ are the vectors $\begin{bmatrix} -6 \\ a \\ 2 \end{bmatrix}$ and $\begin{bmatrix} a \\ a^2 \\ a \end{bmatrix}$ orthogonal?

6. For what values of $a$ is the angle between the vectors $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 0 \\ a \end{bmatrix}$ equal to $60°$?

7. Let $\mathbf{v}$ and $\mathbf{w}$ be vectors of length 1 along the sides of a given angle. Show that the vector $\mathbf{v} + \mathbf{w}$ bisects the angle.

8. If $a$ is a positive real number, find the cosines of the angles between the vector $\mathbf{v} = \begin{bmatrix} a \\ a \\ \vdots \\ a \end{bmatrix}$ and the vectors $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$ in $\mathbb{R}^n$. What are the actual angles in $\mathbb{R}^2$ and $\mathbb{R}^3$?

9. In $\mathbb{R}^3$, find the angle between a diagonal of a cube and one of its edges.

10. Find and prove a formula for $\|\mathbf{u}\|$, where

a. $\mathbf{u} = \begin{bmatrix} \sqrt{1} \\ \sqrt{2} \\ \sqrt{3} \\ \vdots \\ \sqrt{n} \end{bmatrix} \in \mathbb{R}^n$ b. $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ \vdots \\ n \end{bmatrix} \in \mathbb{R}^n$.

(You may want to consult the discussion of mathematical induction in Appendix C.)

11. Prove that equality holds in the Cauchy–Schwarz inequality (i.e., $|\mathbf{x} \bullet \mathbf{y}| = \|\mathbf{x}\|\|\mathbf{y}\|$) if and only if $\mathbf{x} = c\mathbf{y}$ or $\mathbf{y} = k\mathbf{x}$, where $c$ and $k$ are scalars.

12. Under what conditions on two vectors $\mathbf{u}$ and $\mathbf{v}$ is it true that

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2?$$

Justify your answer and interpret the equation geometrically in $\mathbb{R}^2$ and $\mathbb{R}^3$.

13. If we know that $\mathbf{u} \bullet \mathbf{v} = \mathbf{u} \bullet \mathbf{w}$ for vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^n$, does it follow that $\mathbf{v} = \mathbf{w}$? If it does, give a proof. Otherwise, find a set of vectors $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ in some space $\mathbb{R}^n$, for which $\mathbf{u} \bullet \mathbf{v} = \mathbf{u} \bullet \mathbf{w}$ but $\mathbf{v} \neq \mathbf{w}$.

14. Prove that if $\mathbf{u}$ is orthogonal to both $\mathbf{v}$ and $\mathbf{w}$, then $\mathbf{u}$ is orthogonal to $a\mathbf{v} + b\mathbf{w}$ for all scalars $a$ and $b$.

15. Prove that $\|\mathbf{u} + \mathbf{v}\| = \|\mathbf{u} - \mathbf{v}\|$ if and only if $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

16. Suppose that $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$, and $\mathbf{x}$ are vectors in $\mathbb{R}^n$ such that $\mathbf{u}+\mathbf{v}+\mathbf{w}+\mathbf{x}=\mathbf{0}$. Prove that $\mathbf{u}+\mathbf{v}$ is orthogonal to $\mathbf{u}+\mathbf{x}$ if and only if

$$\|\mathbf{u}\|^2+\|\mathbf{w}\|^2=\|\mathbf{v}\|^2+\|\mathbf{x}\|^2.^*$$

[*Hint*: Show that $\mathbf{w}\bullet\mathbf{w}=\|\mathbf{v}\|^2+\|\mathbf{x}\|^2-\|\mathbf{u}\|^2+2(\mathbf{u}\bullet\mathbf{u}+\mathbf{u}\bullet\mathbf{v}+\mathbf{u}\bullet\mathbf{x}+\mathbf{x}\bullet\mathbf{v})$.]

17. If $\mathbf{u}=\begin{bmatrix}u_1\\u_2\\\vdots\\u_n\end{bmatrix}\in\mathbb{R}^n$, define $\mathbf{u}\geq\mathbf{0}$ to mean that $u_i\geq 0$ for $i=1,2,\ldots,n$; also define $\mathbf{u}\geq\mathbf{v}$ to mean $\mathbf{u}-\mathbf{v}\geq\mathbf{0}$, where $\mathbf{u},\mathbf{v}\in\mathbb{R}^n$. (See Exercises 1.1, B4 through B6.) Now let $\mathbf{x},\mathbf{y}$, and $\mathbf{z}$ be vectors in $\mathbb{R}^n$, with $\mathbf{x}>\mathbf{0}$, $\mathbf{y}\geq\mathbf{z}$, and $\mathbf{y}\neq\mathbf{z}$. Prove that $\mathbf{x}\bullet\mathbf{y}>\mathbf{x}\bullet\mathbf{z}$.

18. Define the **distance** $d$ between two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ as

$$d(\mathbf{u},\mathbf{v})=\|\mathbf{u}-\mathbf{v}\|.$$

Prove that
a. $d(\mathbf{u},\mathbf{v})\geq 0$.
b. $d(\mathbf{u},\mathbf{v})=0$ if and only if $\mathbf{u}=\mathbf{v}$.
c. $d(\mathbf{u},\mathbf{v})=d(\mathbf{v},\mathbf{u})$.

19. Using the definition of the *cross product* given in Exercise A10, prove that
a. $\mathbf{u}\times\mathbf{v}=-(\mathbf{v}\times\mathbf{u})$ b. $\mathbf{u}\times(\mathbf{v}+\mathbf{w})=\mathbf{u}\times\mathbf{v}+\mathbf{u}\times\mathbf{w}$.
c. $k(\mathbf{u}\times\mathbf{v})=(k\mathbf{u})\times\mathbf{v}=\mathbf{u}\times(k\mathbf{v})$, where $k$ is a scalar.
d. $\mathbf{u}\times\mathbf{u}=0$ for every $\mathbf{u}\in\mathbb{R}^3$.

20. Using the definition of the *cross product* given in Exercise A10, prove that $\mathbf{u}\times(\mathbf{v}\times\mathbf{w})=(\mathbf{u}\bullet\mathbf{w})\mathbf{v}-(\mathbf{u}\bullet\mathbf{v})\mathbf{w}$.

21. Using the definition of the *cross product* given in Exercise A10, prove that $\mathbf{u}\times\mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$.

22. Using the definition of the *cross product* given in Exercise A10, prove *Lagrange's identity*: $\|\mathbf{u}\times\mathbf{v}\|^2=\|\mathbf{u}\|^2\|\mathbf{v}\|^2-(\mathbf{u}\bullet\mathbf{v})^2$.

---

* This result has geometric significance in $\mathbb{R}^2$. I am indebted to my colleague, Mahmoud Sayrafiezadeh, for calling it to my attention.

## C.

1. In 1968, Shmuel Winograd published* a more efficient method to calculate inner products of vectors in $\mathbb{R}^n$. Efficiency in this situation means fewer multiplications in certain types of problems. If $x$ is a real number, the notation $[[x]]$ denotes the largest integer less than or equal to $x$, for example, $[[\pi]] = 3$ and $[[-\pi]] = -4$. For two vectors $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ in $\mathbb{R}^n$,

$$\mathbf{x}\bullet\mathbf{y} = \begin{cases} \sum_{j=1}^{[[n/2]]} (x_{2j-1} + y_{2j})(x_{2j} + y_{2j-1}) - \alpha - \beta, \text{ for } n \text{ even,} \\ \sum_{j=1}^{[[n/2]]} (x_{2j-1} + y_{2j})(x_{2j} + y_{2j-1}) - \alpha - \beta + x_n y_n, \text{ for } n \text{ odd,} \end{cases}$$

where $\alpha = \sum_{j=1}^{[[n/2]]} x_{2j-1}x_{2j}, \quad \beta = \sum_{j=1}^{[[n/2]]} y_{2j-1}y_{2j}.$

The cleverness of this method lies in the fact that if you have to compute the inner products of a vector $\mathbf{x}$ with several other vectors, the calculation of $\alpha$ (or $\beta$) in Winograd's formula can be done just once (and stored) instead of using the vector $\mathbf{x}$ each time.

Suppose we want to calculate each of the 16 inner products, $\mathbf{u}_i \bullet \mathbf{v}_j$, for $i, j = 1, 2, 3, 4$, where each $\mathbf{u}_i$ and $\mathbf{v}_j$ is a vector in $\mathbb{R}^4$.

a. If you use Definition 1.2.1, how many multiplications and how many additions/subtractions are needed to calculate the 16 inner products?

b. If you use Winograd's formula, how many multiplications and how many additions/subtractions will you need to calculate the 16 inner products? (Remember that you do not have to repeat calculations for $\alpha$ or $\beta$ if you have already done them.)

Compare your answers with those in part (a).

2. Prove the **Cauchy–Schwarz Inequality** (Theorem 1.2.2) for vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ as follows:

a. Assume $\mathbf{y} \neq \mathbf{0}$ and let $t$ be an arbitrary parameter (a real number). Show that $0 \leq \|\mathbf{x} - t\mathbf{y}\|^2 = \|\mathbf{x}\|^2 - 2t(\mathbf{x}\bullet\mathbf{y}) + t^2 \|\mathbf{y}\|^2$.

b. Let $t = (\mathbf{x}\bullet\mathbf{y})/\|\mathbf{y}\|^2$ and show that $0 \leq \|\mathbf{x}\|^2 - \frac{|\mathbf{x}\bullet\mathbf{y}|^2}{\|\mathbf{y}\|^2}$.

c. Use the result of part (b) to establish that $|\mathbf{x}\bullet\mathbf{y}| \leq \|\mathbf{x}\| \, \|\mathbf{y}\|$.

* S. Winograd, A new algorithm for inner product, *IEEE Trans. Comput.* **17** (1968): 693–694.

3. Prove the **Triangle Inequality** for vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$:

$$\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|.$$

[*Hint*: Consider $\|\mathbf{u}+\mathbf{v}\|^2$ and use the Cauchy–Schwarz inequality.]

4. Prove that $\|\mathbf{u}-\mathbf{v}\| \geq \|\mathbf{u}\| - \|\mathbf{v}\|$ for all vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$. [*Hint*: Use the result of Exercise C3.]

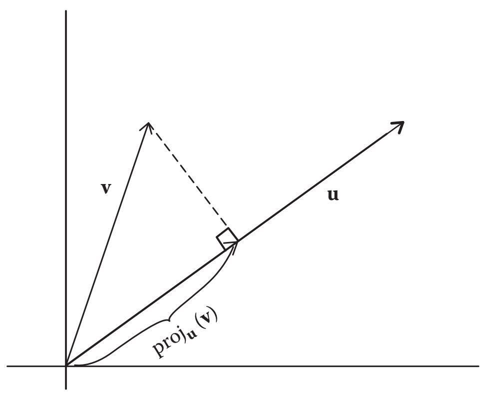
5. Prove the **Parallelogram Law** for vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$:

$$\|\mathbf{u}+\mathbf{v}\|^2 + \|\mathbf{u}-\mathbf{v}\|^2 = 2\|\mathbf{u}\|^2 + 2\|\mathbf{v}\|^2.$$

6. Prove the **Polarization Identity** for vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$:

$$\mathbf{u} \bullet \mathbf{v} = \frac{1}{4}\|\mathbf{u}+\mathbf{v}\|^2 - \frac{1}{4}\|\mathbf{u}-\mathbf{v}\|^2 .$$

*If* $\mathbf{u}$ *and* $\mathbf{v}$ *are vectors in* $\mathbb{R}^n$ *and* $\mathbf{u} \neq \mathbf{0}$, *then the* ***orthogonal projection of*** $\mathbf{v}$ ***onto*** $\mathbf{u}$ *is the vector* $\mathbf{proj_u}(\mathbf{v})$ *defined by* $\mathbf{p} = \mathbf{proj_u}(\mathbf{v}) = \left(\frac{\mathbf{u}\bullet\mathbf{v}}{\mathbf{u}\bullet\mathbf{u}}\right) \cdot \mathbf{u}$. *Here's an illustration of the orthogonal projection of a vector* $\mathbf{v}$ *onto another vector* $\mathbf{u}$ *in* $\mathbb{R}^2$*:*




7. Prove that $\mathbf{u}$ is orthogonal to $\mathbf{v} - \mathbf{proj_u}(\mathbf{v})$ for all vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, where $\mathbf{u} \neq \mathbf{0}$.

8. Prove that $\mathbf{proj_u}(\mathbf{v} - \mathbf{proj_u}(\mathbf{v})) = \mathbf{0}$.

9. Prove that $\mathbf{proj_u}\,(\mathbf{proj_u}(\mathbf{v})) = \mathbf{proj_u}(\mathbf{v})$.

10. Define the **sum norm**, or $\ell_1$ **norm**, of a vector $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ as follows:

$$\|\mathbf{u}\|_1 = |u_1| + |u_2| + \cdots + |u_n|.$$

Assuming the *Triangle Inequality* for real numbers $x$ and $y$, $|x+y| \le |x| + |y|$, prove that $\|\mathbf{u}+\mathbf{v}\|_1 \le \|\mathbf{u}\|_1 + \|\mathbf{v}\|_1$.

11. Define the **max norm**, or $\ell_\infty$ **norm**, of a vector $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ as follows:

$$\begin{aligned}\|\mathbf{u}\|_\infty &= \max\{|u_1|, |u_2|, \ldots, |u_n|\} \\ &= \text{the largest of the numbers } |u_i|.\end{aligned}$$

Assuming the Triangle Inequality (see Exercise C3) for real numbers, prove that

$$\|\mathbf{u}+\mathbf{v}\|_\infty \le \|\mathbf{u}\|_\infty + \|\mathbf{v}\|_\infty.$$

## 1.3 Spanning Sets

Given a set of vectors in some Euclidean space $\mathbb{R}^n$, we know how to combine these vectors using the operations of vector addition and scalar multiplication. The set of all vectors resulting from such combinations is important in both theory and applications.

**Definition 1.3.1**

Given a nonempty finite set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ in $\mathbb{R}^n$, a **linear combination** of these vectors is any vector of the form $a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k$, where $a_1, a_2, \ldots, a_k$ are scalars.

In each space $\mathbb{R}^n$, there are special sets of vectors that play an important role in describing the space. For example, in $\mathbb{R}^2$ the vectors $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ have the significant property that *any* vector $\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix}$ can be written as a linear combination of vectors $\mathbf{e}_1$ and $\mathbf{e}_2$: $\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix} = x\mathbf{e}_1 + y\mathbf{e}_2$. In physics and engineering, the symbols $\mathbf{i}$ and $\mathbf{j}$ (or $\vec{\mathbf{i}}$ and $\vec{\mathbf{j}}$) are often used for $\mathbf{e}_1$ and $\mathbf{e}_2$, respectively. If we extend (stretch) the vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ using multiplication by positive and negative scalars, we get the usual $x$- and $y$-axes in the Euclidean plane (Figure 1.8).

Similarly, any vector $\mathbf{w} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ in $\mathbb{R}^3$ can be expressed uniquely in terms of the vectors $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, and $\mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$:

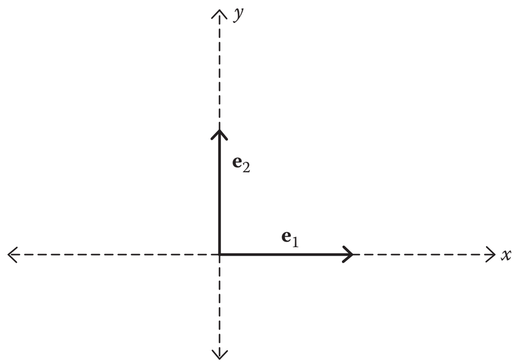



**Figure 1.8** The vectors $\mathbf{e_1}$ and $\mathbf{e_2}$ in $\mathbb{R}^2$.

$\mathbf{w} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3$. In applied courses, these three vectors are often denoted by **i**, **j**, and **k** (or by $\vec{\mathbf{i}}$, $\vec{\mathbf{j}}$, and $\vec{\mathbf{k}}$), respectively.

Generalizing, any vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ in $\mathbb{R}^n$ can be written as a linear combination of the $n$ vectors $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$, $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$: $\mathbf{v} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n$.

The special set $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ will be used often in our analysis of $\mathbb{R}^n$. This set of vectors is a particular example of a **spanning set** for a particular space $\mathbb{R}^n$—appropriately named because such a set reaches across (spans) the entire space when all linear combinations of vectors in the set are formed.

## Definition 1.3.2

Given a nonempty set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ in $\mathbb{R}^n$, the **span** of $S$, denoted by *span*($S$), is the set of all linear combinations of vectors from $S$:

$$span(S) = \{a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k | a_i \in \mathbb{R},\ \mathbf{v}_i \in S,\ i = 1, 2, \ldots, k\}.$$

(Note that some or all of the scalars may be zero.)

A nonempty set $S$ of vectors in $\mathbb{R}^n$ **spans** $\mathbb{R}^n$ (or is a **spanning set** for $\mathbb{R}^n$) if every vector in $\mathbb{R}^n$ is an element of *span*($S$), that is, if every vector in $\mathbb{R}^n$ is a linear combination of vectors in $S$.

Note that *span*($S$) is an infinite set of vectors unless $S = \{\mathbf{0}\}$, in which case *span*($S$) $= S = \{\mathbf{0}\}$.

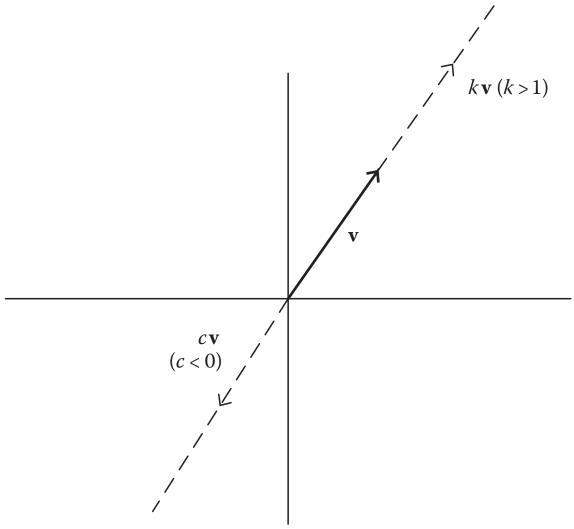



**Figure 1.9** The span of a nonzero vector in $\mathbb{R}^2$.

Given a single nonzero vector $\mathbf{v} = \begin{bmatrix} x \\ y \end{bmatrix}$ in $\mathbb{R}^2$, the span of $\{\mathbf{v}\}$ is the set of all scalar multiples of $\mathbf{v}$: *span*$(\{\mathbf{v}\}) = \left\{ a\mathbf{v} = \begin{bmatrix} ax \\ ay \end{bmatrix} : a \in \mathbb{R} \right\}$. Because $a = 0$ is a possibility, the origin is in the span and we can interpret the span of $\{\mathbf{v}\}$ as *the set of all points on a straight line through the origin* (see Figure 1.9). The slope of this line depends on the original coordinates of $\mathbf{v}$: Slope $= ay/ax = y/x$, assuming that $a \neq 0$ and $x \neq 0$. If $x = 0$, the span is just the $y$-axis, whereas if $y = 0$, the span is the $x$-axis.

Suppose we have two nonzero vectors $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^2$. If one of these vectors is a scalar multiple of the other, then span$(\{\mathbf{v}, \mathbf{w}\}) =$ *span*$(\{\mathbf{v}\}) =$ *span*$(\{\mathbf{w}\})$, again a straight line through the origin. On the other hand, if the vectors do not lie on the same straight line through the origin, their span is *all of* $\mathbb{R}^2$ (see Figure 1.10), a fact we will prove after Theorem 1.4.2.

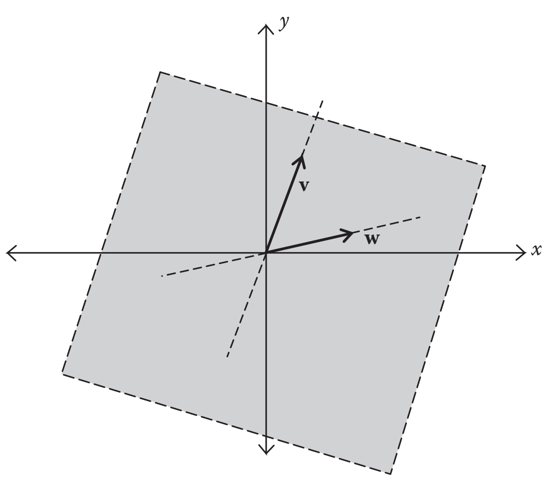



**Figure 1.10** The span of linearly independent vectors in $\mathbb{R}^2$.

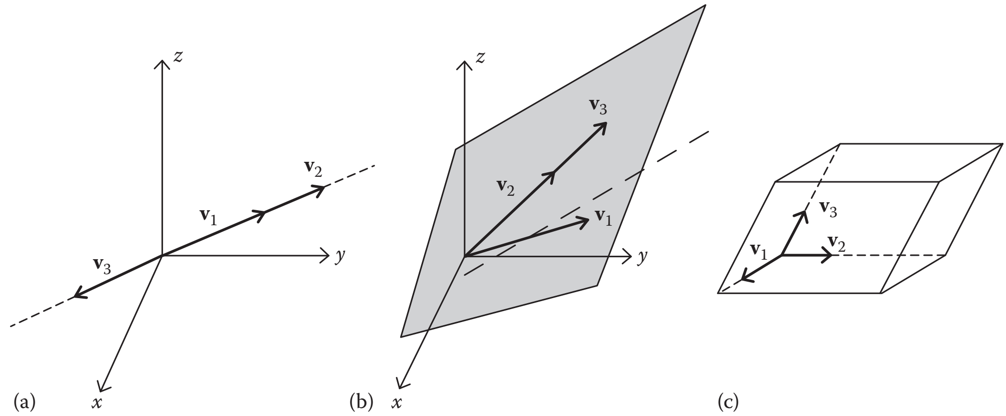



**Figure 1.11** Spans of three nonzero vectors in $\mathbb{R}^3$.

In $\mathbb{R}^3$ the span of a single nonzero vector is a line through the origin in a three-dimensional space. Given two nonzero vectors in $\mathbb{R}^3$, their span is either a line through the origin or a plane passing through the origin, depending on whether the vectors are scalar multiples of each other.

Finally, three nonzero vectors in $\mathbb{R}^3$ span a line, a plane, or all of $\mathbb{R}^3$, depending on whether all three of the vectors lie on the same straight line through the origin, only two of the vectors are collinear, or no two vectors lie on the same straight line, respectively (Figure 1.11a through c, respectively).

## Example 1.3.1: The Span of a Set of Vectors in $\mathbb{R}^2$

(a) Let us determine the span of the three vectors, $\begin{bmatrix} 1 \\ -2 \end{bmatrix}, \begin{bmatrix} -3 \\ 6 \end{bmatrix}, \begin{bmatrix} 4 \\ -8 \end{bmatrix}$. According to Definition 1.3.2, the span of this set of vectors is the set $\left\{ a\begin{bmatrix} 1 \\ -2 \end{bmatrix} + b\begin{bmatrix} -3 \\ 6 \end{bmatrix} + c\begin{bmatrix} 4 \\ -8 \end{bmatrix} : a, b, c \in \mathbb{R} \right\}$ $= \left\{ \begin{bmatrix} a - 3b + 4c \\ -2a + 6b - 8c \end{bmatrix} : a, b, c \in \mathbb{R} \right\}$. If we examine the vectors in the span carefully, we see that no matter how the scalars $a$, $b$, and $c$ vary, the second component is always $-2$ times the first component. This means that we can express the span as $\left\{ \begin{bmatrix} x \\ -2x \end{bmatrix} : x \in \mathbb{R} \right\}$, or as $\left\{ x\begin{bmatrix} 1 \\ -2 \end{bmatrix} : x \in \mathbb{R} \right\}$. Thus the span of our set of three vectors consists of all scalar multiples of one of them, that is, the span is a straight line through the origin with slope equal to $-2$. (Going back to the original set of vectors, we notice that all three vectors lie on the same straight line through the origin.)

(b) Now let us consider the span of the set $S=\left\{\begin{bmatrix}1\\-2\end{bmatrix},\begin{bmatrix}-3\\6\end{bmatrix},\begin{bmatrix}2\\1\end{bmatrix}\right\}$. We have $span(S)=\left\{a\begin{bmatrix}1\\-2\end{bmatrix}+b\begin{bmatrix}-3\\6\end{bmatrix}+c\begin{bmatrix}2\\1\end{bmatrix}:a,b,c\in\mathbb{R}\right\}=\left\{\begin{bmatrix}a-3b+2c\\-2a+6b+c\end{bmatrix}:a,b,c\in\mathbb{R}\right\}$, and there is no apparent connection between the components of the vectors in the span.

However, we can write

$$\begin{aligned} span(S) &= \left\{\begin{bmatrix}a-3b\\-2a+6b\end{bmatrix}+\begin{bmatrix}2c\\c\end{bmatrix}:\ a,b,c\in\mathbb{R}\right\} \\ &= \left\{k_1\begin{bmatrix}1\\-2\end{bmatrix}+k_2\begin{bmatrix}2\\1\end{bmatrix}:\ k_1,k_2\in\mathbb{R}\right\}, \end{aligned}$$

showing that $span\left\{\begin{bmatrix}1\\-2\end{bmatrix},\begin{bmatrix}-3\\6\end{bmatrix},\begin{bmatrix}2\\1\end{bmatrix}\right\}=span\left\{\begin{bmatrix}1\\-2\end{bmatrix},\begin{bmatrix}2\\1\end{bmatrix}\right\}$. (We can see, for example, that $\begin{bmatrix}-3\\6\end{bmatrix}=-3\begin{bmatrix}1\\-2\end{bmatrix}$, so $\begin{bmatrix}-3\\6\end{bmatrix}$ does not contribute to the spanning capability of the original set of vectors.) The vectors $\begin{bmatrix}1\\-2\end{bmatrix}$ and $\begin{bmatrix}2\\1\end{bmatrix}$ do not lie on the same straight line, and we claim that span $\left\{\begin{bmatrix}1\\-2\end{bmatrix},\begin{bmatrix}2\\1\end{bmatrix}\right\}=\mathbb{R}^2$.

To see this, suppose that $\begin{bmatrix}x\\y\end{bmatrix}$ is any vector in $\mathbb{R}^2$. We show that there are real numbers $k_1$ and $k_2$ such that $\begin{bmatrix}x\\y\end{bmatrix}=k_1\begin{bmatrix}1\\-2\end{bmatrix}+k_2\begin{bmatrix}2\\1\end{bmatrix}=\begin{bmatrix}k_1+2k_2\\-2k_1+k_2\end{bmatrix}$. This vector equation is equivalent to the system of two equations in two unknowns, $k_1+2k_2=x,\ -2k_1+k_2=y$; we can solve this system by substitution or elimination to conclude that $k_1=(x-2y)/5$ and $k_2=(2x+y)/5$. (For instance, if $\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}2\\-7\end{bmatrix}$, then $k_1=[2-2(-7)]/5=16/5$, $k_2=[2(2)+(-7)]/5=-3/5$, and $\begin{bmatrix}2\\-7\end{bmatrix}=\frac{16}{5}\begin{bmatrix}1\\-2\end{bmatrix}-\frac{3}{5}\begin{bmatrix}2\\1\end{bmatrix}$.)

## Example 1.3.2: The Span of a Set of Vectors in $\mathbb{R}^3$

Consider the set $S=\left\{\begin{bmatrix}1\\-3\\2\end{bmatrix},\begin{bmatrix}2\\-4\\-1\end{bmatrix},\begin{bmatrix}1\\-5\\7\end{bmatrix}\right\}$. Then

$$\begin{aligned} span(S) &= \left\{ a\begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix} + b\begin{bmatrix} 2 \\ -4 \\ -1 \end{bmatrix} + c\begin{bmatrix} 1 \\ -5 \\ 7 \end{bmatrix} :\ a, b, c \in \mathbb{R} \right\} \\ &= \left\{ \begin{bmatrix} a+2b+c \\ -3a-4b-5c \\ 2a-b+7c \end{bmatrix} :\ a, b, c \in \mathbb{R} \right\} \\ &= \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3 :\ 11x+5y+2z=0 \right\}. \end{aligned}$$

Although not obvious, this last description of the span is easily verified, and it indicates that the span is a plane through the origin.

An examination of the original vectors indicates that they are not collinear. However, there is a relationship among them, for example, $\begin{bmatrix} 1 \\ -5 \\ 7 \end{bmatrix} = 3\begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix} - \begin{bmatrix} 2 \\ -4 \\ -1 \end{bmatrix}$. Finally, we can conclude that the span is not all of $\mathbb{R}^3$ because (for example) the components of the vector, $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, do not satisfy the equation $11x+5y+2z=0$ and so this vector cannot be in the span. (Example 1.3.4 provides a more detailed explanation of nonspanning.)

## Example 1.3.3: A Spanning Set for $\mathbb{R}^3$

We show that the set $S = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ spans $\mathbb{R}^3$ by demonstrating that for any vector $\mathbf{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ in $\mathbb{R}^3$, we can find scalars $a$, $b$, and $c$, such that $\mathbf{v}$ can be written as

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = a\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + b\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + c\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \text{or} \quad \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} a \\ a+b \\ a+b+c \end{bmatrix}.$$

This last vector equation is equivalent to the linear system

$$\begin{aligned} a &= x \\ a+b &= y \\ a+b+c &= z \end{aligned}$$

with the solutions $a = x$, $b = y - a = y - x$ and $c = z - a - b = z - x - (y - x) = z - y$.

Therefore,

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = x\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + (y-x)\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + (z-y)\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

and we have shown that $S$ is a spanning set for $\mathbb{R}^3$. (For instance, if $\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$, then $\begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} = (1)\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + (-3)\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + 5\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.)

## Example 1.3.4: A Nonspanning Set for $\mathbb{R}^3$

Let us examine the vectors $\mathbf{u} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ to see if they span $\mathbb{R}^3$. We will take an arbitrary vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ in $\mathbb{R}^3$ and try to find scalars $a$, $b$, and $c$ such that $\mathbf{x}$ can be written as $a\mathbf{u} + b\mathbf{v} + c\mathbf{w}$. This problem is equivalent to solving the system

(1) $a + 2b = a + 2b = x_1$.
(2) $a + 5b + c = a + 5b + c = x_2$.
(3) $3b + c = 3b + c = x_3$.

Subtracting equation (3) from equation (2) yields $a + 2\,b = x_2 - x_3$. Comparing this result to the original equation (1), we are forced to conclude that $x_1 = x_2 - x_3$. In other words, the only way we can find the scalars $a$, $b$, and $c$ is if the vector $\mathbf{x}$ has the form $\begin{bmatrix} x_2 - x_3 \\ x_2 \\ x_3 \end{bmatrix}$. This means that any vector *not* of this form will not be ''reached'' by the vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$. For example, the vector $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ cannot be written as a linear combination of these three vectors,* although the vector $\begin{bmatrix} -1 \\ 2 \\ 3 \end{bmatrix}$ *can* be expressed this way. Thus $A = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$ is not a

* Just finding such a counterexample constitutes a proof that the set does not span $\mathbb{R}^3$.

spanning set for $\mathbb{R}^3$. We note that there is some redundancy in set $A$, for example, $\begin{bmatrix} 2 \\ 5 \\ 3 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + 3\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$.

## Exercises 1.3

**A.**

1. Give a geometric interpretation of the span of $\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}\right\}$ in $\mathbb{R}^3$.

2. What is the span of the set $\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}\right\}$ in $\mathbb{R}^3$?

3. Let $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 2 \\ t \\ 2t \end{bmatrix}$. Find all values of $t$ (if any) for which $\mathbf{u}$ and $\mathbf{v}$ span $\mathbb{R}^3$.

4. Let $S = \left\{\begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}\right\}$. Describe $span(S)$. Is the vector $\begin{bmatrix} -5 \\ 2 \\ 3 \end{bmatrix}$ in $span(S)$?

5. Show that $span\left\{\begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix}, \begin{bmatrix} -3 \\ 4 \\ 1 \end{bmatrix}\right\} = span\left\{\begin{bmatrix} -1 \\ 3 \\ 7 \end{bmatrix}, \begin{bmatrix} 8 \\ -9 \\ 4 \end{bmatrix}\right\}$.

6. Find a vector $\mathbf{u} \in \mathbb{R}^3$ *not* in the span of $S = \left\{\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}\right\}$.

7. Let $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 4 \\ 3 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 2 \\ 7 \\ 6 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$, and $\mathbf{v}_2 = \begin{bmatrix} -2 \\ 2 \\ 3 \end{bmatrix}$. Show that $span\{\mathbf{u}_1, \mathbf{u}_2\} \subseteq span\{\mathbf{v}_1, \mathbf{v}_2\}$.

8. In $\mathbb{R}^3$, determine the graph of

a. $span\left\{\begin{bmatrix}1\\3\\2\end{bmatrix},\begin{bmatrix}2\\6\\4\end{bmatrix},\begin{bmatrix}-3\\-9\\-6\end{bmatrix}\right\}$ b. $span\left\{\begin{bmatrix}2\\-4\\6\end{bmatrix},\begin{bmatrix}-6\\12\\-18\end{bmatrix},\begin{bmatrix}1\\-2\\3\end{bmatrix}\right\}$.

9. Find the span of
$S=\left\{[1\quad 4\quad 0\quad 0]^{\mathrm{T}},\,[2\quad 0\quad 4\quad 0]^{\mathrm{T}},\,[0\quad 0\quad 4\quad 1]^{\mathrm{T}}\right\}$.

10. Is the vector $\begin{bmatrix}3\\-1\\0\\-1\end{bmatrix}$ in the span of the vectors $\begin{bmatrix}2\\-1\\3\\2\end{bmatrix}$, $\begin{bmatrix}-1\\1\\1\\-3\end{bmatrix}$, and $\begin{bmatrix}1\\1\\9\\5\end{bmatrix}$ in $\mathbb{R}^4$?

**B.**

1. Prove that if $S\subseteq T\subseteq\mathbb{R}^n$, then $span(S)\subseteq span(T)$.

2. Prove that if S and T are nonempty finite sets such that $S$ is a set of vectors in $\mathbb{R}^n$, then $span(span(S))=span(S)$.

3. If $U$ and $V$ are nonempty finite subsets of $\mathbb{R}^n$, show that $span(U)+span(V)\subseteq span(U\cup V)$. [If $A$ and $B$ are sets, then $A+B=\{x|x=a+b$, where $a\in A$ and $b\in B\}$.]

4. For any two nonempty finite subsets $U$ and $V$ of $\mathbb{R}^n$, show that
a. $span(U)\cup span(V)\subseteq span(U\cup V)$.
b. $span(U\cap V)\subseteq span(U)\cap span(V)$.

5. Let $S=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_k\}$ and $T=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_k,\mathbf{w}\}$ be two sets of vectors from $\mathbb{R}^n$ where $\mathbf{w}\neq\mathbf{0}$. Prove that $span(S)=span(T)$ if and only if $\mathbf{w}\in span(S)$.

6. Prove that if $\mathbf{u}$ is orthogonal to every vector in the set $S=\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_k\}$, then $\mathbf{u}$ is orthogonal to every vector in $span(S)$. (This is a generalization of Problem B14, Exercises 1.2.)

**C.**

1. Suppose that $\mathbf{v}\in span\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_k\}$ and each $\mathbf{v}_i\in span\{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_r\}$.
a. Show that $\mathbf{v}\in span\{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_r\}$.
b. State conditions for $span\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_k\}=span\{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_r\}$.

## 1.4 Linear Independence

Let us look at the set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ in $\mathbb{R}^3$, where $\mathbf{v}_1 = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -8 \\ 4 \\ 9 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} 2 \\ 0 \\ -3 \end{bmatrix}$. If we analyze the relationships among the vectors in $S$ (perhaps for a long time), we might discover that $\mathbf{v}_2$ is a linear combination of vectors $\mathbf{v}_1$ and $\mathbf{v}_3$: $\begin{bmatrix} -8 \\ 4 \\ 9 \end{bmatrix} = 2\begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} - 3\begin{bmatrix} 2 \\ 0 \\ -3 \end{bmatrix}$, or $\mathbf{v}_2 = 2\mathbf{v}_1 - 3\mathbf{v}_3$. This algebraic relationship can also be written as $-2\mathbf{v}_1 + \mathbf{v}_2 + 3\mathbf{v}_3 = \mathbf{0}$ (the zero vector). (See Examples 1.3.1(a), 1.3.2, and 1.3.4 for other instances of this kind of redundancy.)

We say that this set of vectors $S$ is *linearly dependent* because one vector can be expressed as a linear combination of other vectors in the set. Let us define this concept precisely in $\mathbb{R}^n$.

**Definition 1.4.1**

A nonempty finite set of vectors $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ in $\mathbb{R}^n$ is called **linearly independent** if the only way that $a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k = \mathbf{0}$, where $a_1, a_2, \ldots, a_k$ are scalars, is if $a_1 = a_2 = \cdots = a_k = 0$. Otherwise, $S$ is called **linearly dependent**.

[**Note that any set containing the zero vector cannot be linearly independent** (Exercise B12).]

If we speak somewhat loosely and say that vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are linearly independent (or dependent), we mean that the set $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is linearly independent (or dependent).

### Example 1.4.1: A Linearly Independent Set in $\mathbb{R}^3$

We consider the set $S = \left\{ \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix} \right\}$ in $\mathbb{R}^3$.

To determine the linear independence or dependence of $S$, we have to investigate the solutions $a_1$, $a_2$, $a_3$ to the equation $a_1\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} + a_2\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + a_3\begin{bmatrix} 1 \\ 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, or $\begin{bmatrix} a_1 - a_2 + a_3 \\ 2a_1 + a_2 + 3a_3 \\ -a_1 - a_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

This vector equation is equivalent to the system of equations

$$\begin{aligned} a_1 - a_2 + a_3 &= 0 \\ 2a_1 + a_2 + 3a_3 &= 0 \\ -a - a_3 &= 0 \end{aligned}$$

Solving this system, we can add the first and third equations to conclude that $a_2 = 0$. Continuing in this way, we find that $a_1$ and $a_3$ must be zero as well. Therefore, the set of vectors $S$ is *linearly independent*.

---

As we saw in the last example, we have to be able to solve a system of linear equations to determine if a set of vectors is linearly independent. Historically, the solution of systems of linear equations was a major motivating factor in the development of linear algebra,* and we will focus on this topic in the next chapter. For now, we continue to learn more about vectors.

We can characterize linear dependence (and hence independence) in an alternative way.

***Theorem 1.4.1***

A set $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ in $\mathbb{R}^n$ that contains at least two vectors is linearly dependent if and only if some vector $\mathbf{v}_j$ (with $j > 1$) is a linear combination of the remaining vectors in the set.

*Proof* First, suppose that some vector $\mathbf{v}_j$ is a linear combination of the remaining vectors in the set:

$$\mathbf{v}_j = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_{j-1}\mathbf{v}_{j-1} + a_{j+1}\mathbf{v}_{j+1} + \cdots + a_k\mathbf{v}_k.$$

Then

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_{j-1}\mathbf{v}_{j-1} + (-1)\mathbf{v}_j + a_{j+1}\mathbf{v}_{j+1} + \cdots + a_k\mathbf{v}_k = \mathbf{0}.$$

Because the coefficient of $\mathbf{v}_j$ in this linear combination is not zero, the set $S$ is linearly dependent.

On the other hand, suppose that we know $S$ is linearly dependent. Then we can find scalars $a_1, a_2, \ldots, a_k$ such that $\sum_{i=1}^{k} a_i\mathbf{v}_i = \mathbf{0}$ with at least one scalar, say $a_j$, not zero. This implies that

$$a_j\mathbf{v}_j = -a_1\mathbf{v}_1 - a_2\mathbf{v}_2 - \cdots - a_{j-1}\mathbf{v}_{j-1} - a_{j+1}\mathbf{v}_{j+1} - \cdots - a_k\mathbf{v}_k,$$

so that $\mathbf{v}_j = \left(\frac{-a_1}{a_j}\right)\mathbf{v}_1 + \left(\frac{-a_2}{a_j}\right)\mathbf{v}_2 + \cdots + \left(\frac{-a_{j-1}}{a_j}\right)\mathbf{v}_{j-1} + \left(\frac{-a_{j+1}}{a_j}\right)\mathbf{v}_{j+1} + \cdots + \left(\frac{-a_k}{a_j}\right)\mathbf{v}_k$. Thus $\mathbf{v}_j$ can be written as a linear combination of the remaining vectors in $S$.

---

* See, for example, V. Katz, Historical ideas in teaching linear algebra, in *Learn from the Masters!*, F. Swetz et al., eds. (Washington, DC: MAA, 1995).

## Example 1.4.2: Linearly Dependent Sets of Vectors in $\mathbb{R}^4$ and $\mathbb{R}^5$

a. The vectors $[1\ \ -1\ \ 1\ \ 2\ \ 1]^T$, $[4\ \ -1\ \ 6\ \ 6\ \ 2]^T$, $[-4\ \ -2\ \ -3\ \ -4\ \ -2]^T$, and $[-2\ \ -1\ \ 1\ \ -2\ \ -2]^T$ are linearly dependent in $\mathbb{R}^5$ because (for example) $[-4\ \ -2\ \ -3\ \ -4\ \ -2]^T = 2[1\ \ -1\ \ 1\ \ 2\ \ 1]^T - [4\ \ -1\ \ 6\ \ 6\ \ 2]^T + [-2\ \ -1\ \ 1\ \ -2\ \ -2]^T$.

b. The vectors $[1\ \ 1\ \ 0\ \ 0]^T$, $[1\ \ 0\ \ 1\ \ 0]^T$, $[1\ \ 0\ \ 0\ \ 1]^T$, $[0\ \ 1\ \ 1\ \ 0]^T$, and $[0\ \ 0\ \ 1\ \ 1]^T$ are linearly dependent in $\mathbb{R}^4$ because $[1\ \ 1\ \ 0\ \ 0]^T = [1\ \ 0\ \ 0\ \ 1]^T + [0\ \ 1\ \ 1\ \ 0]^T - [0\ \ 0\ \ 1\ \ 1]^T$, or equivalently, $(1)[1\ \ 1\ \ 0\ \ 0]^T + (0)[1\ \ 0\ \ 1\ \ 0]^T + (-1)[1\ \ 0\ \ 0\ \ 1]^T + (-1)[0\ \ 1\ \ 1\ \ 0]^T + (1)[0\ \ 0\ \ 1\ \ 1]^T = [0\ \ 0\ \ 0\ \ 0]^T$.

Geometrically, in $\mathbb{R}^2$ or $\mathbb{R}^3$, a set of vectors is linearly independent if and only if no two vectors (viewed as emanating from the origin) lie on the same straight line (Figure 1.12a through c).

At this point, we investigate the connection between linearly independent sets and spanning sets. For any space $\mathbb{R}^n$ the possible relationships are indicated by Figure 1.13. The letters refer to the components of

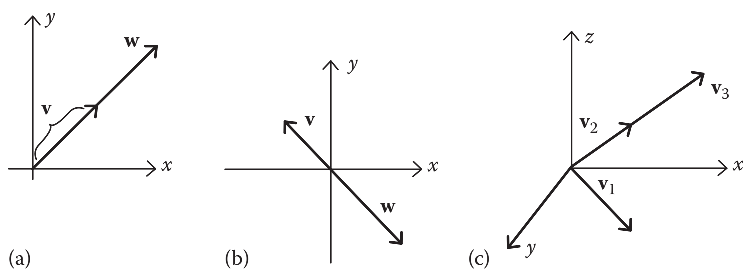



**Figure 1.12** Linearly dependent vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$.

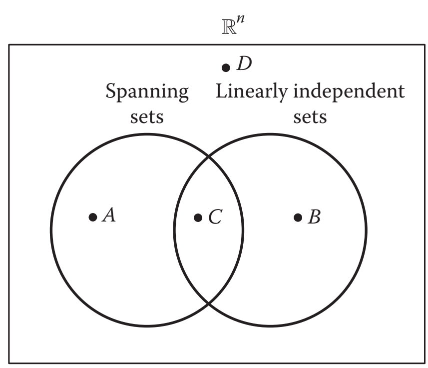



**Figure 1.13** Spanning sets and linearly independent sets in $\mathbb{R}^n$.

Example 1.4.3, which provides specifics in $\mathbb{R}^3$. The verification of these statements is required in Exercises A2 through A5.

## Example 1.4.3: Spanning Sets and Linearly Independent Sets in $\mathbb{R}^3$

a. The set $A = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$ spans $\mathbb{R}^3$ but is not linearly independent.

b. The set $B = \left\{ \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ -4 \\ -1 \end{bmatrix} \right\}$ is linearly independent but does not span $\mathbb{R}^3$.

c. The set $C = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ is linearly independent and also spans $\mathbb{R}^3$.

d. The set $D = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}$ is linearly dependent and does not span $\mathbb{R}^3$.

Despite the possibilities shown in Example 1.4.3, we *can* establish a relationship between linearly independent sets and spanning sets in any Euclidean space $\mathbb{R}^n$: **The size of any linearly independent set is always less than or equal to the size of any spanning set**. (In Example 1.4.3, notice that the linearly independent sets have two or three vectors, whereas any spanning set has three or four vectors.) The proof of this important result is a bit intricate, using what is sometimes called the **Steinitz*** **replacement** (or **exchange**) **technique**; but the next example should illuminate the basic idea. The name of the process comes from the fact that vectors of a spanning set are replaced by (exchanged for) vectors of a linearly independent set.

## Example 1.4.4: A Taste of Theorem 1.4.2

We know that $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, and $\mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ span $\mathbb{R}^3$.

Suppose we believe that $\mathbf{w}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $\mathbf{w}_2 = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$,

* This technique was named for Ernst Steinitz (1871–1928), who gave the first abstract definition of the algebraic structure known as a *field* and made other contributions to algebra.

$\mathbf{w}_3 = \begin{bmatrix} 5 \\ 6 \\ 7 \end{bmatrix}$, and $\mathbf{w}_3 = \begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}$ are linearly independent vectors in $\mathbb{R}^3$.

We can write $\mathbf{w}_1 = \mathbf{e}_1 + 2\mathbf{e}_2 + 3\mathbf{e}_3$. Then $\mathbf{e}_1$ can be expressed as a linear combination of $\mathbf{w}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$: $\mathbf{e}_1 = \mathbf{w}_1 + (-2)\mathbf{e}_2 + (-3)\mathbf{e}_3$. Thus the set $\{\mathbf{w}_1, \mathbf{e}_2, \mathbf{e}_3\}$ spans $\mathbb{R}^3$ because any linear combination of $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ is actually a combination of $\mathbf{w}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$. Note that we have replaced one of the spanning vectors by a vector from the alleged linearly independent set.

Now repeat this process by writing $\mathbf{w}_2 = 3\mathbf{w}_1 + (-2)\mathbf{e}_2 + (-4)\mathbf{e}_3$. Therefore, we have $\mathbf{e}_2 = \frac{3}{2}\mathbf{w}_1 + (-\frac{1}{2})\mathbf{w}_2 + (-2)\mathbf{e}_3$, so the set $\{\mathbf{w}_2 1, \mathbf{w}_2, \mathbf{e}_3\}$ spans $\mathbb{R}^3$. Proceeding one step further, we can write $\mathbf{w}_3$ in terms of $\mathbf{w}_1$, $\mathbf{w}_2$, and $\mathbf{e}_3$, and then express $\mathbf{e}_3$ as a linear combination of $\mathbf{w}_1$, $\mathbf{w}_2$, and $\mathbf{w}_3$. Thus we have shown that $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ spans $\mathbb{R}^3$. But then $\mathbf{w}_4$ must be a linear combination of $\mathbf{w}_1$, $\mathbf{w}_2$, and $\mathbf{w}_3$—a contradiction of the linear independence of $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4\}$ (by Theorem 1.4.1). [From this analysis, we might suspect that no set of four or more vectors in $\mathbb{R}^3$ can be linearly independent. In this example, note that $\mathbf{w}_4 = (-1)\mathbf{w}_2 + 2\mathbf{w}_3$.]

Now we should be ready to tackle the relationship between linearly independent sets and spanning sets in any Euclidean space.

***Theorem 1.4.2***

Suppose that the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ span $\mathbb{R}^n$ and that the nonzero vectors $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k$ in $\mathbb{R}^n$ are linearly independent. Then $k \le m$. (That is, the size of any linearly independent set is always less than or equal to the size of any spanning set.)

*Proof* Because $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ span $\mathbb{R}^n$, every vector in $\mathbb{R}^n$ can be written as a linear combination of the vectors $\mathbf{v}_i$. In particular,

$$\mathbf{w}_1 = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_m\mathbf{v}_m.$$

Because $\mathbf{w}_1 \neq \mathbf{0}$, not all the coefficients $a_i$ are equal to 0. After renumbering $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ if necessary, we can say that $a_1 \neq 0$. Then $\mathbf{v}_1$ can be expressed as a linear combination of $\mathbf{w}_1$ and the remaining vectors $\mathbf{v}_i$: $\mathbf{v}_1 = \left(\frac{1}{a_1}\right)\mathbf{w}_1 + \left(\frac{-a_2}{a_1}\right)\mathbf{v}_2 + \cdots + \left(\frac{-a_m}{a_1}\right)\mathbf{v}_m$. Therefore, the set $\{\mathbf{w}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$, consisting of the $\mathbf{v}_i$'s with $\mathbf{v}_1$ replaced by $\mathbf{w}_1$, spans $\mathbb{R}^n$. We continue this process, replacing the $\mathbf{v}$'s by $\mathbf{w}$'s and relabeling the vectors $\mathbf{v}_i$, if necessary. The key claim, which we will prove by mathematical induction, is that for every $i = 1, 2, \ldots, m-1$, the set $\{\mathbf{w}_1, \mathbf{w}_2, \ldots \mathbf{w}_i, \mathbf{v}_{i+1}, \ldots \mathbf{v}_m\}$ spans $\mathbb{R}^n$.

We have proved the claim for $i = 1$. Now, assuming that for a specific $i$, $\{\mathbf{w}_1, \mathbf{w}_2, \ldots \mathbf{w}_i, \mathbf{v}_{i+1}, \ldots \mathbf{v}_m\}$ spans $\mathbb{R}^n$, we can write

$$\mathbf{w}_{i+1} = (a_1\mathbf{w}_1 + a_2\mathbf{w}_2 + \cdots + a_i\mathbf{w}_i) + (a_{i+1}\mathbf{v}_{i+1} + a_{i+2}\mathbf{v}_{i+2} + \cdots + a_m\mathbf{v}_m).$$

Not all $a_i$'s in the second set of parentheses can be 0 because this would imply that $\mathbf{w}_{i+1}$ is a linear combination of other $\mathbf{w}_k$'s, so that the set of $\mathbf{w}_k$'s is linearly dependent. Say that $a_{i+1} \neq 0$ (if not, just rearrange or relabel the terms in the second set of parentheses). Then

$$\begin{aligned}\mathbf{v}_{i+1} = \frac{\mathbf{w}_{i+1}}{a_{i+1}} &- \left\{\left(\frac{a_1}{a_{i+1}}\right)\mathbf{w}_1 + \cdots + \left(\frac{a_i}{a_{i+1}}\right)\mathbf{w}_i\right\} \\ &- \left\{\left(\frac{a_{i+2}}{a_{i+1}}\right)\mathbf{v}_{i+2} + \cdots + \left(\frac{a_m}{a_{i+1}}\right)\mathbf{v}_m\right\},\end{aligned}$$

which implies that $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{i+1}, \mathbf{v}_{i+2}, \ldots, \mathbf{v}_m\}$ spans $\mathbb{R}^n$. Thus we have proved the claim by induction.

If $k > m$, eventually the $\mathbf{v}_i$'s disappear, having been replaced by the $\mathbf{w}_j$'s. Furthermore, span $\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\} = \mathbb{R}^n$. Then each of the vectors $\mathbf{w}_{m+1}, \mathbf{w}_{m+2}, \ldots, \mathbf{w}_k$ is a linear combination of $\mathbf{w}_1, \mathbf{w}_2, \ldots,$ $\mathbf{w}_m$—contradicting the assumed linear independence of the $\mathbf{w}_i$'s. Therefore, $k$ must be less than or equal to $m$.

We can use Theorem 1.4.2 to make a stronger statement about linearly independent/dependent sets of vectors.

***Theorem 1.4.3***

Suppose that $A = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is a set of vectors in $\mathbb{R}^n$. If $k > n$, that is, if the number of vectors in $A$ exceeds the dimension of the space, then $A$ is linearly dependent.

*Proof* We have observed that $\mathbb{R}^n$ has a spanning set consisting of the $n$ vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$, where $\mathbf{e}_i$ has 1 as component $i$ and zeros elsewhere. Theorem 1.4.2 implies that if $A$ is a set of $k$ linearly independent vectors in $\mathbb{R}^n$, then $k \leq n$. Therefore, any set of $k > n$ vectors must be dependent.

Another way of stating Theorem 1.4.3 is that **if $A = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is a linearly independent subset of $\mathbb{R}^n$, then $k \leq n$.** In the language of logic, this last statement is just the *contrapositive* of Theorem 1.4.3 and so is an equivalent statement. Note that this does *not* say that $k \leq n$ guarantees linear independence.

Now let $\mathbf{v}$ and $\mathbf{w}$ be two nonzero linearly independent vectors in $\mathbb{R}^2$. We want to prove our earlier claim (Section 1.3) that the span of these two vectors is all of $\mathbb{R}^2$. Suppose that there is a vector $\mathbf{z} \in \mathbb{R}^2$ such that $\mathbf{z} \notin \text{span}\{\mathbf{v}, \mathbf{w}\}$. Exercise B1 implies that if $\mathbf{z} \notin \text{span}(\{\mathbf{v}, \mathbf{w}\})$, then $\{\mathbf{v}, \mathbf{w}, \mathbf{z}\}$ must be linearly independent. But this contradicts Theorem 1.4.3

since we would have three vectors in a two-dimensional space. Therefore, $\mathbf{z} \in span\{\mathbf{v}, \mathbf{w}\} = \mathbb{R}^2$. In this same way, we can show that three linearly independent vectors in $\mathbb{R}^3$ must span $\mathbb{R}^3$.

In the next section, we investigate the consequences of having a set of vectors in $\mathbb{R}^n$ that are linearly independent and that span $\mathbb{R}^n$ at the same time.

## Exercises 1.4

### A.

1. Determine whether each of the following sets of vectors is linearly independent in the appropriate space $\mathbb{R}^n$.

a. $\begin{bmatrix} -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \end{bmatrix}$ b. $\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ c. $\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \end{bmatrix}$

d. $\begin{bmatrix} -3 \\ 1 \\ 5 \end{bmatrix}, \begin{bmatrix} -6 \\ -3 \\ 15 \end{bmatrix}$ e. $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

f. $\begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ -4 \\ 3 \end{bmatrix}$

g. $\begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ -4 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 4 \end{bmatrix}$ h. $\begin{bmatrix} 1 \\ 2 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 6 \\ 0 \end{bmatrix}$

i. $\begin{bmatrix} 1 \\ 2 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ j. $\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$

k. $\begin{bmatrix} -1 \\ 2 \\ -3 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$

l. $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$

2. Show that the set $A$ in Example 1.4.3 spans $\mathbb{R}^3$ but is not linearly independent.

3. Show that the set $B$ in Example 1.4.3 is linearly independent but does not span $\mathbb{R}^3$.

4. Show that the set $C$ in Example 1.4.3 is linearly independent and also spans $\mathbb{R}^3$.

5. Show that the set $D$ in Example 1.4.3 is linearly dependent and does not span $\mathbb{R}^3$.

6. a. Show that the vectors $\begin{bmatrix} a \\ b \end{bmatrix}$ and $\begin{bmatrix} c \\ d \end{bmatrix}$ in $\mathbb{R}^2$ are linearly dependent if and only if $ad - bc = 0$.

   b. For what values of $r$ are the vectors $\begin{bmatrix} r \\ 1 \end{bmatrix}$ and $\begin{bmatrix} r+2 \\ r \end{bmatrix}$ linearly independent?

7. Suppose that $U = \{\mathbf{u}_1, \mathbf{u}_2\}$ and $V = \{\mathbf{v}_1, \mathbf{v}_2\}$ are linearly independent subsets of $\mathbb{R}^3$. Give a geometrical description of the intersection $span(U) \cap span(V)$.

**8.** Express a general vector $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ in $\mathbb{R}^3$ as a linear combination of the vectors $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$, and $\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$.

9. Let $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 2 \\ t \\ 2t \end{bmatrix}$. Find all values of $t$ (if any), for which $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent.

10. Determine a *maximal* set of linearly independent vectors from the set $S = \{[1\ \ -1\ \ -4\ \ 0]^{\mathrm{T}}, [1\ \ 1\ \ 2\ \ 4]^{\mathrm{T}}, [2\ \ 1\ \ 1\ \ 6]^{\mathrm{T}}, [2\ \ -1\ \ -5\ \ 2]^{\mathrm{T}}\}$, that is, find a linearly independent subset of $S$ such that adding any other vector from $S$ to it renders the subset linearly dependent.

11. Show that the set $\mathcal{E}_n = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ is linearly independent in $\mathbb{R}^n$. (See Section 1.3 for the definition of the vectors $\mathbf{e}_i, i = 1, 2, \ldots, n$.)

## B.

1. Let $S$ be a linearly independent set of vectors in $\mathbb{R}^n$. Suppose that $\mathbf{v}$ is a vector in $\mathbb{R}^n$ that is not in the span of $S$. Prove that the set $S \cup \{\mathbf{v}\}$ is linearly independent.

2. If $\mathbf{u}_1$ and $\mathbf{u}_2$ are linearly independent in $\mathbb{R}^n$ and $\mathbf{w}_1 = a\mathbf{u}_1 + b\mathbf{u}_2$, $\mathbf{w}_2 = c\mathbf{u}_1 + d\mathbf{u}_2$, show that $\mathbf{w}_1$ and $\mathbf{w}_2$ are linearly independent if and only if $ad \neq bc$.

3. Prove that if $\mathbf{v}_1$,$\mathbf{v}_2$, and $\mathbf{v}_3$ are linearly independent in $\mathbb{R}^n$, then so are the vectors $\mathbf{x} = \mathbf{v}_1 + \mathbf{v}_2, \mathbf{y} = \mathbf{v}_1 + \mathbf{v}_3$, and $\mathbf{z} = \mathbf{v}_2 + \mathbf{v}_3$.

4. Suppose that $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a linearly independent set of vectors in $\mathbb{R}^n$. Are the vectors $\mathbf{v}_1 - \mathbf{v}_2, \mathbf{v}_2 - \mathbf{v}_3$, and $\mathbf{v}_3 - \mathbf{v}_1$ linearly independent? Explain.

5. Suppose that $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a linearly independent set of vectors in $\mathbb{R}^n$. Show that $T = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$ is also linearly independent, where $\mathbf{w}_1 = \mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3$, $\mathbf{w}_2 = \mathbf{v}_2 + \mathbf{v}_3$, and $\mathbf{w}_3 = \mathbf{v}_3$.

6. Suppose that $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a linearly dependent set of vectors in $\mathbb{R}^n$. Is $T = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$, where $\mathbf{w}_1 = \mathbf{v}_1$, $\mathbf{w}_2 = \mathbf{v}_1 + \mathbf{v}_2$, and $\mathbf{w}_3 = \mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3$, linearly dependent or linearly independent? Justify your answer.

7. For what values of the scalar $a$ are the vectors $\begin{bmatrix} a \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ a \\ 1 \end{bmatrix}$, and $\begin{bmatrix} 0 \\ 1 \\ a \end{bmatrix}$ linearly independent, and for what values of $a$ are they linearly dependent? Explain your reasoning.

8. Find an integer $k$ so that the following set is linearly dependent:

$$\left\{ \begin{bmatrix} k \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ k \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ k \end{bmatrix} \right\}.$$

9. Find three vectors in $\mathbb{R}^3$ that are linearly dependent, but are such that any two of them are linearly independent.

10. a. Let $S_1$ and $S_2$ be finite subsets of vectors in $\mathbb{R}^n$ such that $S_1 \subseteq S_2$. If $S_2$ is linearly dependent, show by examples in $\mathbb{R}^3$ that $S_1$ may be either linearly dependent or linearly independent.
    b. Let $S_1$ and $S_2$ be finite subsets of vectors in $\mathbb{R}^n$ such that $S_1 \subseteq S_2$. If $S_1$ is linearly independent, show by examples in $\mathbb{R}^3$ that $S_2$ may be either linearly dependent or linearly independent.

11. Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be a set of vectors in $\mathbb{R}^n$. Show that if one of these vectors is the zero vector, then $S$ is linearly dependent.

12. a. Prove that any subset of a linearly independent set of vectors is linearly independent.
    b. Prove that any set of vectors containing a linearly dependent subset is linearly dependent.

13. Prove that any linearly independent set $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ consisting of $n$ vectors in $\mathbb{R}^n$ must span $\mathbb{R}^n$. [*Hint*: Use the result of Problem B1 of Exercises 1.4.]

**C.**

1. If

$$\begin{aligned}\mathbf{v}_1 &= \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3\\ \mathbf{v}_2 &= \mathbf{u}_1 + \alpha\mathbf{u}_2\\ \mathbf{v}_3 &= \mathbf{u}_2 + \beta\mathbf{u}_3,\end{aligned}$$

where $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ are given linearly independent vectors, find the conditions that must be satisfied by $\alpha$ and $\beta$ to ensure that $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are linearly independent.

## 1.5 Bases

Figure 1.13 indicates that there need be no connection between the concepts of linear independence and spanning. However, a set that is both linearly independent and spanning provides valuable insights into the structure of a Euclidean space.

### Definition 1.5.1

A nonempty set $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ of vectors in $\mathbb{R}^n$ is called a **basis** for $\mathbb{R}^n$ if both the following conditions hold:

(1) The set $S$ is linearly independent.
(2) The set $S$ spans $\mathbb{R}^n$.

In $\mathbb{R}^n$, the vectors $\mathbf{e}_1 = \begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}, \mathbf{e}_2 = \begin{bmatrix}0\\1\\\vdots\\0\end{bmatrix}, \dots, \mathbf{e}_3 = \begin{bmatrix}0\\0\\\vdots\\1\end{bmatrix}$ constitute a special basis called the **standard basis** for $\mathbb{R}^n$. As we have noted before, these vectors span $\mathbb{R}^n$. The proof that these vectors are linearly independent constituted Exercise A11 in Section 1.4. From now on, we will denote the standard basis in $\mathbb{R}^n$ by $\mathcal{E}_n$: $\mathcal{E}_n = \{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$. We note that the vectors in the standard basis for $\mathbb{R}^n$ are **mutually orthogonal**: $\mathbf{e}_i \bullet \mathbf{e}_j = 0$ if $i \neq j$. Also, $\|\mathbf{e}_i\| = 1$, $i = 1, 2, \dots, n$. A set of vectors that are mutually orthogonal and of unit length is called an **orthonormal set**. In particular, a basis consisting of vectors that are mutually orthogonal and of unit length is called an **orthonormal basis**. (See Exercises B5 through B7 for some properties of orthonormal bases.) We will discuss orthonormal sets and orthonormal bases more thoroughly in Chapters 7 and 8.

## Example 1.5.1: A Basis for $\mathbb{R}^3$

We show that the set $B = \left\{ \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 7 \\ 7 \end{bmatrix}, \begin{bmatrix} 4 \\ -1 \\ 0 \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^3$.

First of all, $B$ spans $\mathbb{R}^3$. If $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ is any vector in $\mathbb{R}^3$, then the equation $a \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + b \begin{bmatrix} 1 \\ 7 \\ 7 \end{bmatrix} + c \begin{bmatrix} 4 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ is equivalent to the system of linear equations:

(1) $2a + b + 4c = x_1$.
(2) $a + 7b - c = x_2$.
(3) $a + 7b = x_3$.

Equation (2) − equation (3) yields $-c = x_2 - x_3$, or $c = x_3 - x_2$. Equation (1) − twice equation (3) gives us $-13b + 4c = x_1 - 2x_3$, or (replacing $c$ by $x_3 - x_2$) $b = -\frac{1}{13}(x_1 - 6x_3 + 4x_2)$. Finally, equation (3) implies that $a = x_3 - 7b = x_3 + \frac{7}{13}(x_1 - 6x_3 + 4x_2)$. Thus any vector in $\mathbb{R}^3$ can be expressed as a linear combination of vectors in $B$.

From the formulas for $a$, $b$, and $c$ we have just derived, we see that if $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, then $a = b = c = 0$. This says that $B$ is a linearly independent set of vectors. As a linearly independent spanning set of vectors in $\mathbb{R}^3$, $B$ is a basis for $\mathbb{R}^3$.

The next example shows that $\mathbb{R}^2$ has an orthonormal basis other than the standard basis. The validity and generalization of the constructive process demonstrated in this example is shown in Section 7.4. The crucial concept is that of an *orthogonal projection* (see the explanation given between Problems C6 and C7 of Exercises 1.2).

## Example 1.5.2: Constructing an Orthonormal Basis for $\mathbb{R}^2$

It is easy to show that the set $B = \{\mathbf{v}_1, \mathbf{v}_2\} = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -3 \\ 4 \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^2$. However, $B$ is not an orthonormal basis: The vectors in $B$ are not mutually orthogonal and neither vector in $B$ has unit length. However, we illustrate a process (algorithm) that transforms the vectors of $B$ into an orthonormal basis for $\mathbb{R}^2$.

Let $\mathbf{w}_1 = \mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and calculate the vector

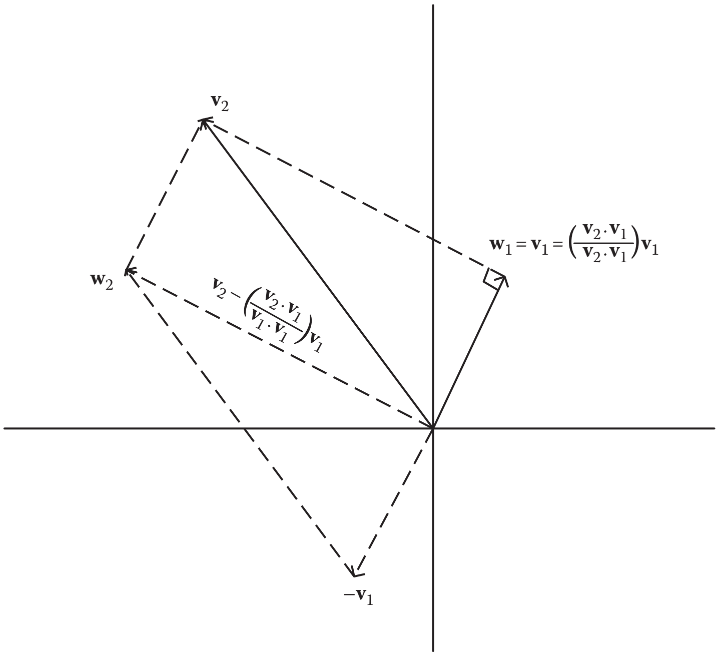



**Figure 1.14** Constructing an orthonormal basis.

$$\mathbf{w}_2 = \mathbf{v}_2 - \left(\frac{\mathbf{v}_2 \bullet \mathbf{v}_1}{\mathbf{v}_1 \bullet \mathbf{v}_1}\right)\mathbf{v}_1 = \begin{bmatrix} -3 \\ 4 \end{bmatrix} - \left(\frac{\begin{bmatrix} -3 \\ 4 \end{bmatrix} \bullet \begin{bmatrix} 1 \\ 2 \end{bmatrix}}{\begin{bmatrix} 1 \\ 2 \end{bmatrix} \bullet \begin{bmatrix} 1 \\ 2 \end{bmatrix}}\right)\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$
$$= \begin{bmatrix} -3 \\ 4 \end{bmatrix} - \frac{5}{5}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} -4 \\ 2 \end{bmatrix}.$$

The vector $\left(\frac{\mathbf{v}_2 \bullet \mathbf{v}_1}{\mathbf{v}_1 \bullet \mathbf{v}_1}\right)\mathbf{v}_1$ is the *orthogonal projection* of $\mathbf{v}_2$ on $\mathbf{v}_1$, and $\mathbf{w}_2 = \mathbf{v}_2 - \left(\frac{\mathbf{v}_2 \bullet \mathbf{v}_1}{\mathbf{v}_1 \bullet \mathbf{v}_1}\right)\mathbf{v}_1$ is a vector orthogonal to $\mathbf{v}_1 = \mathbf{w}_1$ (Figure 1.14).

Algebraically, we have $\mathbf{w}_1 \bullet \mathbf{w}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \bullet \begin{bmatrix} -4 \\ 2 \end{bmatrix} = -4 + 4 = 0.$

We can verify that $\{\mathbf{w}_1, \mathbf{w}_2\} = \left\{\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -4 \\ 2 \end{bmatrix}\right\}$ is a basis for $\mathbb{R}^2$. All that is left is to "normalize" these vectors by dividing each by its length:

$$\hat{\mathbf{w}}_1 = \frac{\mathbf{w}_1}{\|\mathbf{w}_1\|} = \frac{\begin{bmatrix} 1 \\ 2 \end{bmatrix}}{\sqrt{5}} = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix},$$
$$\hat{\mathbf{w}}_2 = \frac{\mathbf{w}_2}{\|\mathbf{w}_2\|} = \begin{bmatrix} \frac{\begin{matrix} -4 \\ 2 \end{matrix}}{\sqrt{20}} \end{bmatrix} = \begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}.$$

The result is that $\{\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2\} = \left\{ \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}, \begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix} \right\}$ is an orthonormal basis for $\mathbb{R}^2$.

---

Examples 1.5.1 and 1.5.2 imply that, for any positive integer value of $n$, the space $\mathbb{R}^n$ may have many bases (the plural of *basis*). In fact, any space $\mathbb{R}^n$ has *infinitely many* bases: It is easy to verify that if the vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ constitute a basis, then the vectors $\alpha\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$, where $\alpha$ is any scalar, also form a basis. However, the next theorem reveals that *any basis for a given* $\mathbb{R}^n$ *must have the same number of elements*. Theorem 1.4.2 implies that this number is less than or equal to $n$. An important consequence of the next theorem is that **the number of vectors in any basis for $\mathbb{R}^n$ must be exactly $n$**.

***Theorem 1.5.1***

Any two bases for $\mathbb{R}^n$ must have the same number of vectors.

*Proof* Suppose $B = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_k\}$ and $T = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r\}$ are two bases for $\mathbb{R}^n$, where $n$ is a given positive integer. We will show that $k = r$.

As bases, both $B$ and $T$ are linearly independent and span $\mathbb{R}^n$. First let us focus on the fact that $B$ is a spanning set for $\mathbb{R}^n$ and that the vectors $\mathbf{v}_i$ in $T$ are linearly independent. Then we can apply Theorem 1.4.2 to conclude that $r \le k$. Reversing the roles of $B$ and $T$—so that $T$ is viewed as spanning $\mathbb{R}^n$ and $B$ is linearly independent—we use Theorem 1.4.2 again to see that $k \le r$. The only way we can have both $r \le k$ and $k \le r$ is if $k = r$.

***Corollary 1.5.1***

Any basis for $\mathbb{R}^n$ must have $n$ elements.

*Proof* The standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ for $\mathbb{R}^n$ contains $n$ elements, and therefore, by Theorem 1.5.1, *every* basis for $\mathbb{R}^n$ contains $n$ vectors.

Later, in Chapter 5, when we discuss the more general (and more abstract) notion of a *vector space V* (instead of just $\mathbb{R}^n$), we will define the *dimension* of such a space as the number of elements in a basis for $V$.

Although any basis for $\mathbb{R}^n$ must contain $n$ elements, not every set of $n$ vectors in $\mathbb{R}^n$ is a basis—see Example 1.4.3(d), for instance. However, as we will see shortly, **every set of *n linearly independent* vectors in $\mathbb{R}^n$ must span $\mathbb{R}^n$ and so must be a basis.**

Theorem 1.4.2 states that the number of vectors in any linearly independent subset of $\mathbb{R}^n$ is less than or equal to the number of vectors in any spanning set for $\mathbb{R}^n$. The next result indicates how we can fill in this "gap" between linearly independent sets and spanning sets.

***Theorem 1.5.2***

If $S$ is a finite spanning set for some Euclidean space $\mathbb{R}^n$ and if $I$ is a linearly independent subset of $\mathbb{R}^n$, such that $I \subseteq S$, then there is a basis $B$ of $\mathbb{R}^n$, such that $I \subseteq B \subseteq S$.

*Proof* Suppose that $I$, a subset of a spanning set $S$, is a linearly independent set. We note first that if $span(I) = \mathbb{R}^n$, then $I$ must be a basis by definition. Now suppose that $span(I) \neq \mathbb{R}^n$. This implies that $span(I) \subset \mathbb{R}^n$ and that $I \subset S$, or else $I$ would be a spanning set. Then there exists at least one element $\mathbf{s}_1 \in S \backslash I$ such that $\mathbf{s}_1 \notin span(I)$: If this last condition were not true, then every element of $S \backslash I$ would belong to $span(I)$, so that $\mathbb{R}^n = span(S) \subseteq span(I)$ and $\mathbb{R}^n = span(I)$, a contradiction.

Now we can see that $I \cup \{\mathbf{s}_1\}$ is a linearly independent set. Otherwise, $\mathbf{s}_1 \in \mathrm{span}(I)$, another contradiction (see Exercises 1.4, Problem B1). If $I \cup \{\mathbf{s}_1\}$ spans $\mathbb{R}^n$, then it is a basis and we can let $B = I \cup \{\mathbf{s}_1\}$. If $I \cup \{\mathbf{s}_1\}$ does not span $\mathbb{R}^n$, we can repeat our argument to find an element $\mathbf{s}_2 \in S \backslash (I \cup \{\mathbf{s}_1\})$ with $I \cup \{\mathbf{s}_1, \mathbf{s}_2\}$ linearly independent. Because $S$ is a finite set by hypothesis, if we continue this process, we see that for some value $m$ the set, $B = I \cup \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_m\}$, is a basis of $\mathbb{R}^n$ with $I \subset B \subseteq S$.

The following consequence of Theorem 1.5.2 is very useful when working with linearly independent vectors, and will be invoked often in later sections. The corollary asserts that we can add enough vectors to any linearly independent subset to produce a basis, that is, an expanded version of the linearly independent set that also spans the space.

***Corollary 1.5.2***

Every linearly independent subset $I$, of a Euclidean space $\mathbb{R}^n$, can be extended to form a basis.

*Proof* Suppose that $I$ is a linearly independent subset of $\mathbb{R}^n$ that is not a basis for $\mathbb{R}^n$, and take $S = I \cup B$, where $B$ is any basis for $\mathbb{R}^n$. Then $S$ is a spanning set and, by Theorem 1.5.2, there is a basis $\hat{B}$ such that $I \subseteq \hat{B} \subseteq I \cup B$. This basis $\hat{B}$ is an extension of $I$.

Of course, the extension method described in Corollary 1.5.2 is not unique. There are other ways to form a basis by extending a linearly independent set.

The *Steinitz replacement technique*, used in the proof of Theorem 1.4.2, actually extended a linearly independent set of vectors to a basis, but we chose not to emphasize this fact at that time.

***Corollary 1.5.3***

Any set of $n$ linearly independent vectors in $\mathbb{R}^n$ is a basis of $\mathbb{R}^n$.

*Proof* Suppose that $I$ is a set of $n$ linearly independent vectors. If $I$ is not a basis, then it can be extended to form a basis for $\mathbb{R}^n$. But then the number of elements in the extended set is greater than or equal to $n+1$, and the set must be linearly dependent by Theorem 1.4.3, a contradiction. Therefore, $I$ must already be a basis.

## Example 1.5.3: Extending a Linearly Independent Set to a Basis

Suppose that we are given the set $S = \left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix} \right\}$ in $\mathbb{R}^3$. It is easy to check that $S$ is linearly independent. However, by Corollary 1.5.1, $S$ cannot be a basis for $\mathbb{R}^3$. Because any three linearly independent vectors in $\mathbb{R}^3$ constitute a basis for $\mathbb{R}^3$ (Corollary 1.5.3), $S$ needs one more appropriate vector.

First we calculate $span(S) = \left\{ \begin{bmatrix} a+2b \\ -a \\ a+3b \end{bmatrix} : a, b \in \mathbb{R} \right\}$. Now we find any vector that does not belong to the span of $S$. For example, $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ does not fit the pattern of the vectors in $span(S)$. (If $a=0$, the third component would have to be 3/2 of the first component.) Then $\hat{S} = \left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$ is a linearly independent set by Problem B1 of Exercises 1.4. (Alternatively, we can consider a linear combination of the three vectors that equals the zero vector and see that all the scalar coefficients must be zero.) Thus $\hat{S}$, as a linearly independent set of three vectors in $\mathbb{R}^3$, must be a basis.

Any vector in $\mathbb{R}^n$ can be represented as a linear combination of basis vectors, and it is important to realize that any such representation is *unique*. The next result has far-reaching consequences; the converse of this theorem is also true (Exercise B1).

***Theorem 1.5.3***

If $B = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for $\mathbb{R}^n$, then every vector $\mathbf{v}$ in $\mathbb{R}^n$ can be written as a linear combination $\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n$ in one and only one way.

*Proof* We give a proof by contradiction. First of all, because $B$ is a basis, any vector $\mathbf{v}$ in $\mathbb{R}^n$ can be written as a linear combination of vectors in $B$:

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n.$$

Now suppose that $\mathbf{v}$ has another representation in terms of the vectors in $B$:

$$\mathbf{v} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_n\mathbf{v}_n.$$

Then

$$\begin{aligned} 0 = \mathbf{v} - \mathbf{v} &= a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n - (b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_n\mathbf{v}_n) \\ &= (a_1 - b_1)\mathbf{v}_1 + (a_2 - b_2)\mathbf{v}_2 + \cdots + (a_n - b_n)\mathbf{v}_n. \end{aligned}$$

But the linear independence of the vectors $\mathbf{v}_i$ implies that $(a_1 - b_1) = (a_2 - b_2) = \cdots = (a_n - b_n) = 0$, that is, $a_1 = b_1, a_2 = b_2, \ldots, a_n = b_n$, so $\mathbf{v}$ has a unique representation as a linear combination of basis vectors.

We should notice that the proofs of Theorems 1.4.1 through 1.4.3 and 1.5.1 through 1.5.3 do not make explicit use of the fact that the vectors are $n$-tuples. These demonstrations use only definitions and the algebraic properties of vectors contained in Theorem 1.1.1. We shall see this abstract algebraic approach in its full glory beginning in Chapter 5.

## Example 1.5.4: RGB Color Space

In terms of the perception of light by the human eye, experiments have shown that every color can be duplicated by using mixtures of the three primary colors, red, green, and blue.* There is, however, a difference between mixtures of *light* (as used on a computer screen) and mixtures of *pigment* an artist uses.†

In studies of color matching, involving the projection of light onto a screen, the standard basis vectors, $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, and $\mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, represent the projection of one foot-candle of **red**, **green**, and **blue**, respectively, onto a screen. The zero vector represents the color **black**, whereas the vector $[\,1 \quad 1 \quad 1\,]^{\mathrm{T}}$ represents **white**. The scalar multiple $c\mathbf{e}_i$ represents the projection of $c$ foot-candles of the color represented by $\mathbf{e}_i$, that is, the *intensity* or *brightness* of the light is changed

---

* The theory was presented by Thomas Young in 1802 and developed further by Hermann von Helmholtz in the 1850s. The theory was finally proved in a 1983 experiment by Dartnall, Bowmaker, and Mollon.

† See, for example, R.G. Gonzalez and R.E. Woods, *Digital Image Processing*, 2nd edn., Chap. 6 (Upper Saddle River, NJ: Prentice-Hall, 2002).