.....

.....

Linear Algebra

الجبر الخطي

(لطلاب كلية الحاسبات)

(MTH1441)

(2026-2027)

الإسم : .....

الشعبة: .....

الخطة الدراسية لمقرر الجبر الخطي (طلاب كلية الحاسبات) : MTH1441

Weeks	Chapters	Sections Numbers	Topics
1	**Introduction to vectors** Strang, G. (2023). Introduction to Linear Algebra (6th ed.)	**1.1**	**Vectors and linear combinations** Examples: *1.1A, 1.1 B, 1.1 C.* Exercises: 1, 3, 5, 7, 9, *15, 17, 21, 23, and 25.*
2		**1.2**	***Lengths and dot products*** Examples: *1, 2, 3, 4, and 5.* Exercises: *1, 3, 5, 7, 11, 14, 19, and 20.*
3	**Chapter 1: Linear Equations in linear algebra** Lay, D. (2021). Linear Algebra and Its Applications, Global Edition (6th ed.)	**1.1**	**Systems of linear equations** Examples: *1, 2, and 3.* Exercises: *1, 3, 5, 7, 11, 13, 19, 20, and 22* + practice problems 1, 2, 4.
		1.2	**Row reduction and Echelon Forms** Examples: *1, 2, 3, and 4.* Exercises: *1, 3, 7, 9, 11, 19, 21, 23,* and *(25-to-34)* + practice problems 1, 2.
4		**1.3**	**Vector equations** Examples: *1, 2, 3, 4, 5, and 6.* Exercises: *2, 4, 6, 8, 10, 12, 14, 16, 18, 20,* and *(23-to-32).* + practice problem 2.
		1.4	**The matrix equation Ax = b** Examples: *1, 2, 3, 4, and 5.* Exercises: *2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22,* and *(23-to-34).*+ practice problems 1, 2.
5		**1.5**	**Solution sets of linear systems** Examples: *1, 2, and 3.* Exercises: *1, 3, 5, 7, 9, 11, 19, 21, 23,* and *(27-to-36)* + practice problem 1, 2.
		1.7	**Linear independence** Examples: *1, 2, and 3.* Exercises: *2, 4, 6, 8, 10, 12, 14, 16, 18, 20, (21-to-28).* + practice problems 1.
6		**1.8**	**Introduction to linear transformations** Examples: *1, 2, 3, 4, and 5.* Exercises: *2, 4, 6, 10, 12, 14, 16, 19, 20,* and *(21-to-30).* + practice problem 1, 2.
		1.9	**The matrix of a linear transformation** Examples: *1, 2, 3, and 4.* Exercises: *2, 4, 6, 8, 10, 15, 16, 18, 20, 22,* and *(23-to-32).* + practice problems 1, 2.
7	**Chapter 2: Matrix algebra** Lay, D. (2021). Linear Algebra and Its Applications, Global Edition (6th ed.)	**2.1**	**Matrix operations** Examples: *1, 2, 3, 4, 5, 6, 7, and 8.* Exercises: *1, 3, 5, 7, 9, 11,* and *(15-to-24)* + practice problem 1.
		2.2	**The inverse of a matrix** Examples: *1, 2, 4,5, 6, and 7.* Exercises: *2, 4, 6, 8, 10, and (11-to-20).* + practice problems 1, 2, 3.
8	**Chapter 3: Determinants** Lay, D. (2021). Linear Algebra and Its Applications, Global Edition (6th ed.)	**3.1**	**Introduction to determinants** Examples: *1, 2, and 3.* Exercises: *2, 4, 6, 10, 12, 14, 16, 19, 20, 26, 32,* and *(39-to-42).*+ practice problem 1.
		3.2	**Properties of determinants** Examples: *1, 2, 3, 4, and 5.* Exercises: *2, 4, 6, 10, 12, 14,* and *(27-to-34).* + practice problem 1, 2, 3.

		3.3	**Cramer's rule, volume, and linear transformations** Examples: *1, 2, and 3.* Exercises: *1, 3, 5, 7, 9, 11, 13, 15.*
9	**Midterm Exam**		
10	**Chapter 4: Vector spaces** Lay, D. (2021). Linear Algebra and Its Applications, Global Edition (6th ed.)	**4.1**	**Vector spaces and subspaces** Examples: *1, 2, 4, 6, 7, 11, and 12.* Exercises: *1, 3, 9, 10, 11, 12, 13, 15, 17,* and *(23-to-32).*
		4.2	**Null spaces, column spaces, row spaces, and linear transformations** Examples: 1, 2, 3, 4, and 5. Exercises: 1, 3, 5, 7, 9, 11, 13, 15, 19, and (25-to-38).
11		**4.3**	**Linearly independent sets, bases** Examples: 1, 2, 3, 4, 5, 6, 7, and 8. Exercises: 1, 3, 5, 7, 9, 13, 15, and (21-to-32).
		4.5	**The dimension of a vector space** Examples: 1, 2, 3, 4, 5, and 6. Exercises: 1, 3, 5, 7, 9, 11, 15, and (17-to-26).
12	**Chapter 5: Eigenvalues and eigenvectors** Lay, D. (2021). Linear Algebra and Its Applications, Global Edition (6th ed.)	**5.1**	**Eigenvectors and eigenvalues** Examples: *1, 2, 3, 4, and 5.* Exercises: *2, 4, 6, 8, 10, 15, 16, 18, 20,* and *(21-to-30).* + practice problems 1, 2.
		5.2	**The characteristic equation** Examples: *1, 3, and 4.* Exercises: *1, 3, 5, 7, 9, 11, 13, 15, 17,* and *(21-to-30)* + practice problem 1.
13		**5.3**	**Diagonalization** Examples: *1, 2, 3, 4, and 5.* Exercises: *2, 4, 6, 8, 10, 12, 14, 16, 18, 20, and (21-to-28).* + practice problems 1, 2.
	Chapter 6: Orthogonality and least squares Lay, D. (2021). Linear Algebra and Its Applications, Global Edition (6th ed.)	**6.1**	**Inner product, length, and orthogonality** Examples: *1, 2, 3, 4, and 5.* Exercises: *2, 4, 6, 8, 10, 12, 14, 16, 18,* and *(19-to-28).* + practice problems 1, 2.
14		**6.2**	**Orthogonal sets** Examples: *1, 2, 3, and 4.* Exercises: *1, 3, 5, 7, 9, 11, 13, 15, 17,19,* and *(23-to-32)* + practice problem 1, 2, 3.
		6.3	**Orthogonal projections** Examples: *1, 2, 3, and 4.* Exercises: *2, 4, 6, 8, 10, 12, 14, and (21-to-30).* + practice problems 1.
15		**6.6**	**Machine learning and linear models** Examples: *1, and 2.* Exercises: 1, *2, 3, 4, 5, 6, 7, 8, 9, and 10.*
	Chapter 9: Optimization Lay, D. (2021). Linear Algebra and Its Applications, Global Edition (6th ed.)	**9.2**	**Linear programming-Geometric method** Examples: *3, 4, and 5.* Exercises: *3, 4, 5, 6, 7, 8, 9, and 10.*

References:

1- **Lay, D. (2021). Linear Algebra and Its Applications, Global Edition (6th ed.). Pearson Education Limited**
2- **Strang, G. (2023). Introduction to Linear Algebra (6th ed.). Wellesley-Cambridge Press.**

Chapter 1

Introduction to Vectors

The heart of linear algebra is in two operations—both with vectors. We add vectors to get $\boldsymbol{v} + \boldsymbol{w}$. We multiply them by numbers c and d to get $c\boldsymbol{v}$ and $d\boldsymbol{w}$. Combining those two operations (adding $c\boldsymbol{v}$ to $d\boldsymbol{w}$) gives the ***linear combination*** $c\boldsymbol{v} + d\boldsymbol{w}$.

Linear combination $$c\boldsymbol{v} + d\boldsymbol{w} = c\begin{bmatrix} 1 \\ 1 \end{bmatrix} + d\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} c + 2d \\ c + 3d \end{bmatrix}$$

Example $\boldsymbol{v} + \boldsymbol{w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ is the combination with $c = d = 1$

Linear combinations are all-important in this subject! Sometimes we want one particular combination, the specific choice $c = 2$ and $d = 1$ that produces $c\boldsymbol{v} + d\boldsymbol{w} = (4, 5)$. Other times we want *all the combinations* of $\boldsymbol{v}$ and $\boldsymbol{w}$ (coming from all c and d).

The vectors $c\boldsymbol{v}$ lie along a line. When $\boldsymbol{w}$ is not on that line, **the combinations $c\boldsymbol{v} + d\boldsymbol{w}$ fill the whole two-dimensional plane**. Starting from four vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}, \boldsymbol{z}$ in four-dimensional space, their combinations $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w} + f\boldsymbol{z}$ are likely to fill the space—but not always. The vectors and their combinations could lie in a plane or on a line.

Chapter 1 explains these central ideas, on which everything builds. We start with two-dimensional vectors and three-dimensional vectors, which are reasonable to draw. Then we move into higher dimensions. The really impressive feature of linear algebra is how smoothly it takes that step into n-dimensional space. Your mental picture stays completely correct, even if drawing a ten-dimensional vector is impossible.

This is where the book is going (into n-dimensional space). The first steps are the operations in Sections 1.1 and 1.2. Then Section 1.3 outlines three fundamental ideas.

1.1 ***Vector addition $\boldsymbol{v} + \boldsymbol{w}$ and linear combinations $c\boldsymbol{v} + d\boldsymbol{w}$.***

1.2 ***The dot product $\boldsymbol{v} \cdot \boldsymbol{w}$ of two vectors and the length $\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}$.***

1.3 ***Matrices A, linear equations $A\boldsymbol{x} = \boldsymbol{b}$, solutions $\boldsymbol{x} = A^{-1}\boldsymbol{b}$.***

1.1 Vectors and Linear Combinations

1 $3\boldsymbol{v} + 5\boldsymbol{w}$ is a typical **linear combination** $c\boldsymbol{v} + d\boldsymbol{w}$ of the vectors $\boldsymbol{v}$ and $\boldsymbol{w}$.

2 For $\boldsymbol{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ that combination is $3\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 5\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3+10 \\ 3+15 \end{bmatrix} = \begin{bmatrix} 13 \\ 18 \end{bmatrix}$.

3 The vector $\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 3 \end{bmatrix}$ goes across to $x = 2$ and up to $y = 3$ in the xy plane.

4 The combinations $c\begin{bmatrix} 1 \\ 1 \end{bmatrix} + d\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ fill the whole xy plane. They produce every $\begin{bmatrix} x \\ y \end{bmatrix}$.

5 The combinations $c\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + d\begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$ fill a **plane** in xyz space. Same plane for $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$.

6 But $\begin{matrix} c + 2d = 1 \\ c + 3d = 0 \\ c + 4d = 0 \end{matrix}$ has no solution because its right side $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ is not on that plane.

"You can't add apples and oranges." In a strange way, this is the reason for vectors. We have two separate numbers v_1 and v_2. That pair produces a ***two-dimensional vector*** $\boldsymbol{v}$:

Column vector $\boldsymbol{v}$ $\qquad \boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ $\qquad \begin{matrix} v_1 = \text{first component of } \boldsymbol{v} \\ v_2 = \text{second component of } \boldsymbol{v} \end{matrix}$

We write $\boldsymbol{v}$ as a ***column***, not as a row. The main point so far is to have a single letter $\boldsymbol{v}$ (in ***boldface italic***) for this pair of numbers v_1 and v_2 (in *lightface italic*).

Even if we don't add v_1 to v_2, we do ***add vectors***. The first components of $\boldsymbol{v}$ and $\boldsymbol{w}$ stay separate from the second components:

VECTOR ADDITION $\qquad \boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ add to $\boldsymbol{v} + \boldsymbol{w} = \begin{bmatrix} v_1 + w_1 \\ v_2 + w_2 \end{bmatrix}$.

Subtraction follows the same idea: *The components of* $\boldsymbol{v} - \boldsymbol{w}$ *are* $v_1 - w_1$ *and* $v_2 - w_2$.

The other basic operation is *scalar multiplication*. Vectors can be multiplied by 2 or by -1 or by any number c. To find $2\boldsymbol{v}$, multiply each component of $\boldsymbol{v}$ by 2:

SCALAR MULTIPLICATION $\qquad 2\boldsymbol{v} = \begin{bmatrix} 2v_1 \\ 2v_2 \end{bmatrix} = \boldsymbol{v} + \boldsymbol{v} \quad -\boldsymbol{v} = \begin{bmatrix} -v_1 \\ -v_2 \end{bmatrix}$.

The components of $c\boldsymbol{v}$ are cv_1 and cv_2. The number c is called a "scalar".

Notice that the sum of $-\boldsymbol{v}$ and $\boldsymbol{v}$ is the zero vector. This is $\mathbf{0}$, which is not the same as the number zero! The vector $\mathbf{0}$ has components 0 and 0. Forgive me for hammering away at the difference between a vector and its components. Linear algebra is built on these operations $\boldsymbol{v} + \boldsymbol{w}$ and $c\boldsymbol{v}$ and $d\boldsymbol{w}$—***adding vectors and multiplying by scalars***.

Linear Combinations

Now we combine addition with scalar multiplication to produce a "**linear combination**" of $\boldsymbol{v}$ and $\boldsymbol{w}$. Multiply $\boldsymbol{v}$ by c and multiply $\boldsymbol{w}$ by d. Then add $c\boldsymbol{v} + d\boldsymbol{w}$.

The sum of $c\boldsymbol{v}$ and $d\boldsymbol{w}$ is a *linear combination $c\boldsymbol{v} + d\boldsymbol{w}$.*

Four special linear combinations are: sum, difference, zero, and a scalar multiple $c\boldsymbol{v}$:

$1\boldsymbol{v} + 1\boldsymbol{w}$	=	sum of vectors in Figure 1.1a
$1\boldsymbol{v} - 1\boldsymbol{w}$	=	difference of vectors in Figure 1.1b
$0\boldsymbol{v} + 0\boldsymbol{w}$	=	***zero vector***
$c\boldsymbol{v} + 0\boldsymbol{w}$	=	vector $c\boldsymbol{v}$ in the direction of $\boldsymbol{v}$

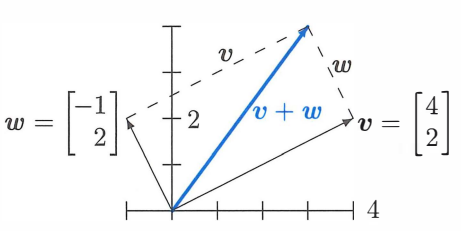
The zero vector is always a possible combination (its coefficients are zero). Every time we see a "space" of vectors, that zero vector will be included. This big view, taking *all* the combinations of $\boldsymbol{v}$ and $\boldsymbol{w}$, is linear algebra at work.

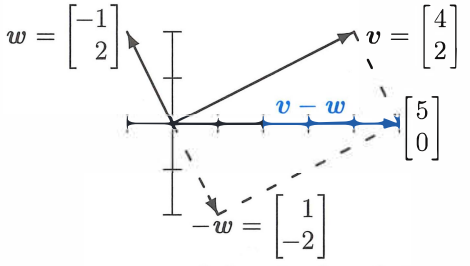
The figures show how you can visualize vectors. For algebra, we just need the components (like 4 and 2). That vector $\boldsymbol{v}$ is represented by an arrow. The arrow goes $v_1 = 4$ units to the right and $v_2 = 2$ units up. It ends at the point whose x, y coordinates are $4, 2$. This point is another representation of the vector—so we have three ways to describe $\boldsymbol{v}$:

Represent vector $\boldsymbol{v}$	Two numbers	Arrow from $(0, 0)$	Point in the plane

We add using the numbers. We visualize $\boldsymbol{v} + \boldsymbol{w}$ using arrows:

Vector addition (head to tail) ***At the end of $\boldsymbol{v}$, place the start of $\boldsymbol{w}$.***





$$\boldsymbol{v} + \boldsymbol{w} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} + \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \qquad \boldsymbol{v} - \boldsymbol{w} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} - \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$$

Figure 1.1: Vector addition $\boldsymbol{v} + \boldsymbol{w} = (3,\ 4)$ produces the diagonal of a parallelogram. The reverse of $\boldsymbol{w}$ is $-\boldsymbol{w}$. The linear combination on the right is $\boldsymbol{v} - \boldsymbol{w} = (5,\ 0)$.

We travel along $\boldsymbol{v}$ and then along $\boldsymbol{w}$. Or we take the diagonal shortcut along $\boldsymbol{v} + \boldsymbol{w}$. We could also go along $\boldsymbol{w}$ and then $\boldsymbol{v}$. In other words, $\boldsymbol{w} + \boldsymbol{v}$ **gives the same answer as** $\boldsymbol{v} + \boldsymbol{w}$. These are different ways along the parallelogram (in this example it is a rectangle).

Vectors in Three Dimensions

A vector with two components corresponds to a point in the xy plane. The components of $\boldsymbol{v}$ are the coordinates of the point: $x = v_1$ and $y = v_2$. The arrow ends at this point (v_1, v_2), when it starts from $(0, 0)$. Now we allow vectors to have three components (v_1, v_2, v_3).

The xy plane is replaced by three-dimensional xyz space. Here are typical vectors (still column vectors but with three components):

$$\boldsymbol{v} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{w} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \quad \text{and} \quad \boldsymbol{v} + \boldsymbol{w} = \begin{bmatrix} 3 \\ 4 \\ 3 \end{bmatrix}.$$

The vector $\boldsymbol{v}$ corresponds to an arrow in 3-space. Usually the arrow starts at the "origin", where the xyz axes meet and the coordinates are $(0, 0, 0)$. The arrow ends at the point with coordinates v_1, v_2, v_3. There is a perfect match between the ***column vector*** and the ***arrow from the origin*** and the ***point where the arrow ends***.

The vector (x, y) in the plane is different from $(x, y, 0)$ in 3-space !

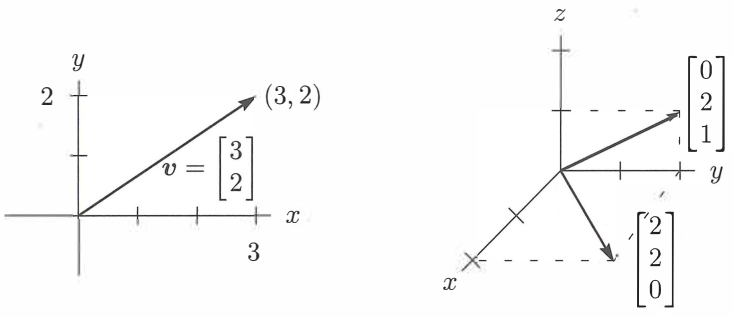


Figure 1.2: Vectors $\begin{bmatrix} x \\ y \end{bmatrix}$ and $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$ correspond to points (x, y) and (x, y, z).

From now on $\boldsymbol{v} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$ ***is also written as*** $\boldsymbol{v} = (1, 1, -1)$.

The reason for the row form (in parentheses) is to save space. But $\boldsymbol{v} = (1, 1, -1)$ is not a row vector! It is in actuality a column vector, just temporarily lying down. The row vector $\begin{bmatrix} 1 & 1 & -1 \end{bmatrix}$ is absolutely different, even though it has the same three components. That 1 by 3 row vector is the "transpose" of the 3 by 1 column vector $\boldsymbol{v}$.

In three dimensions, $\boldsymbol{v} + \boldsymbol{w}$ is still found a component at a time. The sum has components $v_1 + w_1$ and $v_2 + w_2$ and $v_3 + w_3$. You see how to add vectors in 4 or 5 or n dimensions. When $\boldsymbol{w}$ starts at the end of $\boldsymbol{v}$, the third side is $\boldsymbol{v} + \boldsymbol{w}$. The other way around the parallelogram is $\boldsymbol{w} + \boldsymbol{v}$. Question: Do the four sides all lie in the same plane? *Yes*. And the sum $\boldsymbol{v} + \boldsymbol{w} - \boldsymbol{v} - \boldsymbol{w}$ goes completely around to produce the _____ vector.

A typical linear combination of three vectors in three dimensions is $\boldsymbol{u} + 4\boldsymbol{v} - 2\boldsymbol{w}$:

Linear combination
Multiply by $1, 4, -2$
Then add

$$\begin{bmatrix}1\\0\\3\end{bmatrix} + 4\begin{bmatrix}1\\2\\1\end{bmatrix} - 2\begin{bmatrix}2\\3\\-1\end{bmatrix} = \begin{bmatrix}1\\2\\9\end{bmatrix}.$$

The Important Questions

For one vector $\boldsymbol{u}$, the only linear combinations are the multiples $c\boldsymbol{u}$. For two vectors, the combinations are $c\boldsymbol{u} + d\boldsymbol{v}$. For three vectors, the combinations are $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}$. Will you take the big step from *one* combination to **all combinations**? Every c and d and e are allowed. Suppose the vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ are in three-dimensional space:

1. What is the picture of *all* combinations $c\boldsymbol{u}$?
2. What is the picture of *all* combinations $c\boldsymbol{u} + d\boldsymbol{v}$?
3. What is the picture of *all* combinations $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}$?

The answers depend on the particular vectors $\boldsymbol{u}, \boldsymbol{v}$, and $\boldsymbol{w}$. If they were zero vectors (a very extreme case), then every combination would be zero. If they are typical nonzero vectors (components chosen at random), here are the three answers. This is the key to our subject:

1. The combinations $c\boldsymbol{u}$ fill a ***line through*** $(\mathbf{0}, \mathbf{0}, \mathbf{0})$.
2. The combinations $c\boldsymbol{u} + d\boldsymbol{v}$ fill a ***plane through*** $(\mathbf{0}, \mathbf{0}, \mathbf{0})$.
3. The combinations $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}$ fill ***three-dimensional space***.

The zero vector $(0, 0, 0)$ is on the line because c can be zero. It is on the plane because c and d could both be zero. The line of vectors $c\boldsymbol{u}$ is infinitely long (forward and backward). It is the plane of all $c\boldsymbol{u} + d\boldsymbol{v}$ (combining two vectors in three-dimensional space) that I especially ask you to think about.

Adding all $c\boldsymbol{u}$ on one line to all $d\boldsymbol{v}$ on the other line fills in the plane in Figure 1.3.

When we include a third vector $\boldsymbol{w}$, the multiples $e\boldsymbol{w}$ give a third line. **Suppose that third line is not in the plane of $\boldsymbol{u}$ and $\boldsymbol{v}$.** Then combining all $e\boldsymbol{w}$ with all $c\boldsymbol{u} + d\boldsymbol{v}$ fills up the whole three-dimensional space.

This is the typical situation! **Line**, then **plane**, then **space**. But other possibilities exist. When $\boldsymbol{w}$ happens to be $c\boldsymbol{u} + d\boldsymbol{v}$, that third vector $\boldsymbol{w}$ is in the plane of the first two. The combinations of $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ will not go outside that $\boldsymbol{uv}$ plane. We do not get the full three-dimensional space. Please think about the special cases in Problem 1.

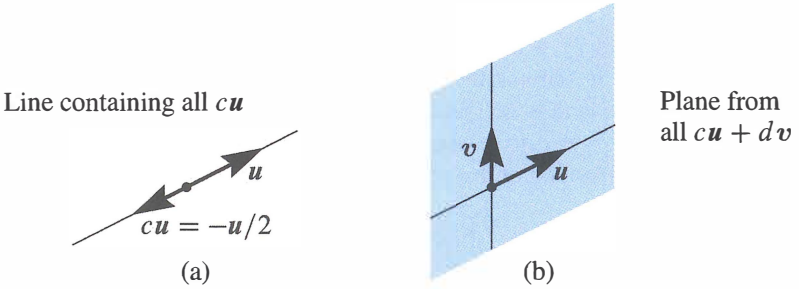


Figure 1.3: (a) Line through $\boldsymbol{u}$. (b) The plane containing the lines through $\boldsymbol{u}$ and $\boldsymbol{v}$.

■ REVIEW OF THE KEY IDEAS ■

1. A vector $\boldsymbol{v}$ in two-dimensional space has two components v_1 and v_2.
2. $\boldsymbol{v} + \boldsymbol{w} = (v_1 + w_1,\, v_2 + w_2)$ and $c\boldsymbol{v} = (cv_1, cv_2)$ are found a component at a time.
3. A linear combination of three vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ and $\boldsymbol{w}$ is $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}$.
4. Take *all* linear combinations of $\boldsymbol{u}$, or $\boldsymbol{u}$ and $\boldsymbol{v}$, or $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$. In three dimensions, those combinations typically fill a line, then a plane, then the whole space $\mathbf{R}^3$.

■ WORKED EXAMPLES ■

1.1 A The linear combinations of $\boldsymbol{v} = (1, 1, 0)$ and $\boldsymbol{w} = (0, 1, 1)$ fill a plane in $\mathbf{R}^3$. *Describe that plane.* Find a vector that is *not* a combination of $\boldsymbol{v}$ and $\boldsymbol{w}$—not on the plane.

Solution The plane of $\boldsymbol{v}$ and $\boldsymbol{w}$ contains all combinations $c\boldsymbol{v} + d\boldsymbol{w}$. The vectors in that plane allow any c and d. The plane of Figure 1.3 fills in between the two lines.

$$\text{Combinations} \quad c\boldsymbol{v} + d\boldsymbol{w} = c\begin{bmatrix}1\\1\\0\end{bmatrix} + d\begin{bmatrix}0\\1\\1\end{bmatrix} = \begin{bmatrix}c\\c+d\\d\end{bmatrix} \text{ fill a plane.}$$

Four vectors in that plane are $(0, 0, 0)$ and $(2, 3, 1)$ and $(5, 7, 2)$ and $(\pi, 2\pi, \pi)$. The second component $c + d$ is always the sum of the first and third components. Like most vectors, $(1, 2, 3)$ *is* ***not*** *in the plane, because* $2 \neq 1 + 3$.

Another description of this plane through $(0, 0, 0)$ is to know that $\boldsymbol{n} = (1, -1, 1)$ is **perpendicular** to the plane. Section 1.2 will confirm that $90°$ angle by testing dot products: $\boldsymbol{v} \cdot \boldsymbol{n} = 0$ and $\boldsymbol{w} \cdot \boldsymbol{n} = 0$. Perpendicular vectors have zero dot products.

1.1 B For $\boldsymbol{v} = (1, 0)$ and $\boldsymbol{w} = (0, 1)$, describe all points $c\boldsymbol{v}$ with $(\mathbf{1})$ *whole numbers* c $(\mathbf{2})$ *nonnegative* numbers $c \geq 0$. Then add all vectors $d\boldsymbol{w}$ and describe all $c\boldsymbol{v} + d\boldsymbol{w}$.

Solution

$(\mathbf{1})$ The vectors $c\boldsymbol{v} = (c, 0)$ with whole numbers c are **equally spaced points** along the x axis (the direction of $\boldsymbol{v}$). They include $(-2, 0), (-1, 0), (0, 0), (1, 0), (2, 0)$.

$(\mathbf{2})$ The vectors $c\boldsymbol{v}$ with $c \geq 0$ fill a ***half-line***. It is the positive x axis. This half-line starts at $(0, 0)$ where $c = 0$. It includes $(100, 0)$ and $(\pi, 0)$ but not $(-100, 0)$.

$(\mathbf{1}')$ Adding all vectors $d\boldsymbol{w} = (0, d)$ puts a vertical line through those equally spaced $c\boldsymbol{v}$. We have infinitely many ***parallel lines*** from (*whole number c, any number d*).

$(\mathbf{2}')$ Adding all vectors $d\boldsymbol{w}$ puts a vertical line through every $c\boldsymbol{v}$ on the half-line. Now we have a ***half-plane***. The right half of the xy plane has any $x \geq 0$ and any y.

1.1 C Find two equations for c and d so that **the linear combination $c\boldsymbol{v} + d\boldsymbol{w}$ equals $\boldsymbol{b}$**:

$$\boldsymbol{v} = \begin{bmatrix} 2 \\ -1 \end{bmatrix} \qquad \boldsymbol{w} = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Solution In applying mathematics, many problems have two parts:

1 *Modeling part* Express the problem by a set of equations.

2 *Computational part* Solve those equations by a fast and accurate algorithm.

Here we are only asked for the first part (the equations). Chapter 2 is devoted to the second part (the solution). Our example fits into a fundamental model for linear algebra:

$$\text{Find } n \text{ numbers } \quad c_1, \ldots, c_n \quad \text{ so that } \quad c_1\boldsymbol{v}_1 + \cdots + c_n\boldsymbol{v}_n = \boldsymbol{b}.$$

For $n = 2$ we will find a formula for the $\boldsymbol{c}$'s. The "elimination method" in Chapter 2 succeeds far beyond $n = 1000$. For n greater than 1 billion, see Chapter 11. Here $n = 2$:

Vector equation $c\boldsymbol{v} + d\boldsymbol{w} = \boldsymbol{b}$

$$c\begin{bmatrix} 2 \\ -1 \end{bmatrix} + d\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

The required equations for c and d just come from the two components separately:

Two ordinary equations

$$\begin{aligned} 2c - d &= 1 \\ -c + 2d &= 0 \end{aligned}$$

Each equation produces a line. The two lines cross at the solution $c = \frac{2}{3}, d = \frac{1}{3}$. Why not see this also as a **matrix equation**, since that is where we are going:

2 by 2 matrix

$$\begin{bmatrix} \mathbf{2} & \mathbf{-1} \\ \mathbf{-1} & \mathbf{2} \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Problem Set 1.1

Problems 1–9 are about addition of vectors and linear combinations.

1 Describe geometrically (line, plane, or all of $\mathbf{R}^3$) all linear combinations of

(a) $\begin{bmatrix}1\\2\\3\end{bmatrix}$ and $\begin{bmatrix}3\\6\\9\end{bmatrix}$ (b) $\begin{bmatrix}1\\0\\0\end{bmatrix}$ and $\begin{bmatrix}0\\2\\3\end{bmatrix}$ (c) $\begin{bmatrix}2\\0\\0\end{bmatrix}$ and $\begin{bmatrix}0\\2\\2\end{bmatrix}$ and $\begin{bmatrix}2\\2\\3\end{bmatrix}$

2 Draw $\boldsymbol{v}=\begin{bmatrix}4\\1\end{bmatrix}$ and $\boldsymbol{w}=\begin{bmatrix}-2\\2\end{bmatrix}$ and $\boldsymbol{v}+\boldsymbol{w}$ and $\boldsymbol{v}-\boldsymbol{w}$ in a single xy plane.

3 If $\boldsymbol{v}+\boldsymbol{w}=\begin{bmatrix}5\\1\end{bmatrix}$ and $\boldsymbol{v}-\boldsymbol{w}=\begin{bmatrix}1\\5\end{bmatrix}$, compute and draw the vectors $\boldsymbol{v}$ and $\boldsymbol{w}$.

4 From $\boldsymbol{v}=\begin{bmatrix}2\\1\end{bmatrix}$ and $\boldsymbol{w}=\begin{bmatrix}1\\2\end{bmatrix}$, find the components of $3\boldsymbol{v}+\boldsymbol{w}$ and $c\boldsymbol{v}+d\boldsymbol{w}$.

5 Compute $\boldsymbol{u}+\boldsymbol{v}+\boldsymbol{w}$ and $2\boldsymbol{u}+2\boldsymbol{v}+\boldsymbol{w}$. How do you know $\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}$ lie in a plane?

These lie in a plane because $\boldsymbol{w}=c\boldsymbol{u}+d\boldsymbol{v}$. Find c and d

$$\boldsymbol{u}=\begin{bmatrix}1\\2\\3\end{bmatrix},\quad \boldsymbol{v}=\begin{bmatrix}-3\\1\\-2\end{bmatrix},\quad \boldsymbol{w}=\begin{bmatrix}2\\-3\\-1\end{bmatrix}.$$

6 Every combination of $\boldsymbol{v}=(1,-2,1)$ and $\boldsymbol{w}=(0,1,-1)$ has components that add to _____. Find c and d so that $c\boldsymbol{v}+d\boldsymbol{w}=(3,3,-6)$. Why is $(3,3,6)$ impossible?

7 In the xy plane mark all nine of these linear combinations:

$$c\begin{bmatrix}2\\1\end{bmatrix}+d\begin{bmatrix}0\\1\end{bmatrix}\quad \text{with}\quad c=0,1,2\quad \text{and}\quad d=0,1,2.$$

8 The parallelogram in Figure 1.1 has diagonal $\boldsymbol{v}+\boldsymbol{w}$. What is its other diagonal? What is the sum of the two diagonals? Draw that vector sum.

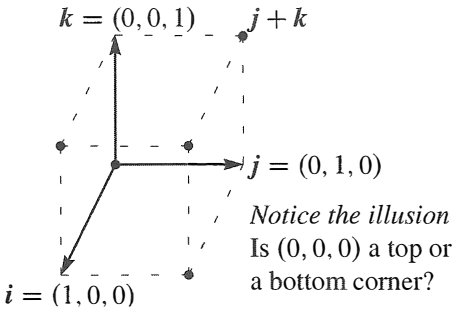
9 If three corners of a parallelogram are $(1,1)$, $(4,2)$, and $(1,3)$, what are all three of the possible fourth corners? Draw two of them.

Problems 10–14 are about special vectors on cubes and clocks in Figure 1.4.

10 Which point of the cube is $\boldsymbol{i}+\boldsymbol{j}$? Which point is the vector sum of $\boldsymbol{i}=(1,0,0)$ and $\boldsymbol{j}=(0,1,0)$ and $\boldsymbol{k}=(0,0,1)$? Describe all points (x,y,z) in the cube.

11 Four corners of this unit cube are $(0,0,0)$, $(1,0,0)$, $(0,1,0)$, $(0,0,1)$. What are the other four corners? Find the coordinates of the center point of the cube. The center points of the six faces are _____. The cube has how many edges?

12 *Review Question.* In xyz space, where is the plane of all linear combinations of $\boldsymbol{i}=(1,0,0)$ and $\boldsymbol{i}+\boldsymbol{j}=(1,1,0)$?



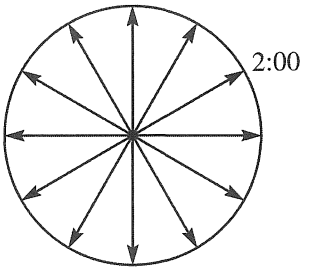


Figure 1.4: Unit cube from $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ and twelve clock vectors.

13 (a) What is the sum V of the twelve vectors that go from the center of a clock to the hours 1:00, 2:00, ..., 12:00?

(b) If the 2:00 vector is removed, why do the 11 remaining vectors add to 8:00?

(c) What are the x, y components of that 2:00 vector $\boldsymbol{v} = (\cos\theta, \sin\theta)$?

14 Suppose the twelve vectors start from 6:00 at the bottom instead of $(0, 0)$ at the center. The vector to 12:00 is doubled to $(0, 2)$. The new twelve vectors add to ____.

Problems 15–19 go further with linear combinations of v and w (Figure 1.5a).

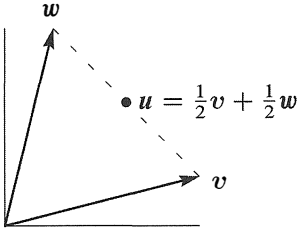
15 Figure 1.5a shows $\frac{1}{2}\,\boldsymbol{v} + \frac{1}{2}\,\boldsymbol{w}$. Mark the points $\frac{3}{4}\,\boldsymbol{v} + \frac{1}{4}\,\boldsymbol{w}$ and $\frac{1}{4}\,\boldsymbol{v} + \frac{1}{4}\,\boldsymbol{w}$ and $\boldsymbol{v} + \boldsymbol{w}$.

16 Mark the point $-\boldsymbol{v} + 2\boldsymbol{w}$ and any other combination $c\boldsymbol{v} + d\boldsymbol{w}$ with $c + d = 1$. Draw the line of all combinations that have $c + d = 1$.

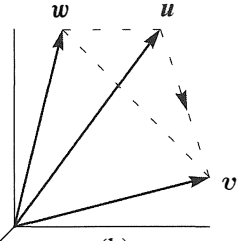
17 Locate $\frac{1}{3}\,\boldsymbol{v} + \frac{1}{3}\,\boldsymbol{w}$ and $\frac{2}{3}\,\boldsymbol{v} + \frac{2}{3}\,\boldsymbol{w}$. The combinations $c\boldsymbol{v} + c\boldsymbol{w}$ fill out what line?

18 Restricted by $0 \le c \le 1$ and $0 \le d \le 1$, shade in all combinations $c\boldsymbol{v} + d\boldsymbol{w}$.

19 Restricted only by $c \ge 0$ and $d \ge 0$ draw the "cone" of all combinations $c\boldsymbol{v} + d\boldsymbol{w}$.



(a)



(b)

Figure 1.5: Problems **15–19** in a plane Problems **20–25** in 3-dimensional space

Problems 20–25 deal with u, v, w in three-dimensional space (see Figure 1.5b).

20 Locate $\frac{1}{3}\boldsymbol{u} + \frac{1}{3}\boldsymbol{v} + \frac{1}{3}\boldsymbol{w}$ and $\frac{1}{2}\boldsymbol{u} + \frac{1}{2}\boldsymbol{w}$ in Figure 1.5b. Challenge problem: Under what restrictions on c, d, e, will the combinations $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}$ fill in the dashed triangle? To stay in the triangle, one requirement is $c \geq 0, d \geq 0, e \geq 0$.

21 The three sides of the dashed triangle are $\boldsymbol{v} - \boldsymbol{u}$ and $\boldsymbol{w} - \boldsymbol{v}$ and $\boldsymbol{u} - \boldsymbol{w}$. Their sum is _____. Draw the head-to-tail addition around a plane triangle of $(3, 1)$ plus $(-1, 1)$ plus $(-2, -2)$.

22 Shade in the pyramid of combinations $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}$ with $c \geq 0$, $d \geq 0$, $e \geq 0$ and $c + d + e \leq 1$. Mark the vector $\frac{1}{2}(\boldsymbol{u} + \boldsymbol{v} + \boldsymbol{w})$ as inside or outside this pyramid.

23 If you look at *all* combinations of those $\boldsymbol{u}, \boldsymbol{v}$, and $\boldsymbol{w}$, is there any vector that can't be produced from $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}$? Different answer if $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ are all in _____.

24 Which vectors are combinations of $\boldsymbol{u}$ and $\boldsymbol{v}$, and *also* combinations of $\boldsymbol{v}$ and $\boldsymbol{w}$?

25 Draw vectors $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ so that their combinations $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}$ fill only a line. Find vectors $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ so that their combinations $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w}$ fill only a plane.

26 What combination $c\begin{bmatrix}\mathbf{1}\\\mathbf{2}\end{bmatrix} + d\begin{bmatrix}\mathbf{3}\\\mathbf{1}\end{bmatrix}$ produces $\begin{bmatrix}\mathbf{14}\\\mathbf{8}\end{bmatrix}$? Express this question as two equations for the coefficients c and d in the linear combination.

Challenge Problems

27 How many corners does a cube have in 4 dimensions? How many 3D faces? How many edges? A typical corner is $(0, 0, 1, 0)$. A typical edge goes to $(0, 1, 0, 0)$.

28 Find vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ so that $\boldsymbol{v} + \boldsymbol{w} = (4, 5, 6)$ and $\boldsymbol{v} - \boldsymbol{w} = (2, 5, 8)$. This is a question with _____ unknown numbers, and an equal number of equations to find those numbers.

29 Find *two different combinations* of the three vectors $\boldsymbol{u} = (1, 3)$ and $\boldsymbol{v} = (2, 7)$ and $\boldsymbol{w} = (1, 5)$ that produce $\boldsymbol{b} = (0, 1)$. Slightly delicate question: If I take any three vectors $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ in the plane, will there always be two different combinations that produce $\boldsymbol{b} = (0, 1)$?

30 The linear combinations of $\boldsymbol{v} = (a, b)$ and $\boldsymbol{w} = (c, d)$ fill the plane unless _____. Find four vectors $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$, $\boldsymbol{z}$ with four components each so that their combinations $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w} + f\boldsymbol{z}$ produce all vectors (b_1, b_2, b_3, b_4) in four-dimensional space.

31 Write down three equations for c, d, e so that $c\boldsymbol{u} + d\boldsymbol{v} + e\boldsymbol{w} = \boldsymbol{b}$. Can you somehow find c, d, e for this $\boldsymbol{b}$?

$$\boldsymbol{u} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} \quad \boldsymbol{v} = \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix} \quad \boldsymbol{w} = \begin{bmatrix} 0 \\ -1 \\ 2 \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

1.2 Lengths and Dot Products

1 The "dot product" of $\boldsymbol{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} 4 \\ 5 \end{bmatrix}$ is $\boldsymbol{v} \cdot \boldsymbol{w} = (1)(4) + (2)(5) = 4 + 10 = \mathbf{14}$.

2 $\boldsymbol{v} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} 4 \\ -4 \\ 4 \end{bmatrix}$ are perpendicular because $\boldsymbol{v} \cdot \boldsymbol{w}$ is zero:
$(1)(4) + (3)(-4) + (2)(4) = \mathbf{0}$.

3 The length squared of $\boldsymbol{v} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$ is $\boldsymbol{v} \cdot \boldsymbol{v} = 1 + 9 + 4 = 14$. **The length is** $||\boldsymbol{v}|| = \sqrt{\mathbf{14}}$.

4 Then $\boldsymbol{u} = \dfrac{\boldsymbol{v}}{||\boldsymbol{v}||} = \dfrac{\boldsymbol{v}}{\sqrt{14}} = \dfrac{1}{\sqrt{14}} \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$ has length $||\boldsymbol{u}|| = \mathbf{1}$. Check $\dfrac{1}{14} + \dfrac{9}{14} + \dfrac{4}{14} = 1$.

5 The angle θ between $\boldsymbol{v}$ and $\boldsymbol{w}$ has $\cos\theta = \dfrac{\boldsymbol{v} \cdot \boldsymbol{w}}{||\boldsymbol{v}||\ ||\boldsymbol{w}||}$.

6 The angle between $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ has $\cos\theta = \dfrac{1}{(1)(\sqrt{2})}$. That angle is $\theta = 45°$.

7 All angles have $|\cos\theta| \le 1$. So all vectors have $\boxed{|\boldsymbol{v} \cdot \boldsymbol{w}| \le ||\boldsymbol{v}||\ ||\boldsymbol{w}||.}$

The first section backed off from multiplying vectors. Now we go forward to define the *"dot product"* of $\boldsymbol{v}$ and $\boldsymbol{w}$. This multiplication involves the separate products $v_1 w_1$ and $v_2 w_2$, but it doesn't stop there. Those two numbers are added to produce one number $\boldsymbol{v} \cdot \boldsymbol{w}$.

This is the geometry section (lengths of vectors and cosines of angles between them).

The ***dot product*** or ***inner product*** of $\boldsymbol{v} = (v_1, v_2)$ and $\boldsymbol{w} = (w_1, w_2)$ is the number $\boldsymbol{v} \cdot \boldsymbol{w}$:

$$\boldsymbol{v} \cdot \boldsymbol{w} = v_1 w_1 + v_2 w_2. \tag{1}$$

Example 1 The vectors $\boldsymbol{v} = (4, 2)$ and $\boldsymbol{w} = (-1, 2)$ have a *zero* dot product:

Dot product is zero
Perpendicular vectors

$$\begin{bmatrix} 4 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 2 \end{bmatrix} = -4 + 4 = 0.$$

In mathematics, zero is always a special number. For dot products, it means that *these two vectors are perpendicular*. The angle between them is 90°. When we drew them in Figure 1.1, we saw a rectangle (not just any parallelogram). The clearest example of perpendicular vectors is $\boldsymbol{i} = (1, 0)$ along the x axis and $\boldsymbol{j} = (0, 1)$ up the y axis. Again the dot product is $\boldsymbol{i} \cdot \boldsymbol{j} = 0 + 0 = 0$. Those vectors $\boldsymbol{i}$ and $\boldsymbol{j}$ form a right angle.

The dot product of $\boldsymbol{v} = (1, 2)$ and $\boldsymbol{w} = (3, 1)$ is 5. Soon $\boldsymbol{v} \cdot \boldsymbol{w}$ will reveal the angle between $\boldsymbol{v}$ and $\boldsymbol{w}$ (not 90°). Please check that $\boldsymbol{w} \cdot \boldsymbol{v}$ is also 5.

The dot product $\boldsymbol{w} \cdot \boldsymbol{v}$ equals $\boldsymbol{v} \cdot \boldsymbol{w}$. The order of $\boldsymbol{v}$ and $\boldsymbol{w}$ makes no difference.

Example 2 Put a weight of 4 at the point $x = -1$ (left of zero) and a weight of 2 at the point $x = 2$ (right of zero). The x axis will balance on the center point (like a see-saw). The weights balance because the dot product is $(4)(-1) + (2)(2) = 0$.

This example is typical of engineering and science. The vector of weights is $(w_1, w_2) = (4, 2)$. The vector of distances from the center is $(v_1, v_2) = (-1, 2)$. The weights times the distances, $w_1 v_1$ and $w_2 v_2$, give the "moments". The equation for the see-saw to balance is $w_1 v_1 + w_2 v_2 = 0$.

Example 3 Dot products enter in economics and business. We have three goods to buy and sell. Their prices are (p_1, p_2, p_3) for each unit—this is the "price vector" $\boldsymbol{p}$. The quantities we buy or sell are (q_1, q_2, q_3)—positive when we sell, negative when we buy. *Selling q_1 units at the price p_1 brings in $q_1 p_1$.* The total income (quantities q times prices p) is ***the dot product $\boldsymbol{q} \cdot \boldsymbol{p}$ in three dimensions***:

$$\textbf{\textit{Income}} = (q_1, q_2, q_3) \cdot (p_1, p_2, p_3) = q_1 p_1 + q_2 p_2 + q_3 p_3 = \textbf{\textit{dot product}}.$$

A zero dot product means that "the books balance". Total sales equal total purchases if $\boldsymbol{q} \cdot \boldsymbol{p} = 0$. Then $\boldsymbol{p}$ is perpendicular to $\boldsymbol{q}$ (in three-dimensional space). A supermarket with thousands of goods goes quickly into high dimensions.

Small note: Spreadsheets have become essential in management. They compute linear combinations and dot products. What you see on the screen is a matrix.

Main point For $\boldsymbol{v} \cdot \boldsymbol{w}$, multiply each v_i times w_i. Then $\boldsymbol{v} \cdot \boldsymbol{w} = v_1 w_1 + \cdots + v_n w_n$.

Lengths and Unit Vectors

An important case is the dot product of a vector *with itself*. In this case $\boldsymbol{v}$ equals $\boldsymbol{w}$. When the vector is $\boldsymbol{v} = (1, 2, 3)$, the dot product with itself is $\boldsymbol{v} \cdot \boldsymbol{v} = \|\boldsymbol{v}\|^2 = 14$:

Dot product $\boldsymbol{v} \cdot \boldsymbol{v}$
Length squared

$$\|\boldsymbol{v}\|^2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 1 + 4 + 9 = \mathbf{14}.$$

Instead of a 90° angle between vectors we have 0°. The answer is not zero because $\boldsymbol{v}$ is not perpendicular to itself. The dot product $\boldsymbol{v} \cdot \boldsymbol{v}$ gives the *length of $\boldsymbol{v}$ squared*.

DEFINITION The ***length*** $\|\boldsymbol{v}\|$ of a vector $\boldsymbol{v}$ is the square root of $\boldsymbol{v} \cdot \boldsymbol{v}$:

$$\textbf{length} = \|\boldsymbol{v}\| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}} = \left(v_1^2 + v_2^2 + \cdots + v_n^2\right)^{1/2}.$$

In two dimensions the length is $\sqrt{v_1^2 + v_2^2}$. In three dimensions it is $\sqrt{v_1^2 + v_2^2 + v_3^2}$. By the calculation above, the length of $\boldsymbol{v} = (1, 2, 3)$ is $\|\boldsymbol{v}\| = \sqrt{14}$.

Here $\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}$ is just the ordinary length of the arrow that represents the vector. If the components are 1 and 2, the arrow is the third side of a right triangle (Figure 1.6). The Pythagoras formula $a^2 + b^2 = c^2$ connects the three sides: $1^2 + 2^2 = \|\boldsymbol{v}\|^2$.

For the length of $\boldsymbol{v} = (1, 2, 3)$, we used the right triangle formula twice. The vector $(1, 2, 0)$ in the base has length $\sqrt{5}$. This base vector is perpendicular to $(0, 0, 3)$ that goes straight up. So the diagonal of the box has length $\|\boldsymbol{v}\| = \sqrt{5 + 9} = \sqrt{14}$.

The length of a four-dimensional vector would be $\sqrt{v_1^2 + v_2^2 + v_3^2 + v_4^2}$. Thus the vector $(1, 1, 1, 1)$ has length $\sqrt{1^2 + 1^2 + 1^2 + 1^2} = 2$. This is the diagonal through a unit cube in four-dimensional space. That diagonal in n dimensions has length $\sqrt{n}$.

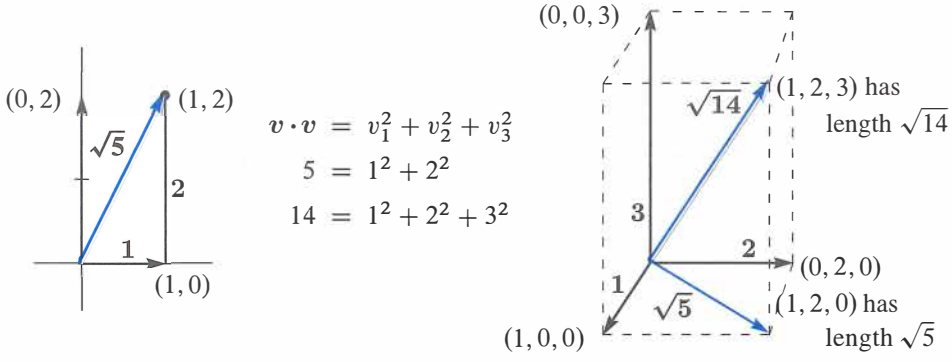


Figure 1.6: The length $\sqrt{\boldsymbol{v} \cdot \boldsymbol{v}}$ of two-dimensional and three-dimensional vectors.

The word "**unit**" is always indicating that some measurement equals "one". The unit price is the price for one item. A unit cube has sides of length one. A unit circle is a circle with radius one. Now we see the meaning of a "unit vector".

DEFINITION ***A unit vector $\boldsymbol{u}$ is a vector whose length equals one.*** Then $\boldsymbol{u} \cdot \boldsymbol{u} = 1$.

An example in four dimensions is $\boldsymbol{u} = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$. Then $\boldsymbol{u} \cdot \boldsymbol{u}$ is $\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = 1$. We divided $\boldsymbol{v} = (1, 1, 1, 1)$ by its length $\|\boldsymbol{v}\| = 2$ to get this unit vector.

Example 4 The standard unit vectors along the x and y axes are written $\boldsymbol{i}$ and $\boldsymbol{j}$. In the xy plane, the unit vector that makes an angle "theta" with the x axis is $(\cos\theta, \sin\theta)$:

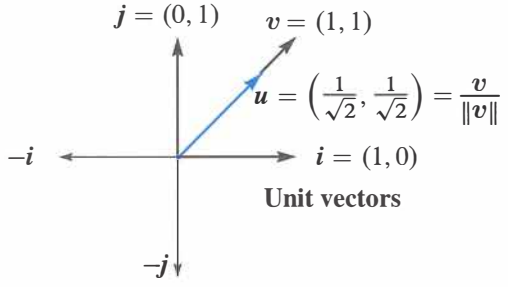
$$\textbf{Unit vectors} \quad \boldsymbol{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{u} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}.$$

When $\theta = 0$, the horizontal vector $\boldsymbol{u}$ is $\boldsymbol{i}$. When $\theta = 90°$ (or $\frac{\pi}{2}$ radians), the vertical vector is $\boldsymbol{j}$. At any angle, the components $\cos\theta$ and $\sin\theta$ produce $\boldsymbol{u} \cdot \boldsymbol{u} = 1$ because

$\cos^2\theta + \sin^2\theta = 1$. These vectors reach out to the unit circle in Figure 1.7. Thus $\cos\theta$ and $\sin\theta$ are simply the coordinates of that point at angle θ on the unit circle.

Since $(2, 2, 1)$ has length 3, the vector $\left(\frac{2}{3}, \frac{2}{3}, \frac{1}{3}\right)$ has length 1. Check that $\boldsymbol{u} \cdot \boldsymbol{u} = \frac{4}{9} + \frac{4}{9} + \frac{1}{9} = 1$. For a unit vector, **divide any nonzero vector $\boldsymbol{v}$ by its length $\|\boldsymbol{v}\|$.**

Unit vector $\boldsymbol{u} = \boldsymbol{v}/\|\boldsymbol{v}\|$ **is a unit vector in the same direction as $\boldsymbol{v}$.**



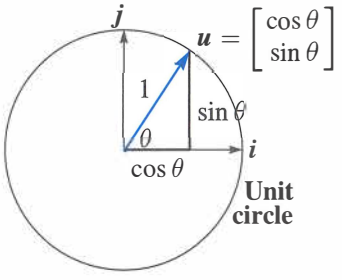


Figure 1.7: The coordinate vectors $\boldsymbol{i}$ and $\boldsymbol{j}$. The unit vector $\boldsymbol{u}$ at angle 45° (left) divides $\boldsymbol{v} = (1, 1)$ by its length $\|\boldsymbol{v}\| = \sqrt{2}$. The unit vector $\boldsymbol{u} = (\cos\theta, \sin\theta)$ is at angle θ.

The Angle Between Two Vectors

We stated that perpendicular vectors have $\boldsymbol{v} \cdot \boldsymbol{w} = 0$. The dot product is zero when the angle is 90°. To explain this, we have to connect angles to dot products. Then we show how $\boldsymbol{v} \cdot \boldsymbol{w}$ finds the angle between any two nonzero vectors $\boldsymbol{v}$ and $\boldsymbol{w}$.

Right angles ***The dot product is $\boldsymbol{v} \cdot \boldsymbol{w} = 0$ when $\boldsymbol{v}$ is perpendicular to $\boldsymbol{w}$.***

Proof When $\boldsymbol{v}$ and $\boldsymbol{w}$ are perpendicular, they form two sides of a right triangle. The third side is $\boldsymbol{v} - \boldsymbol{w}$ (the hypotenuse going across in Figure 1.8). The *Pythagoras Law* for the sides of a right triangle is $a^2 + b^2 = c^2$:

$$\textbf{Perpendicular vectors} \quad \|\boldsymbol{v}\|^2 + \|\boldsymbol{w}\|^2 = \|\boldsymbol{v} - \boldsymbol{w}\|^2 \tag{2}$$

Writing out the formulas for those lengths in two dimensions, this equation is

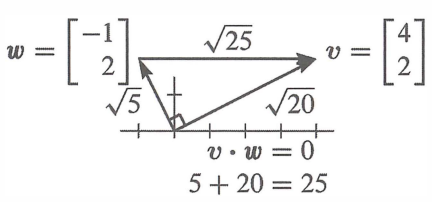
$$\textbf{Pythagoras} \quad (v_1^2 + v_2^2) + (w_1^2 + w_2^2) = (v_1 - w_1)^2 + (v_2 - w_2)^2. \tag{3}$$

The right side begins with $v_1^2 - 2v_1w_1 + w_1^2$. Then v_1^2 and w_1^2 are on both sides of the equation and they cancel, leaving $-2v_1w_1$. Also v_2^2 and w_2^2 cancel, leaving $-2v_2w_2$. (In three dimensions there would be $-2v_3w_3$.) Now divide by -2 to see $\boldsymbol{v} - \boldsymbol{w} = 0$:

$$0 = -2v_1w_1 - 2v_2w_2 \quad \text{which leads to} \quad \boldsymbol{v_1w_1 + v_2w_2 = 0}. \tag{4}$$

Conclusion Right angles produce $\boldsymbol{v} \cdot \boldsymbol{w} = 0$. The dot product is zero when the angle is $\theta = 90°$. Then $\cos\theta = 0$. The zero vector $\boldsymbol{v} = \boldsymbol{0}$ is perpendicular to every vector $\boldsymbol{w}$ because $\boldsymbol{0} \cdot \boldsymbol{w}$ is always zero.

Now suppose $v \cdot w$ is **not zero**. It may be positive, it may be negative. The sign of $v \cdot w$ immediately tells whether we are below or above a right angle. The angle is less than $90°$ when $v \cdot w$ is positive. The angle is above $90°$ when $v \cdot w$ is negative. The right side of Figure 1.8 shows a typical vector $v = (3, 1)$. The angle with $w = (1, 3)$ is less than $90°$ because $v \cdot w = 6$ is positive.



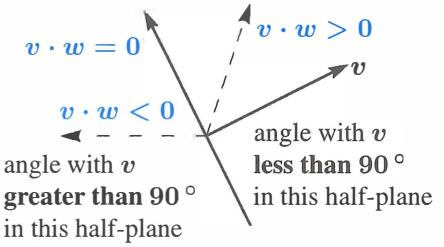


Figure 1.8: Perpendicular vectors have $v \cdot w = 0$. Then $\|v\|^2 + \|w\|^2 = \|v - w\|^2$.

The borderline is where vectors are perpendicular to v. On that dividing line between plus and minus, $(1, -3)$ is perpendicular to $(3, 1)$. The dot product is zero.

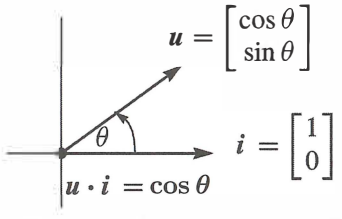
The dot product reveals the exact angle θ. For unit vectors u and U, the sign of $u \cdot U$ tells whether $\theta < 90°$ or $\theta > 90°$. More than that, ***the dot product $u \cdot U$ is the cosine of θ.*** This remains true in n dimensions.

Unit vectors u and U at angle θ have $u \cdot U = \cos\theta$. **Certainly** $|u \cdot U| \leq 1$.

Remember that $\cos\theta$ is never greater than 1. It is never less than -1. *The dot product of unit vectors is between* -1 *and* 1. **The cosine of θ is revealed by $u \cdot U$.**

Figure 1.9 shows this clearly when the vectors are $u = (\cos\theta, \sin\theta)$ and $i = (1, 0)$. The dot product is $u \cdot i = \cos\theta$. That is the cosine of the angle between them.

After rotation through any angle α, these are still unit vectors. The vector $i = (1, 0)$ rotates to $(\cos\alpha, \sin\alpha)$. The vector u rotates to $(\cos\beta, \sin\beta)$ with $\beta = \alpha + \theta$. Their dot product is $\cos\alpha\cos\beta + \sin\alpha\sin\beta$. From trigonometry this is $\cos(\beta - \alpha) = \cos\theta$.



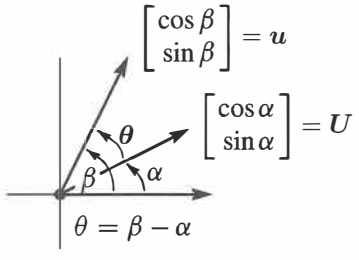


Figure 1.9: Unit vectors: $u \cdot U$ is the cosine of θ (the angle between).

What if v and w are not unit vectors? Divide by their lengths to get $u = v/\|v\|$ and $U = w/\|w\|$. Then the dot product of those unit vectors u and U gives $\cos\theta$.

COSINE FORMULA If v and w are nonzero vectors then $\dfrac{v \cdot w}{\|v\|\,\|w\|} = \cos\theta.$ (5)

Whatever the angle, this dot product of $v/\|v\|$ with $w/\|w\|$ never exceeds one. That is the ***"Schwarz inequality"*** $|v \cdot w| \le \|v\|\,\|w\|$ for dot products—or more correctly the Cauchy-Schwarz-Buniakowsky inequality. It was found in France and Germany and Russia (and maybe elsewhere—it is the most important inequality in mathematics).

Since $|\cos\theta|$ never exceeds 1, the cosine formula gives two great inequalities:

SCHWARZ INEQUALITY $|v \cdot w| \le \|v\|\,\|w\|$

TRIANGLE INEQUALITY $\|v + w\| \le \|v\| + \|w\|$

Example 5 Find $\cos\theta$ for $v = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $w = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and check both inequalities.

Solution The dot product is $v \cdot w = 4$. Both v and w have length $\sqrt{5}$. The cosine is $4/5$.

$$\cos\theta = \frac{v \cdot w}{\|v\|\,\|w\|} = \frac{4}{\sqrt{5}\sqrt{5}} = \frac{4}{5}.$$

By the Schwarz inequality, $v \cdot w = 4$ is less than $\|v\|\,\|w\| = 5$. By the triangle inequality, side $3 = \|v + w\|$ is less than side 1 + side 2. For $v + w = (3, 3)$ the three sides are $\sqrt{18} < \sqrt{5} + \sqrt{5}$. Square this triangle inequality to get $18 < 20$.

Example 6 The dot product of $v = (a, b)$ and $w = (b, a)$ is $2ab$. Both lengths are $\sqrt{a^2 + b^2}$. The Schwarz inequality $v \cdot w \le \|v\|\,\|w\|$ says that $\mathbf{2ab \le a^2 + b^2}$.

This is more famous if we write $x = a^2$ and $y = b^2$. The "geometric mean" $\sqrt{xy}$ is not larger than the "arithmetic mean" = average $\frac{1}{2}(x + y)$.

$$\underset{\text{mean}}{\textbf{Geometric}} \le \underset{\text{mean}}{\textbf{Arithmetic}} \qquad ab \le \frac{a^2 + b^2}{2} \quad \text{becomes} \quad \sqrt{xy} \le \frac{x + y}{2}.$$

Example 5 had $a = 2$ and $b = 1$. So $x = 4$ and $y = 1$. The geometric mean $\sqrt{xy} = 2$ is below the arithmetic mean $\frac{1}{2}(1 + 4) = 2.5$.

Notes on Computing

MATLAB, Python *and* Julia *work directly with whole vectors, not their components.* When v and w have been defined, $v + w$ is immediately understood. Input v and w as rows—the prime $'$ transposes them to columns. $2v + 3w$ becomes $2 * v + 3 * w$. The result will be printed unless the line ends in a semicolon.

MATLAB $v = [2 \quad 3 \quad 4]' \;;\; w = [1 \quad 1 \quad 1]' \;;\; u = 2 * v + 3 * w$

The dot product $v \cdot w$ is ***a row vector times a column vector (use * instead of $\cdot$)***:

Instead of $\begin{bmatrix}1\\2\end{bmatrix} \cdot \begin{bmatrix}3\\4\end{bmatrix}$ we more often see $[1\ 2]\begin{bmatrix}3\\4\end{bmatrix}$ or $v' * w$

The length of v is known to MATLAB as norm (v). This is sqrt $(v' * v)$. Then find the cosine from the dot product $v' * w$ and the angle (in radians) that has that cosine:

Cosine formula cosine $= v' * w/(\text{norm}\,(v) * \text{norm}\,(w))$
The arc cosine angle $=$ acos (cosine)

An M-file would create a new function **cosine** (v, w). Python and Julia are open source.

■ REVIEW OF THE KEY IDEAS ■

1. The dot product $v \cdot w$ multiplies each component v_i by w_i and adds all $v_i w_i$.
2. The length $\|v\|$ is the square root of $v \cdot v$. Then $u = v/\|v\|$ is a ***unit vector***: length 1.
3. The dot product is $v \cdot w = 0$ when vectors v and w are perpendicular.
4. The cosine of θ (the angle between any nonzero v and w) never exceeds 1:

Cosine $\cos\theta = \dfrac{v \cdot w}{\|v\|\|w\|}$ ***Schwarz inequality*** $|v \cdot w| \le \|v\|\,\|w\|$.

■ WORKED EXAMPLES ■

1.2 A For the vectors $v = (3, 4)$ and $w = (4, 3)$ test the Schwarz inequality on $v \cdot w$ and the triangle inequality on $\|v + w\|$. Find $\cos\theta$ for the angle between v and w. Which v and w give *equality* $|v \cdot w| = \|v\|\,\|w\|$ and $\|v + w\| = \|v\| + \|w\|$?

Solution The dot product is $v \cdot w = (3)(4) + (4)(3) = 24$. The length of v is $\|v\| = \sqrt{9 + 16} = 5$ and also $\|w\| = 5$. The sum $v + w = (7, 7)$ has length $7\sqrt{2} < 10$.

Schwarz inequality $|v \cdot w| \le \|v\|\,\|w\|$ is $24 < 25$.

Triangle inequality $\|v + w\| \le \|v\| + \|w\|$ is $7\sqrt{2} < 5 + 5$.

Cosine of angle $\cos\theta = \frac{24}{25}$ Thin angle from $v = (3, 4)$ to $w = (4, 3)$

Equality: One vector is a multiple of the other as in $w = cv$. Then the angle is $0°$ or $180°$. In this case $|\cos\theta| = 1$ and $|v \cdot w|$ *equals* $\|v\|\,\|w\|$. If the angle is $0°$, as in $w = 2v$, then $\|v + w\| = \|v\| + \|w\|$ (both sides give $3\|v\|$). This $v, 2v, 3v$ triangle is flat!

1.2 B Find a unit vector $\boldsymbol{u}$ in the direction of $\boldsymbol{v} = (3, 4)$. Find a unit vector $\boldsymbol{U}$ that is perpendicular to $\boldsymbol{u}$. How many possibilities for $\boldsymbol{U}$?

Solution For a unit vector $\boldsymbol{u}$, divide $\boldsymbol{v}$ by its length $\|\boldsymbol{v}\| = 5$. For a perpendicular vector $\boldsymbol{V}$ we can choose $(-4, 3)$ since the dot product $\boldsymbol{v} \cdot \boldsymbol{V}$ is $(3)(-4) + (4)(3) = 0$. For a *unit* vector perpendicular to $\boldsymbol{u}$, divide $\boldsymbol{V}$ by its length $\|\boldsymbol{V}\|$:

$$\boldsymbol{u} = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|} = \left(\frac{3}{5}, \frac{4}{5}\right) \qquad \boldsymbol{U} = \frac{\boldsymbol{V}}{\|\boldsymbol{V}\|} = \left(-\frac{4}{5}, \frac{3}{5}\right) \qquad \boldsymbol{u} \cdot \boldsymbol{U} = 0$$

The only other perpendicular unit vector would be $-\boldsymbol{U} = (\frac{4}{5}, -\frac{3}{5})$.

1.2 C Find a vector $\boldsymbol{x} = (c, d)$ that has dot products $\boldsymbol{x} \cdot \boldsymbol{r} = 1$ and $\boldsymbol{x} \cdot \boldsymbol{s} = 0$ with two given vectors $\boldsymbol{r} = (2, -1)$ and $\boldsymbol{s} = (-1, 2)$.

Solution Those two dot products give linear equations for c and d. Then $\boldsymbol{x} = (c, d)$.

$\boldsymbol{x} \cdot \boldsymbol{r} = 1$	is	$2c - d = 1$	**The same equations as**
$\boldsymbol{x} \cdot \boldsymbol{s} = 0$	is	$-c + 2d = 0$	**in Worked Example 1.1 C**

Comment on n equations for $\boldsymbol{x} = (x_1, \ldots, x_n)$ in n-dimensional space

Section 1.1 would start with columns $\boldsymbol{v}_j$. The goal is to produce $x_1\boldsymbol{v}_1 + \cdots + x_n\boldsymbol{v}_n = \boldsymbol{b}$. This section would start from rows $\boldsymbol{r}_i$. Now the goal is to find $\boldsymbol{x}$ with $\boldsymbol{x} \cdot \boldsymbol{r}_i = b_i$.

Soon the $\boldsymbol{v}$'s will be the columns of a matrix A, and the $\boldsymbol{r}$'s will be the rows of A. Then the (one and only) problem will be to solve $A\boldsymbol{x} = \boldsymbol{b}$.

Problem Set 1.2

1 Calculate the dot products $\boldsymbol{u} \cdot \boldsymbol{v}$ and $\boldsymbol{u} \cdot \boldsymbol{w}$ and $\boldsymbol{u} \cdot (\boldsymbol{v} + \boldsymbol{w})$ and $\boldsymbol{w} \cdot \boldsymbol{v}$:

$$\boldsymbol{u} = \begin{bmatrix} -.6 \\ .8 \end{bmatrix} \qquad \boldsymbol{v} = \begin{bmatrix} 4 \\ 3 \end{bmatrix} \qquad \boldsymbol{w} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

2 Compute the lengths $\|\boldsymbol{u}\|$ and $\|\boldsymbol{v}\|$ and $\|\boldsymbol{w}\|$ of those vectors. Check the Schwarz inequalities $|\boldsymbol{u} \cdot \boldsymbol{v}| \le \|\boldsymbol{u}\|\,\|\boldsymbol{v}\|$ and $|\boldsymbol{v} \cdot \boldsymbol{w}| \le \|\boldsymbol{v}\|\,\|\boldsymbol{w}\|$.

3 Find unit vectors in the directions of $\boldsymbol{v}$ and $\boldsymbol{w}$ in Problem 1, and the cosine of the angle θ. Choose vectors $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ that make $0°$, $90°$, and $180°$ angles with $\boldsymbol{w}$.

4 For any *unit* vectors $\boldsymbol{v}$ and $\boldsymbol{w}$, find the dot products (actual numbers) of

(a) $\boldsymbol{v}$ and $-\boldsymbol{v}$ (b) $\boldsymbol{v} + \boldsymbol{w}$ and $\boldsymbol{v} - \boldsymbol{w}$ (c) $\boldsymbol{v} - 2\boldsymbol{w}$ and $\boldsymbol{v} + 2\boldsymbol{w}$

5 Find unit vectors $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ in the directions of $\boldsymbol{v} = (1, 3)$ and $\boldsymbol{w} = (2, 1, 2)$. Find unit vectors $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ that are perpendicular to $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$.

6 (a) Describe every vector $\boldsymbol{w} = (w_1, w_2)$ that is perpendicular to $\boldsymbol{v} = (2, -1)$.

(b) All vectors perpendicular to $\boldsymbol{V} = (1, 1, 1)$ lie on a _____ in 3 dimensions.

(c) The vectors perpendicular to both $(1, 1, 1)$ and $(1, 2, 3)$ lie on a _____.

7 Find the angle θ (from its cosine) between these pairs of vectors:

(a) $\boldsymbol{v} = \begin{bmatrix} 1 \\ \sqrt{3} \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ (b) $\boldsymbol{v} = \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix}$

(c) $\boldsymbol{v} = \begin{bmatrix} 1 \\ \sqrt{3} \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} -1 \\ \sqrt{3} \end{bmatrix}$ (d) $\boldsymbol{v} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and $\boldsymbol{w} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$.

8 True or false (give a reason if true or find a counterexample if false):

(a) If $\boldsymbol{u} = (1, 1, 1)$ is perpendicular to $\boldsymbol{v}$ and $\boldsymbol{w}$, then $\boldsymbol{v}$ is parallel to $\boldsymbol{w}$.

(b) If $\boldsymbol{u}$ is perpendicular to $\boldsymbol{v}$ and $\boldsymbol{w}$, then $\boldsymbol{u}$ is perpendicular to $\boldsymbol{v} + 2\boldsymbol{w}$.

(c) If $\boldsymbol{u}$ and $\boldsymbol{v}$ are perpendicular unit vectors then $\|\boldsymbol{u} - \boldsymbol{v}\| = \sqrt{2}$. *Yes*!

9 The slopes of the arrows from $(0, 0)$ to (v_1, v_2) and (w_1, w_2) are v_2/v_1 and w_2/w_1. **Suppose the product v_2w_2/v_1w_1 of those slopes is -1.** Show that $\boldsymbol{v} \cdot \boldsymbol{w} = 0$ and the vectors are perpendicular. (The line $y = 4x$ is perpendicular to $y = -\frac{1}{4}x$.)

10 Draw arrows from $(0, 0)$ to the points $\boldsymbol{v} = (1, 2)$ and $\boldsymbol{w} = (-2, 1)$. Multiply their slopes. That answer is a signal that $\boldsymbol{v} \cdot \boldsymbol{w} = 0$ and the arrows are _____.

11 If $\boldsymbol{v} \cdot \boldsymbol{w}$ is negative, what does this say about the angle between $\boldsymbol{v}$ and $\boldsymbol{w}$? Draw a 3-dimensional vector $\boldsymbol{v}$ (an arrow), and show where to find all $\boldsymbol{w}$'s with $\boldsymbol{v} \cdot \boldsymbol{w} < 0$.

12 With $\boldsymbol{v} = (1, 1)$ and $\boldsymbol{w} = (1, 5)$ choose a number c so that $\boldsymbol{w} - c\boldsymbol{v}$ is perpendicular to $\boldsymbol{v}$. Then find the formula for c starting from *any* nonzero $\boldsymbol{v}$ and $\boldsymbol{w}$.

13 Find nonzero vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ that are perpendicular to $(1, 0, 1)$ and to each other.

14 Find nonzero vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$ that are perpendicular to $(1, 1, 1, 1)$ and to each other.

15 The geometric mean of $x = 2$ and $y = 8$ is $\sqrt{xy} = 4$. The arithmetic mean is larger: $\frac{1}{2}(x + y) =$ _____. This would come in Example 6 from the Schwarz inequality for $\boldsymbol{v} = (\sqrt{2}, \sqrt{8})$ and $\boldsymbol{w} = (\sqrt{8}, \sqrt{2})$. Find $\cos\theta$ for this $\boldsymbol{v}$ and $\boldsymbol{w}$.

16 **How long is the vector $\boldsymbol{v} = (1, 1, \ldots, 1)$ in 9 dimensions**? Find a unit vector $\boldsymbol{u}$ in the same direction as $\boldsymbol{v}$ and a unit vector $\boldsymbol{w}$ that is perpendicular to $\boldsymbol{v}$.

17 What are the cosines of the angles α, β, θ between the vector $(1, 0, -1)$ and the unit vectors $\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ along the axes? Check the formula $\cos^2\alpha + \cos^2\beta + \cos^2\theta = 1$.

Problems 18–28 lead to the main facts about lengths and angles in triangles.

18 The parallelogram with sides $\boldsymbol{v} = (4, 2)$ and $\boldsymbol{w} = (-1, 2)$ is a rectangle. Check the Pythagoras formula $a^2 + b^2 = c^2$ which is for ***right triangles only***:

$$(\text{length of } \boldsymbol{v})^2 + (\text{length of } \boldsymbol{w})^2 = (\text{length of } \boldsymbol{v} + \boldsymbol{w})^2.$$

19 (Rules for dot products) These equations are simple but useful:

(1) $\boldsymbol{v} \cdot \boldsymbol{w} = \boldsymbol{w} \cdot \boldsymbol{v}$ **(2)** $\boldsymbol{u} \cdot (\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{u} \cdot \boldsymbol{v} + \boldsymbol{u} \cdot \boldsymbol{w}$ **(3)** $(c\boldsymbol{v}) \cdot \boldsymbol{w} = c(\boldsymbol{v} \cdot \boldsymbol{w})$

Use **(2)** with $\boldsymbol{u} = \boldsymbol{v} + \boldsymbol{w}$ to prove $\|\boldsymbol{v} + \boldsymbol{w}\|^2 = \boldsymbol{v} \cdot \boldsymbol{v} + 2\boldsymbol{v} \cdot \boldsymbol{w} + \boldsymbol{w} \cdot \boldsymbol{w}$.

20 The "Law of Cosines" comes from $(\boldsymbol{v} - \boldsymbol{w}) \cdot (\boldsymbol{v} - \boldsymbol{w}) = \boldsymbol{v} \cdot \boldsymbol{v} - 2\boldsymbol{v} \cdot \boldsymbol{w} + \boldsymbol{w} \cdot \boldsymbol{w}$:

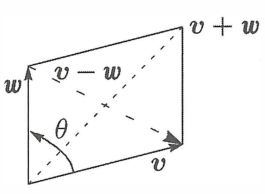
Cosine Law $\|\boldsymbol{v} - \boldsymbol{w}\|^2 = \|\boldsymbol{v}\|^2 - 2\|\boldsymbol{v}\|\,\|\boldsymbol{w}\| \cos\theta + \|\boldsymbol{w}\|^2.$

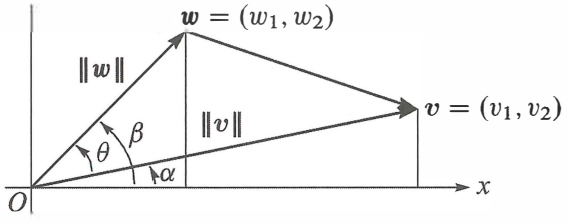
Draw a triangle with sides $\boldsymbol{v}$ and $\boldsymbol{w}$ and $\boldsymbol{v} - \boldsymbol{w}$. Which of the angles is θ ?

21 The ***triangle inequality*** says: (length of $\boldsymbol{v} + \boldsymbol{w}$) $\leq$ (length of $\boldsymbol{v}$) + (length of $\boldsymbol{w}$).

Problem 19 found $\|\boldsymbol{v} + \boldsymbol{w}\|^2 = \|\boldsymbol{v}\|^2 + 2\boldsymbol{v} \cdot \boldsymbol{w} + \|\boldsymbol{w}\|^2$. Increase that $\boldsymbol{v} \cdot \boldsymbol{w}$ to $\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|$ to show that $\|$**side 3**$\|$ can not exceed $\|$**side 1**$\| + \|$**side 2**$\|$:

Triangle inequality $\|\boldsymbol{v} + \boldsymbol{w}\|^2 \leq (\|\boldsymbol{v}\| + \|\boldsymbol{w}\|)^2$ **or** $\|\boldsymbol{v} + \boldsymbol{w}\| \leq \|\boldsymbol{v}\| + \|\boldsymbol{w}\|.$





22 The Schwarz inequality $|\boldsymbol{v} \cdot \boldsymbol{w}| \leq \|\boldsymbol{v}\|\,\|\boldsymbol{w}\|$ by algebra instead of trigonometry:

(a) Multiply out both sides of $(v_1w_1 + v_2w_2)^2 \leq (v_1^2 + v_2^2)(w_1^2 + w_2^2)$.

(b) Show that the difference between those two sides equals $(v_1w_2 - v_2w_1)^2$. This cannot be negative since it is a square—so the inequality is true.

23 The figure shows that $\cos\alpha = v_1/\|\boldsymbol{v}\|$ and $\sin\alpha = v_2/\|\boldsymbol{v}\|$. Similarly $\cos\beta$ is _____ and $\sin\beta$ is _____. The angle θ is $\beta - \alpha$. Substitute into the trigonometry formula $\cos\beta\cos\alpha + \sin\beta\sin\alpha$ for $\cos(\beta - \alpha)$ to find $\cos\theta = \boldsymbol{v} \cdot \boldsymbol{w}/\|\boldsymbol{v}\|\,\|\boldsymbol{w}\|$.

24 One-line proof of the inequality $|\boldsymbol{u} \cdot \boldsymbol{U}| \leq 1$ for unit vectors (u_1, u_2) and (U_1, U_2):

$$|\boldsymbol{u} \cdot \boldsymbol{U}| \leq |u_1|\,|U_1| + |u_2|\,|U_2| \leq \frac{u_1^2 + U_1^2}{2} + \frac{u_2^2 + U_2^2}{2} = 1.$$

Put $(u_1, u_2) = (.6, .8)$ and $(U_1, U_2) = (.8, .6)$ in that whole line and find $\cos\theta$.

25 Why is $|\cos\theta|$ never greater than 1 in the first place?

26 (*Recommended*) Draw a parallelogram

27 Parallelogram with two sides $\boldsymbol{v}$ and $\boldsymbol{w}$. Show that the squared diagonal lengths $\|\boldsymbol{v} + \boldsymbol{w}\|^2 + \|\boldsymbol{v} - \boldsymbol{w}\|^2$ add to the sum of four squared side lengths $2\|\boldsymbol{v}\|^2 + 2\|\boldsymbol{w}\|^2$.

28 If $\boldsymbol{v} = (1, 2)$ draw all vectors $\boldsymbol{w} = (x, y)$ in the xy plane with $\boldsymbol{v} \cdot \boldsymbol{w} = x + 2y = 5$. Why do those $\boldsymbol{w}$'s lie along a line? Which is the shortest $\boldsymbol{w}$?

29 (*Recommended*) If $||\boldsymbol{v}|| = 5$ and $||\boldsymbol{w}|| = 3$, what are the smallest and largest possible values of $||\boldsymbol{v} - \boldsymbol{w}||$? What are the smallest and largest possible values of $\boldsymbol{v} \cdot \boldsymbol{w}$?

Challenge Problems

30 Can three vectors in the xy plane have $\boldsymbol{u} \cdot \boldsymbol{v} < 0$ and $\boldsymbol{v} \cdot \boldsymbol{w} < 0$ and $\boldsymbol{u} \cdot \boldsymbol{w} < 0$? I don't know how many vectors in xyz space can have all negative dot products. (Four of those vectors in the plane would certainly be impossible . . .).

31 Pick any numbers that add to $x + y + z = 0$. Find the angle between your vector $\boldsymbol{v} = (x, y, z)$ and the vector $\boldsymbol{w} = (z, x, y)$. Challenge question: Explain why $\boldsymbol{v} \cdot \boldsymbol{w}/\|\boldsymbol{v}\|\|\boldsymbol{w}\|$ is always $-\frac{1}{2}$.

32 How could you prove $\sqrt[3]{xyz} \leq \frac{1}{3}(x+y+z)$ (geometric mean $\leq$ arithmetic mean) ?

33 Find 4 perpendicular unit vectors of the form $\left(\pm\frac{1}{2}, \pm\frac{1}{2}, \pm\frac{1}{2}, \pm\frac{1}{2}\right)$: Choose $+$ or $-$.

34 Using $\boldsymbol{v} = \mathsf{randn}(3, 1)$ in MATLAB, create a random unit vector $\boldsymbol{u} = \boldsymbol{v}/\|\boldsymbol{v}\|$. Using $V = \mathsf{randn}(3, 30)$ create 30 more random unit vectors $\boldsymbol{U}_j$. What is the average size of the dot products $|\boldsymbol{u} \cdot \boldsymbol{U}_j|$? In calculus, the average is $\int_0^\pi |\cos\theta| d\theta/\pi = 2/\pi$.

1.1 Systems of Linear Equations

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

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b \tag{1}$$

where b and the **coefficients** $a_1, \ldots, a_n$ are real or complex numbers, usually known in advance. The subscript n may be any positive integer. In textbook examples and exercises, n is normally between 2 and 5. In real-life problems, n might be 50 or 5000, or even larger.

The equations

$$4x_1 - 5x_2 + 2 = x_1 \quad \text{and} \quad x_2 = 2(\sqrt{6} - x_1) + x_3$$

are both linear because they can be rearranged algebraically as in equation (1):

$$3x_1 - 5x_2 = -2 \quad \text{and} \quad 2x_1 + x_2 - x_3 = 2\sqrt{6}$$

The equations

$$4x_1 - 5x_2 = x_1x_2 \quad \text{and} \quad x_2 = 2\sqrt{x_1} - 6$$

are not linear because of the presence of x_1x_2 in the first equation and $\sqrt{x_1}$ in the second.

A **system of linear equations** (or a **linear system**) is a collection of one or more linear equations involving the same variables—say, $x_1, \ldots, x_n$. An example is

$$\begin{aligned} 2x_1 - x_2 + 1.5x_3 &= 8 \\ x_1 \qquad - \quad 4x_3 &= -7 \end{aligned} \tag{2}$$

A **solution** of the system is a list $(s_1, s_2, \ldots, s_n)$ of numbers that makes each equation a true statement when the values $s_1, \ldots, s_n$ are substituted for $x_1, \ldots, x_n$, respectively. For instance, $(5, 6.5, 3)$ is a solution of system (2) because, when these values are substituted in (2) for x_1, x_2, x_3, respectively, the equations simplify to $8 = 8$ and $-7 = -7$.

The set of all possible solutions is called the **solution set** of the linear system. Two linear systems are called **equivalent** if they have the same solution set. That is, each solution of the first system is a solution of the second system, and each solution of the second system is a solution of the first.

Finding the solution set of a system of two linear equations in two variables is easy because it amounts to finding the intersection of two lines. A typical problem is

$$\begin{aligned} x_1 - 2x_2 &= -1 \\ -x_1 + 3x_2 &= 3 \end{aligned}$$

The graphs of these equations are lines, which we denote by ℓ_1 and ℓ_2. A pair of numbers (x_1, x_2) satisfies *both* equations in the system if and only if the point (x_1, x_2) lies on both ℓ_1 and ℓ_2. In the system above, the solution is the single point $(3, 2)$, as you can easily verify. See Figure 1.

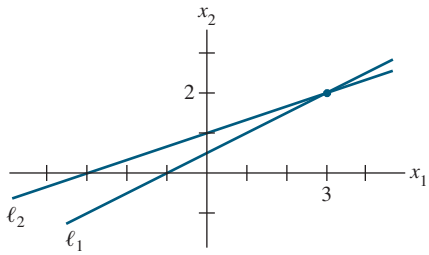


FIGURE 1 Exactly one solution.

Of course, two lines need not intersect in a single point—they could be parallel, or they could coincide and hence "intersect" at every point on the line. Figure 2 shows the graphs that correspond to the following systems:

$$\text{(a)} \quad \begin{aligned} x_1 - 2x_2 &= -1 \\ -x_1 + 2x_2 &= 3 \end{aligned} \qquad \text{(b)} \quad \begin{aligned} x_1 - 2x_2 &= -1 \\ -x_1 + 2x_2 &= 1 \end{aligned}$$

Figures 1 and 2 illustrate the following general fact about linear systems, to be verified in Section 1.2.

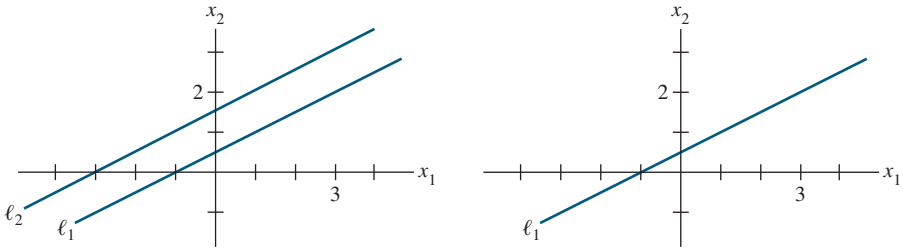


FIGURE 2 (a) No solution. (b) Infinitely many solutions.

A system of linear equations has

1. no solution, or
2. exactly one solution, or
3. infinitely many solutions.

A system of linear equations is said to be **consistent** if it has either one solution or infinitely many solutions; a system is **inconsistent** if it has no solution.

Matrix Notation

The essential information of a linear system can be recorded compactly in a rectangular array called a **matrix**. Given the system

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ 2x_2 - 8x_3 &= 8 \\ 5x_1 \quad - 5x_3 &= 10 \end{aligned} \tag{3}$$

with the coefficients of each variable aligned in columns, the matrix

$$\begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & -8 \\ 5 & 0 & -5 \end{bmatrix}$$

is called the **coefficient matrix** (or **matrix of coefficients**) of the system (3), and the matrix

$$\begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix} \tag{4}$$

is called the **augmented matrix** of the system. (The second row here contains a zero because the second equation could be written as $0 \cdot x_1 + 2x_2 - 8x_3 = 8$.) An augmented matrix of a system consists of the coefficient matrix with an added column containing the constants from the respective right sides of the equations.

The **size** of a matrix tells how many rows and columns it has. The augmented matrix (4) above has 3 rows and 4 columns and is called a 3×4 (read "3 by 4") matrix. If m and n are positive integers, an $\boldsymbol{m \times n}$ **matrix** is a rectangular array of numbers with m rows and n columns. (The number of rows always comes first.) Matrix notation will simplify the calculations in the examples that follow.

Solving a Linear System

This section and the next describe an algorithm, or a systematic procedure, for solving linear systems. The basic strategy is *to replace one system with an equivalent system (that is one with the same solution set) that is easier to solve*.

Roughly speaking, use the x_1 term in the first equation of a system to eliminate the x_1 terms in the other equations. Then use the x_2 term in the second equation to eliminate the x_2 terms in the other equations, and so on, until you finally obtain a very simple equivalent system of equations.

Three basic operations are used to simplify a linear system: Replace one equation by the sum of itself and a multiple of another equation, interchange two equations, and multiply all the terms in an equation by a nonzero constant. After the first example, you will see why these three operations do not change the solution set of the system.

EXAMPLE 1 Solve system (3).

SOLUTION The elimination procedure is shown here with and without matrix notation, and the results are placed side by side for comparison:

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ 2x_2 - 8x_3 &= 8 \\ 5x_1 - 5x_3 &= 10 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 5 & 0 & -5 & 10 \end{bmatrix}$$

Keep x_1 in the first equation and eliminate it from the other equations. To do so, add -5 times equation 1 to equation 3. After some practice, this type of calculation is usually performed mentally:

$$\begin{array}{r} -5 \cdot [\text{equation 1}] \\ +\,[\text{equation 3}] \\ \hline [\text{new equation 3}] \end{array} \qquad \begin{array}{r} -5x_1 + 10x_2 - 5x_3 = 0 \\ 5x_1 \qquad\quad - 5x_3 = 10 \\ \hline 10x_2 - 10x_3 = 10 \end{array}$$

The result of this calculation is written in place of the original third equation:

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ 2x_2 - 8x_3 &= 8 \\ 10x_2 - 10x_3 &= 10 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & 10 & -10 & 10 \end{bmatrix}$$

Now, multiply equation 2 by $\frac{1}{2}$ in order to obtain 1 as the coefficient for x_2. (This calculation will simplify the arithmetic in the next step.)

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ x_2 - 4x_3 &= 4 \\ 10x_2 - 10x_3 &= 10 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 10 & -10 & 10 \end{bmatrix}$$

Use the x_2 in equation 2 to eliminate the $10x_2$ in equation 3. The "mental" computation is

$$\begin{array}{r} -10 \cdot [\text{equation 2}] \\ +\,[\text{equation 3}] \\ \hline [\text{new equation 3}] \end{array} \qquad \begin{array}{r} -10x_2 + 40x_3 = -40 \\ 10x_2 - 10x_3 = 10 \\ \hline 30x_3 = -30 \end{array}$$

The result of this calculation is written in place of the previous third equation (row):

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ x_2 - 4x_3 &= 4 \\ 30x_3 &= -30 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 30 & -30 \end{bmatrix}$$

Now, multiply equation 3 by $\frac{1}{30}$ in order to obtain 1 as the coefficient for x_3. (This calculation will simplify the arithmetic in the next step.)

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ x_2 - 4x_3 &= 4 \\ x_3 &= -1 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

The new system has a *triangular* form (the intuitive term *triangular* will be replaced by a precise term in the next section):

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ x_2 - 4x_3 &= 4 \\ x_3 &= -1 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

Eventually, you want to eliminate the $-2x_2$ term from equation 1, but it is more efficient to use the x_3 in equation 3 first, to eliminate the $-4x_3$ and $+x_3$ terms in equations 2 and 1. The two "mental" calculations are

$$\begin{array}{r} 4 \cdot [\text{equation 3}] \\ +\,[\text{equation 2}] \\ \hline [\text{new equation 2}] \end{array} \quad \begin{array}{r} 4x_3 = -4 \\ x_2 - 4x_3 = 4 \\ \hline x_2 \qquad = 0 \end{array} \qquad \begin{array}{r} -1 \cdot [\text{equation 3}] \\ +\,[\text{equation 1}] \\ \hline [\text{new equation 1}] \end{array} \quad \begin{array}{r} -x_3 = 1 \\ x_1 - 2x_2 + x_3 = 0 \\ \hline x_1 - 2x_2 \qquad = 1 \end{array}$$

It is convenient to combine the results of these two operations:

$$\begin{aligned} x_1 - 2x_2 &= 1 \\ x_2 &= 0 \\ x_3 &= -1 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

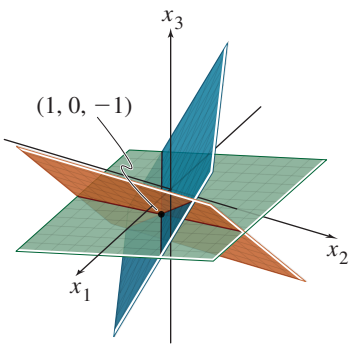
Now, having cleaned out the column above the x_3 in equation 3, move back to the x_2 in equation 2 and use it to eliminate the $-2x_2$ above it. Because of the previous work with x_3, there is now no arithmetic involving x_3 terms. Add 2 times equation 2 to equation 1 and obtain the system:

$$\begin{aligned} x_1 &= 1 \\ x_2 &= 0 \\ x_3 &= -1 \end{aligned} \qquad \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

The work is essentially done. It shows that the only solution of the original system is $(1, 0, -1)$. However, since there are so many calculations involved, it is a good practice to check the work. To verify that $(1, 0, -1)$ *is* a solution, substitute these values into the left side of the original system, and compute:

$$\begin{aligned} 1(1) - 2(0) + 1(-1) &= 1 - 0 - 1 = 0 \\ 2(0) - 8(-1) &= 0 + 8 = 8 \\ 5(1) - 5(-1) &= 5 + 5 = 10 \end{aligned}$$

The results agree with the right side of the original system, so $(1, 0, -1)$ is a solution of the system. ■



Each of the original equations determines a plane in three-dimensional space. The point $(1, 0, -1)$ lies in all three planes.

Example 1 illustrates how operations on equations in a linear system correspond to operations on the appropriate rows of the augmented matrix. The three basic operations listed earlier correspond to the following operations on the augmented matrix.

ELEMENTARY ROW OPERATIONS

1. (Replacement) Replace one row by the sum of itself and a multiple of another row.[1]
2. (Interchange) Interchange two rows.
3. (Scaling) Multiply all entries in a row by a nonzero constant.

[1] A common paraphrase of row replacement is "Add to one row a multiple of another row."

Row operations can be applied to any matrix, not merely to one that arises as the augmented matrix of a linear system. Two matrices are called **row equivalent** if there is a sequence of elementary row operations that transforms one matrix into the other.

It is important to note that row operations are *reversible*. If two rows are interchanged, they can be returned to their original positions by another interchange. If a row is scaled by a nonzero constant c, then multiplying the new row by $1/c$ produces the original row. Finally, consider a replacement operation involving two rows—say, rows 1 and 2—and suppose that c times row 1 is added to row 2 to produce a new row 2. To "reverse" this operation, add $-c$ times row 1 to (new) row 2 and obtain the original row 2. See Exercises 39–42 at the end of this section.

At the moment, we are interested in row operations on the augmented matrix of a system of linear equations. Suppose a system is changed to a new one via row operations. By considering each type of row operation, you can see that any solution of the original system remains a solution of the new system. Conversely, since the original system can be produced via row operations on the new system, each solution of the new system is also a solution of the original system. This discussion justifies the following statement.

If the augmented matrices of two linear systems are row equivalent, then the two systems have the same solution set.

Though Example 1 is lengthy, you will find that after some practice, the calculations go quickly. Row operations in the text and exercises will usually be extremely easy to perform, allowing you to focus on the underlying concepts. Still, you must learn to perform row operations accurately because they will be used throughout the text.

The rest of this section shows how to use row operations to determine the size of a solution set, without completely solving the linear system.

Existence and Uniqueness Questions

Section 1.2 will show why a solution set for a linear system contains either no solutions, one solution, or infinitely many solutions. Answers to the following two questions will determine the nature of the solution set for a linear system.

To determine which possibility is true for a particular system, we ask two questions.

TWO FUNDAMENTAL QUESTIONS ABOUT A LINEAR SYSTEM

1. Is the system consistent; that is, does at least one solution *exist*?
2. If a solution exists, is it the *only* one; that is, is the solution *unique*?

These two questions will appear throughout the text, in many different guises. This section and the next will show how to answer these questions via row operations on the augmented matrix.

EXAMPLE 2 Determine if the following system is consistent:

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ 2x_2 - 8x_3 &= 8 \\ 5x_1 - 5x_3 &= 10 \end{aligned}$$

SOLUTION This is the system from Example 1. Suppose that we have performed the row operations necessary to obtain the triangular form

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 0 \\ x_2 - 4x_3 &= 4 \\ x_3 &= -1 \end{aligned} \qquad \begin{bmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

At this point, we know x_3. Were we to substitute the value of x_3 into equation 2, we could compute x_2 and hence could determine x_1 from equation 1. So a solution exists; the system is consistent. (In fact, x_2 is uniquely determined by equation 2 since x_3 has only one possible value, and x_1 is therefore uniquely determined by equation 1. So the solution is unique.) ■

EXAMPLE 3 Determine if the following system is consistent:

$$\begin{aligned} x_2 - 4x_3 &= 8 \\ 2x_1 - 3x_2 + 2x_3 &= 1 \\ 4x_1 - 8x_2 + 12x_3 &= 1 \end{aligned} \tag{5}$$

SOLUTION The augmented matrix is

$$\begin{bmatrix} 0 & 1 & -4 & 8 \\ 2 & -3 & 2 & 1 \\ 4 & -8 & 12 & 1 \end{bmatrix}$$

To obtain an x_1 in the first equation, interchange rows 1 and 2:

$$\begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 4 & -8 & 12 & 1 \end{bmatrix}$$

To eliminate the $4x_1$ term in the third equation, add -2 times row 1 to row 3:

$$\begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & -2 & 8 & -1 \end{bmatrix} \tag{6}$$

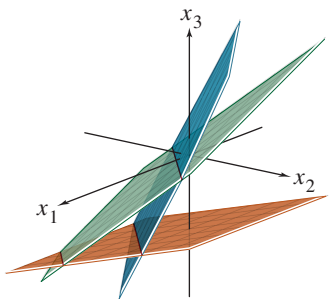
Next, use the x_2 term in the second equation to eliminate the $-2x_2$ term from the third equation. Add 2 times row 2 to row 3:

$$\begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & 15 \end{bmatrix} \tag{7}$$

The augmented matrix is now in triangular form. To interpret it correctly, go back to equation notation:

$$\begin{aligned} 2x_1 - 3x_2 + 2x_3 &= 1 \\ x_2 - 4x_3 &= 8 \\ 0 &= 15 \end{aligned} \tag{8}$$

The equation $0 = 15$ is a short form of $0x_1 + 0x_2 + 0x_3 = 15$. This system in triangular form obviously has a built-in contradiction. There are no values of x_1, x_2, x_3 that satisfy (8) because the equation $0 = 15$ is never true. Since (8) and (5) have the same solution set, the original system is inconsistent (it has no solution). ■



The system is inconsistent because there is no point that lies on all three planes.

Pay close attention to the augmented matrix in (7). Its last row is typical of an inconsistent system in triangular form.

Reasonable Answers

Once you have one or more solutions to a system of equations, remember to check your answer by substituting the solution you found back into the original equation. For example, if you found $(2, 1, -1)$ was a solution to the system of equations

$$\begin{aligned} x_1 - 2x_2 + x_3 &= 2 \\ x_1 - 2x_3 &= -2 \\ x_2 + x_3 &= 3 \end{aligned}$$

you could substitute your solution into the original equations to get

$$\begin{aligned} 2 - 2(1) + (-1) &= -1 \neq 2 \\ 2 - 2(-1) &= 4 \neq -2 \\ 1 + (-1) &= 0 \neq 3 \end{aligned}$$

It is now clear that there must have been an error in your original calculations. If upon rechecking your arithmetic, you find the answer $(2, 1, 2)$, you can see that

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

and you can now be confident you have a correct solution to the given system of equations.

Numerical Note

In real-world problems, systems of linear equations are solved by a computer. For a square coefficient matrix, computer programs nearly always use the elimination algorithm given here and in Section 1.2, modified slightly for improved accuracy.

The vast majority of linear algebra problems in business and industry are solved with programs that use *floating point arithmetic*. Numbers are represented as decimals $\pm.d_1 \cdots d_p \times 10^r$, where r is an integer and the number p of digits to the right of the decimal point is usually between 8 and 16. Arithmetic with such numbers typically is inexact, because the result must be rounded (or truncated) to the number of digits stored. "Roundoff error" is also introduced when a number such as $1/3$ is entered into the computer, since its decimal representation must be approximated by a finite number of digits. Fortunately, inaccuracies in floating point arithmetic seldom cause problems. The numerical notes in this book will occasionally warn of issues that you may need to consider later in your career.

Practice Problems

Throughout the text, practice problems should be attempted before working the exercises. Solutions appear after each exercise set.

1. State in words the next elementary row operation that should be performed on the system in order to solve it. [More than one answer is possible in (a).]

Practice Problems (Continued)

a. $$\begin{aligned} x_1 + 4x_2 - 2x_3 + 8x_4 &= 12 \\ x_2 - 7x_3 + 2x_4 &= -4 \\ 5x_3 - x_4 &= 7 \\ x_3 + 3x_4 &= -5 \end{aligned}$$

b. $$\begin{aligned} x_1 - 3x_2 + 5x_3 - 2x_4 &= 0 \\ x_2 + 8x_3 &= -4 \\ 2x_3 &= 3 \\ x_4 &= 1 \end{aligned}$$

2. The augmented matrix of a linear system has been transformed by row operations into the form below. Determine if the system is consistent.

$$\begin{bmatrix} 1 & 5 & 2 & -6 \\ 0 & 4 & -7 & 2 \\ 0 & 0 & 5 & 0 \end{bmatrix}$$

3. Is $(3, 4, -2)$ a solution of the following system?

$$\begin{aligned} 5x_1 - x_2 + 2x_3 &= 7 \\ -2x_1 + 6x_2 + 9x_3 &= 0 \\ -7x_1 + 5x_2 - 3x_3 &= -7 \end{aligned}$$

4. For what values of h and k is the following system consistent?

$$\begin{aligned} 2x_1 - x_2 &= h \\ -6x_1 + 3x_2 &= k \end{aligned}$$

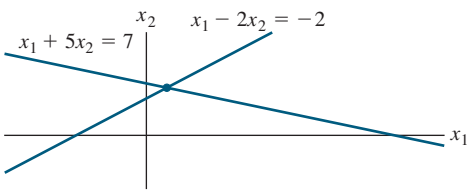
1.1 Exercises

Solve each system in Exercises 1–4 by using elementary row operations on the equations or on the augmented matrix. Follow the systematic elimination procedure described in this section.

1. $$\begin{aligned} x_1 + 5x_2 &= 7 \\ -2x_1 - 7x_2 &= -5 \end{aligned}$$

2. $$\begin{aligned} 2x_1 + 4x_2 &= -4 \\ 5x_1 + 7x_2 &= 11 \end{aligned}$$

3. Find the point (x_1, x_2) that lies on the line $x_1 + 5x_2 = 7$ and on the line $x_1 - 2x_2 = -2$. See the figure.



4. Find the point of intersection of the lines $x_1 - 5x_2 = 1$ and $3x_1 - 7x_2 = 5$.

Consider each matrix in Exercises 5 and 6 as the augmented matrix of a linear system. State in words the next two elementary row operations that should be performed in the process of solving the system.

5. $$\begin{bmatrix} 1 & 3 & -4 & 0 & 9 \\ 1 & 1 & 5 & 0 & -8 \\ 0 & 0 & 1 & 0 & 7 \\ 0 & 0 & 0 & 1 & -6 \end{bmatrix}$$

6. $$\begin{bmatrix} 1 & -6 & 4 & 0 & -1 \\ 0 & 2 & -7 & 0 & 4 \\ 0 & 0 & 1 & 2 & -3 \\ 0 & 0 & 3 & 1 & 6 \end{bmatrix}$$

In Exercises 7–10, the augmented matrix of a linear system has been reduced by row operations to the form shown. In each case, continue the appropriate row operations and describe the solution set of the original system.

7. $$\begin{bmatrix} 1 & 7 & 3 & -4 \\ 0 & 1 & -1 & 3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{bmatrix}$$

8. $$\begin{bmatrix} 1 & 1 & 5 & 0 \\ 0 & 1 & 9 & 0 \\ 0 & 0 & 7 & -7 \end{bmatrix}$$

9. $$\begin{bmatrix} 1 & -1 & 0 & 0 & -4 \\ 0 & 1 & -3 & 0 & -7 \\ 0 & 0 & 1 & -3 & -1 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

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

Solve the systems in Exercises 11–14.

11.
$$\begin{aligned} x_2 + 4x_3 &= -4 \\ x_1 + 3x_2 + 3x_3 &= -2 \\ 3x_1 + 7x_2 + 5x_3 &= 6 \end{aligned}$$

12.
$$\begin{aligned} x_1 - 3x_2 + 4x_3 &= -4 \\ 3x_1 - 7x_2 + 7x_3 &= -8 \\ -4x_1 + 6x_2 + 2x_3 &= 4 \end{aligned}$$

13.
$$\begin{aligned} x_1 - 3x_3 &= 8 \\ 2x_1 + 2x_2 + 9x_3 &= 7 \\ x_2 + 5x_3 &= -2 \end{aligned}$$

14.
$$\begin{aligned} x_1 - 3x_2 &= 5 \\ -x_1 + x_2 + 5x_3 &= 2 \\ x_2 + x_3 &= 0 \end{aligned}$$

15. Verify that the solution you found to Exercise 11 is correct by substituting the values you obtained back into the original equations.

16. Verify that the solution you found to Exercise 12 is correct by substituting the values you obtained back into the original equations.

17. Verify that the solution you found to Exercise 13 is correct by substituting the values you obtained back into the original equations.

18. Verify that the solution you found to Exercise 14 is correct by substituting the values you obtained back into the original equations.

Determine if the systems in Exercises 19 and 20 are consistent. Do not completely solve the systems.

19.
$$\begin{aligned} x_1 + 3x_3 &= 2 \\ x_2 - 3x_4 &= 3 \\ -2x_2 + 3x_3 + 2x_4 &= 1 \\ 3x_1 + 7x_4 &= -5 \end{aligned}$$

20.
$$\begin{aligned} x_1 - 2x_4 &= -3 \\ 2x_2 + 2x_3 &= 0 \\ x_3 + 3x_4 &= 1 \\ -2x_1 + 3x_2 + 2x_3 + x_4 &= 5 \end{aligned}$$

21. Do the three lines $x_1 - 4x_2 = 1$, $2x_1 - x_2 = -3$, and $-x_1 - 3x_2 = 4$ have a common point of intersection? Explain.

22. Do the three planes $x_1 + 2x_2 + x_3 = 4$, $x_2 - x_3 = 1$, and $x_1 + 3x_2 = 0$ have at least one common point of intersection? Explain.

In Exercises 23–26, determine the value(s) of h such that the matrix is the augmented matrix of a consistent linear system.

23. $\begin{bmatrix} 1 & h & 4 \\ 3 & 6 & 8 \end{bmatrix}$ **24.** $\begin{bmatrix} 1 & h & -3 \\ -2 & 4 & 6 \end{bmatrix}$

25. $\begin{bmatrix} 1 & 3 & -2 \\ -4 & h & 8 \end{bmatrix}$ **26.** $\begin{bmatrix} 3 & -4 & h \\ -6 & 8 & 9 \end{bmatrix}$

In Exercises 27–34, key statements from this section are either quoted directly, restated slightly (but still true), or altered in some way that makes them false in some cases. Mark each statement True or False, and *justify* your answer. (If true, give the approximate location where a similar statement appears, or refer to a definition or theorem. If false, give the location of a statement that has been quoted or used incorrectly, or cite an example that shows the statement is not true in all cases.) Similar true/false questions will appear in many sections of the text and will be flagged with a **(T/F)** at the beginning of the question.

27. **(T/F)** Every elementary row operation is reversible.

28. **(T/F)** Elementary row operations on an augmented matrix never change the solution set of the associated linear system.

29. **(T/F)** A 5×6 matrix has six rows.

30. **(T/F)** Two matrices are row equivalent if they have the same number of rows.

31. **(T/F)** The solution set of a linear system involving variables $x_1, \ldots, x_n$ is a list of numbers $(s_1, \ldots, s_n)$ that makes each equation in the system a true statement when the values $s_1, \ldots, s_n$ are substituted for $x_1, \ldots, x_n$, respectively.

32. **(T/F)** An inconsistent system has more than one solution.

33. **(T/F)** Two fundamental questions about a linear system involve existence and uniqueness.

34. **(T/F)** Two linear systems are equivalent if they have the same solution set.

35. Find an equation involving g, h, and k that makes this augmented matrix correspond to a consistent system:

$$\begin{bmatrix} 1 & -3 & 5 & g \\ 0 & 2 & -3 & h \\ -3 & 5 & -9 & k \end{bmatrix}$$

36. Construct three different augmented matrices for linear systems whose solution set is $x_1 = -2$, $x_2 = 1$, $x_3 = 0$.

37. Suppose the system below is consistent for all possible values of f and g. What can you say about the coefficients c and d? Justify your answer.

$$\begin{aligned} x_1 + 5x_2 &= f \\ cx_1 + dx_2 &= g \end{aligned}$$

38. Suppose a, b, c, and d are constants such that a is not zero and the system below is consistent for all possible values of f and g. What can you say about the numbers a, b, c, and d? Justify your answer.

$$ax_1 + bx_2 = f$$
$$cx_1 + dx_2 = g$$

In Exercises 39–42, find the elementary row operation that transforms the first matrix into the second, and then find the reverse row operation that transforms the second matrix into the first.

39. $\begin{bmatrix} 0 & -2 & 5 \\ 1 & 4 & -7 \\ 3 & -1 & 6 \end{bmatrix}, \begin{bmatrix} 1 & 4 & -7 \\ 0 & -2 & 5 \\ 3 & -1 & 6 \end{bmatrix}$

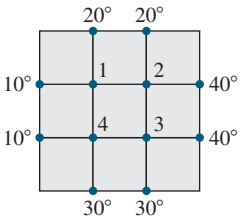
40. $\begin{bmatrix} 1 & 3 & -4 \\ 0 & -2 & 6 \\ 0 & -5 & 9 \end{bmatrix}, \begin{bmatrix} 1 & 3 & -4 \\ 0 & 1 & -3 \\ 0 & -5 & 9 \end{bmatrix}$

41. $\begin{bmatrix} 1 & -3 & 2 & 0 \\ 0 & 4 & -5 & 6 \\ 5 & -7 & 8 & -9 \end{bmatrix}, \begin{bmatrix} 1 & -3 & 2 & 0 \\ 0 & 4 & -5 & 6 \\ 0 & 8 & -2 & -9 \end{bmatrix}$

42. $\begin{bmatrix} 1 & 2 & -5 & 0 \\ 0 & 1 & -3 & -2 \\ 0 & -3 & 9 & 5 \end{bmatrix}, \begin{bmatrix} 1 & 2 & -5 & 0 \\ 0 & 1 & -3 & -2 \\ 0 & 0 & 0 & -1 \end{bmatrix}$

An important concern in the study of heat transfer is to determine the steady-state temperature distribution of a thin plate when the temperature around the boundary is known. Assume the plate shown in the figure represents a cross section of a metal beam, with negligible heat flow in the direction perpendicular to the plate. Let $T_1, \ldots, T_4$ denote the temperatures at the four interior nodes of the mesh in the figure. The temperature at a node is approximately equal to the average of the four nearest nodes—to the left, above, to the right, and below.[2] For instance,

$$T_1 = (10 + 20 + T_2 + T_4)/4, \quad \text{or} \quad 4T_1 - T_2 - T_4 = 30$$



43. Write a system of four equations whose solution gives estimates for the temperatures $T_1, \ldots, T_4$.

44. Solve the system of equations from Exercise 43. [*Hint:* To speed up the calculations, interchange rows 1 and 4 before starting "replace" operations.]

[2] See Frank M. White, *Heat and Mass Transfer* (Reading, MA: Addison-Wesley Publishing, 1991), pp. 145–149.

Solutions to Practice Problems

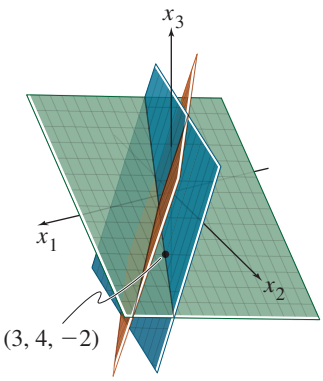
1. a. For "hand computation," the best choice is to interchange equations 3 and 4. Another possibility is to multiply equation 3 by $1/5$. Or, replace equation 4 by its sum with $-1/5$ times row 3. (In any case, do not use the x_2 in equation 2 to eliminate the $4x_2$ in equation 1. Wait until a triangular form has been reached and the x_3 terms and x_4 terms have been eliminated from the first two equations.)

b. The system is in triangular form. Further simplification begins with the x_4 in the fourth equation. Use the x_4 to eliminate all x_4 terms above it. The appropriate step now is to add 2 times equation 4 to equation 1. (After that, move to equation 3, multiply it by $1/2$, and then use the equation to eliminate the x_3 terms above it.)

2. The system corresponding to the augmented matrix is

$$\begin{aligned} x_1 + 5x_2 + 2x_3 &= -6 \\ 4x_2 - 7x_3 &= 2 \\ 5x_3 &= 0 \end{aligned}$$

The third equation makes $x_3 = 0$, which is certainly an allowable value for x_3. After eliminating the x_3 terms in equations 1 and 2, you could go on to solve for unique values for x_2 and x_1. Hence a solution exists, and it is unique. Contrast this situation with that in Example 3.



Since $(3, 4, -2)$ satisfies the first two equations, it is on the line of the intersection of the first two planes. Since $(3, 4, -2)$ does not satisfy all three equations, it does not lie on all three planes.

3. It is easy to check if a specific list of numbers is a solution. Set $x_1 = 3$, $x_2 = 4$, and $x_3 = -2$, and find that

$$\begin{aligned} 5(3) - (4) + 2(-2) &= 15 - 4 - 4 = 7 \\ -2(3) + 6(4) + 9(-2) &= -6 + 24 - 18 = 0 \\ -7(3) + 5(4) - 3(-2) &= -21 + 20 + 6 = 5 \end{aligned}$$

Although the first two equations are satisfied, the third is not, so $(3, 4, -2)$ is not a solution of the system. Notice the use of parentheses when making the substitutions. They are strongly recommended as a guard against arithmetic errors.

4. When the second equation is replaced by its sum with 3 times the first equation, the system becomes

$$\begin{aligned} 2x_1 - x_2 &= h \\ 0 &= k + 3h \end{aligned}$$

If $k + 3h$ is nonzero, the system has no solution. The system is consistent for any values of h and k that make $k + 3h = 0$.

1.2 Row Reduction and Echelon Forms

This section refines the method of Section 1.1 into a row reduction algorithm that will enable us to analyze any system of linear equations.[1] By using only the first part of the algorithm, we will be able to answer the fundamental existence and uniqueness questions posed in Section 1.1.

The algorithm applies to any matrix, whether or not the matrix is viewed as an augmented matrix for a linear system. So the first part of this section concerns an arbitrary rectangular matrix and begins by introducing two important classes of matrices that include the "triangular" matrices of Section 1.1. In the definitions that follow, a *nonzero* row or column in a matrix means a row or column that contains at least one nonzero entry; a **leading entry** of a row refers to the leftmost nonzero entry (in a nonzero row).

DEFINITION

A rectangular matrix is in **echelon form** (or **row echelon form**) if it has the following three properties:

1. All nonzero rows are above any rows of all zeros.
2. Each leading entry of a row is in a column to the right of the leading entry of the row above it.
3. All entries in a column below a leading entry are zeros.

If a matrix in echelon form satisfies the following additional conditions, then it is in **reduced echelon form** (or **reduced row echelon form**):

4. The leading entry in each nonzero row is 1.
5. Each leading 1 is the only nonzero entry in its column.

[1] The algorithm here is a variant of what is commonly called *Gaussian elimination*. A similar elimination method for linear systems was used by Chinese mathematicians in about 250 B.C. The process was unknown in Western culture until the nineteenth century, when a famous German mathematician, Carl Friedrich Gauss, discovered it. A German engineer, Wilhelm Jordan, popularized the algorithm in an 1888 text on geodesy.

An **echelon matrix** (respectively, **reduced echelon matrix**) is one that is in echelon form (respectively, reduced echelon form). Property 2 says that the leading entries form an *echelon* ("steplike") pattern that moves down and to the right through the matrix. Property 3 is a simple consequence of property 2, but we include it for emphasis.

The "triangular" matrices of Section 1.1, such as

$$\begin{bmatrix} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & 5/2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 & 0 & 29 \\ 0 & 1 & 0 & 16 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

are in echelon form. In fact, the second matrix is in reduced echelon form. Here are additional examples.

EXAMPLE 1 The following matrices are in echelon form. The leading entries (■) may have any nonzero value; the starred entries (∗) may have any value (including zero).

$$\begin{bmatrix} ■ & * & * & * \\ 0 & ■ & * & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & ■ & * & * & * & * & * & * & * & * \\ 0 & 0 & 0 & ■ & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & ■ & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & ■ & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & ■ & * \end{bmatrix}$$

The following matrices are in reduced echelon form because the leading entries are 1's, and there are 0's below *and above* each leading 1.

$$\begin{bmatrix} 1 & 0 & * & * \\ 0 & 1 & * & * \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & * & 0 & 0 & 0 & * & * & 0 & * \\ 0 & 0 & 0 & 1 & 0 & 0 & * & * & 0 & * \\ 0 & 0 & 0 & 0 & 1 & 0 & * & * & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 1 & * & * & 0 & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & * \end{bmatrix}$$

■

Any nonzero matrix may be **row reduced** (that is, transformed by elementary row operations) into more than one matrix in echelon form, using different sequences of row operations. However, the reduced echelon form one obtains from a matrix is unique. The following theorem is proved in Appendix A at the end of the text.

THEOREM 1

Uniqueness of the Reduced Echelon Form

Each matrix is row equivalent to one and only one reduced echelon matrix.

If a matrix A is row equivalent to an echelon matrix U, we call U **an echelon form** (or row echelon form) **of** A; if U is in reduced echelon form, we call U **the reduced echelon form of** A. [Most matrix programs and calculators with matrix capabilities use the abbreviation RREF for reduced (row) echelon form. Some use REF for (row) echelon form.]

Pivot Positions

When row operations on a matrix produce an echelon form, further row operations to obtain the reduced echelon form do not change the positions of the leading entries. Since the reduced echelon form is unique, *the leading entries are always in the same positions*

in any echelon form obtained from a given matrix. These leading entries correspond to leading 1's in the reduced echelon form.

DEFINITION

A **pivot position** in a matrix A is a location in A that corresponds to a leading 1 in the reduced echelon form of A. A **pivot column** is a column of A that contains a pivot position.

In Example 1, the squares (■) identify the pivot positions. Many fundamental concepts in the first four chapters will be connected in one way or another with pivot positions in a matrix.

EXAMPLE 2 Row reduce the matrix A below to echelon form, and locate the pivot columns of A.

$$A = \begin{bmatrix} 0 & -3 & -6 & 4 & 9 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 1 & 4 & 5 & -9 & -7 \end{bmatrix}$$

SOLUTION Use the same basic strategy as in Section 1.1. The top of the leftmost nonzero column is the first pivot position. A nonzero entry, or *pivot*, must be placed in this position. A good choice is to interchange rows 1 and 4 (because the mental computations in the next step will not involve fractions).

$$\begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix}$$

Pivot

Pivot column

Create zeros below the pivot, 1, by adding multiples of the first row to the rows below, and obtain matrix (1) below. The pivot position in the second row must be as far left as possible—namely in the second column. Choose the 2 in this position as the next pivot.

$$\begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 5 & 10 & -15 & -15 \\ 0 & -3 & -6 & 4 & 9 \end{bmatrix} \tag{1}$$

Pivot

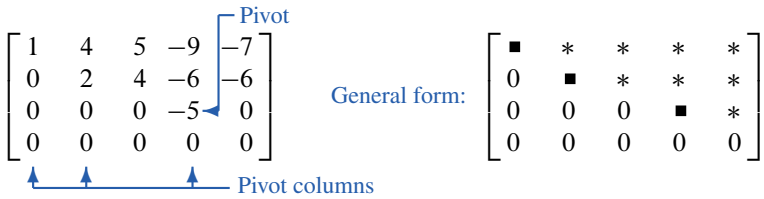
Next pivot column

Add $-5/2$ times row 2 to row 3, and add $3/2$ times row 2 to row 4.

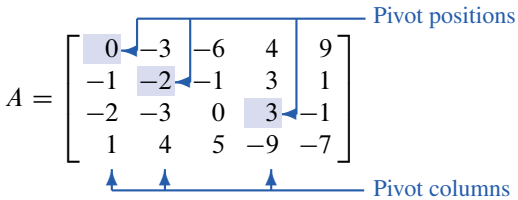
$$\begin{bmatrix} 1 & 4 & 5 & -9 & -7 \\ 0 & 2 & 4 & -6 & -6 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -5 & 0 \end{bmatrix} \tag{2}$$

The matrix in (2) is different from any encountered in Section 1.1. There is no way to create a leading entry in column 3! (We can't use row 1 or 2 because doing so would

destroy the echelon arrangement of the leading entries already produced.) However, if we interchange rows 3 and 4, we can produce a leading entry in column 4.



The matrix is in echelon form and thus reveals that columns 1, 2, and 4 of A are pivot columns.



$$A = \begin{bmatrix} 0 & -3 & -6 & 4 & 9 \\ -1 & -2 & -1 & 3 & 1 \\ -2 & -3 & 0 & 3 & -1 \\ 1 & 4 & 5 & -9 & -7 \end{bmatrix} \tag{3}$$

A **pivot**, as illustrated in Example 2, is a nonzero number in a pivot position that is used as needed to create zeros via row operations. The pivots in Example 2 were 1, 2, and −5. Notice that these numbers are not the same as the actual elements of A in the highlighted pivot positions shown in (3).

With Example 2 as a guide, we are ready to describe an efficient procedure for transforming a matrix into an echelon or reduced echelon matrix. Careful study and mastery of this procedure now will pay rich dividends later in the course.

The Row Reduction Algorithm

The algorithm that follows consists of four steps, and it produces a matrix in echelon form. A fifth step produces a matrix in reduced echelon form. We illustrate the algorithm by an example.

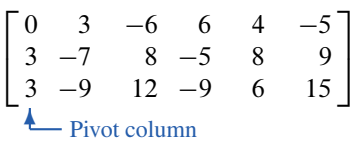
EXAMPLE 3 Apply elementary row operations to transform the following matrix first into echelon form and then into reduced echelon form:

$$\begin{bmatrix} 0 & 3 & -6 & 6 & 4 & -5 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 3 & -9 & 12 & -9 & 6 & 15 \end{bmatrix}$$

SOLUTION

Step 1

Begin with the leftmost nonzero column. This is a pivot column. The pivot position is at the top.



Step 2

Select a nonzero entry in the pivot column as a pivot. If necessary, interchange rows to move this entry into the pivot position.

Interchange rows 1 and 3. (We could have interchanged rows 1 and 2 instead.)

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 3 & -7 & 8 & -5 & 8 & 9 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix}$$

Pivot

Step 3

Use row replacement operations to create zeros in all positions below the pivot.

As a preliminary step, we could divide the top row by the pivot, 3. But with two 3's in column 1, it is just as easy to add -1 times row 1 to row 2.

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix}$$

Pivot

Step 4

Cover (or ignore) the row containing the pivot position and cover all rows, if any, above it. Apply steps 1–3 to the submatrix that remains. Repeat the process until there are no more nonzero rows to modify.

With row 1 covered, step 1 shows that column 2 is the next pivot column; for step 2, select as a pivot the "top" entry in that column.

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 3 & -6 & 6 & 4 & -5 \end{bmatrix}$$

Pivot

New pivot column

For step 3, we could insert an optional step of dividing the "top" row of the submatrix by the pivot, 2. Instead, we add $-3/2$ times the "top" row to the row below. This produces

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix}$$

When we cover the row containing the second pivot position for step 4, we are left with a new submatrix having only one row:

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix} \quad \text{Pivot}$$

Steps 1–3 require no work for this submatrix, and we have reached an echelon form of the full matrix. If we want the reduced echelon form, we perform one more step.

Step 5

Beginning with the rightmost pivot and working upward and to the left, create zeros above each pivot. If a pivot is not 1, make it 1 by a scaling operation.

The rightmost pivot is in row 3. Create zeros above it, adding suitable multiples of row 3 to rows 2 and 1.

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 0 & -9 \\ 0 & 2 & -4 & 4 & 0 & -14 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix} \begin{matrix} \\ \leftarrow \text{Row } 1 + (-6) \cdot \text{row } 3 \\ \leftarrow \text{Row } 2 + (-2) \cdot \text{row } 3 \end{matrix}$$

The next pivot is in row 2. Scale this row, dividing by the pivot.

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 0 & -9 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix} \leftarrow \text{Row scaled by } \tfrac{1}{2}$$

Create a zero in column 2 by adding 9 times row 2 to row 1.

$$\begin{bmatrix} 3 & 0 & -6 & 9 & 0 & -72 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix} \leftarrow \text{Row } 1 + (9) \cdot \text{row } 2$$

Finally, scale row 1, dividing by the pivot, 3.

$$\begin{bmatrix} 1 & 0 & -2 & 3 & 0 & -24 \\ 0 & 1 & -2 & 2 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix} \leftarrow \text{Row scaled by } \tfrac{1}{3}$$

This is the reduced echelon form of the original matrix. ■

The combination of steps 1–4 is called the **forward phase** of the row reduction algorithm. Step 5, which produces the unique reduced echelon form, is called the **backward phase**.

Numerical Note

In step 2 on page 41, a computer program usually selects as a pivot the entry in a column having the largest absolute value. This strategy, called **partial pivoting**, is used because it reduces roundoff errors in the calculations.

Solutions of Linear Systems

The row reduction algorithm leads directly to an explicit description of the solution set of a linear system when the algorithm is applied to the augmented matrix of the system.

Suppose, for example, that the augmented matrix of a linear system has been changed into the equivalent *reduced* echelon form

$$\begin{bmatrix} 1 & 0 & -5 & 1 \\ 0 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

There are three variables because the augmented matrix has four columns. The associated system of equations is

$$\begin{aligned} x_1 \quad & -5x_3 = 1 \\ x_2 + \; & x_3 = 4 \\ 0 \; & = 0 \end{aligned} \tag{4}$$

The variables x_1 and x_2 corresponding to pivot columns in the matrix are called **basic variables**.[2] The other variable, x_3, is called a **free variable**.

Whenever a system is consistent, as in (4), the solution set can be described explicitly by solving the *reduced* system of equations for the basic variables in terms of the free variables. This operation is possible because the reduced echelon form places each basic variable in one and only one equation. In (4), solve the first equation for x_1 and the second for x_2. (Ignore the third equation; it offers no restriction on the variables.)

$$\begin{cases} x_1 = 1 + 5x_3 \\ x_2 = 4 - x_3 \\ x_3 \text{ is free} \end{cases} \tag{5}$$

The statement "x_3 is free" means that you are free to choose any value for x_3. Once that is done, the formulas in (5) determine the values for x_1 and x_2. For instance, when $x_3 = 0$, the solution is $(1, 4, 0)$; when $x_3 = 1$, the solution is $(6, 3, 1)$. *Each different choice of x_3 determines a (different) solution of the system, and every solution of the system is determined by a choice of x_3.*

EXAMPLE 4 Find the general solution of the linear system whose augmented matrix has been reduced to

$$\begin{bmatrix} 1 & 6 & 2 & -5 & -2 & -4 \\ 0 & 0 & 2 & -8 & -1 & 3 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix}$$

SOLUTION The matrix is in echelon form, but we want the reduced echelon form before solving for the basic variables. The row reduction is completed next. The symbol $\sim$ before a matrix indicates that the matrix is row equivalent to the preceding matrix.

$$\begin{bmatrix} 1 & 6 & 2 & -5 & -2 & -4 \\ 0 & 0 & 2 & -8 & -1 & 3 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & 6 & 2 & -5 & 0 & 10 \\ 0 & 0 & 2 & -8 & 0 & 10 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 6 & 2 & -5 & 0 & 10 \\ 0 & 0 & 1 & -4 & 0 & 5 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix} \sim \begin{bmatrix} 1 & 6 & 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & -4 & 0 & 5 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix}$$

[2] Some texts use the term *leading variables* because they correspond to the columns containing leading entries.

There are five variables because the augmented matrix has six columns. The associated system now is

$$\begin{aligned} x_1 + 6x_2 \quad + 3x_4 \quad &= 0 \\ x_3 - 4x_4 \quad &= 5 \\ x_5 &= 7 \end{aligned} \tag{6}$$

The pivot columns of the matrix are 1, 3, and 5, so the basic variables are x_1, x_3, and x_5. The remaining variables, x_2 and x_4, must be free. Solve for the basic variables to obtain the general solution:

$$\begin{cases} x_1 = -6x_2 - 3x_4 \\ x_2 \text{ is free} \\ x_3 = 5 + 4x_4 \\ x_4 \text{ is free} \\ x_5 = 7 \end{cases} \tag{7}$$

Note that the value of x_5 is already fixed by the third equation in system (6). ■

Parametric Descriptions of Solution Sets

The descriptions in (5) and (7) are *parametric descriptions* of solution sets in which the free variables act as parameters. *Solving a system* amounts to finding a parametric description of the solution set or determining that the solution set is empty.

Whenever a system is consistent and has free variables, the solution set has many parametric descriptions. For instance, in system (4), we may add 5 times equation 2 to equation 1 and obtain the equivalent system

$$\begin{aligned} x_1 + 5x_2 \quad &= 21 \\ x_2 + x_3 &= 4 \end{aligned}$$

We could treat x_2 as a parameter and solve for x_1 and x_3 in terms of x_2, and we would have an accurate description of the solution set. However, to be consistent, we make the (arbitrary) convention of always using the free variables as the parameters for describing a solution set. (The answer section at the end of the text also reflects this convention.)

Whenever a system is inconsistent, the solution set is empty, even when the system has free variables. In this case, the solution set has *no* parametric representation.

Back-Substitution

Consider the following system, whose augmented matrix is in echelon form but is *not* in reduced echelon form:

$$\begin{aligned} x_1 - 7x_2 + 2x_3 - 5x_4 + 8x_5 &= 10 \\ x_2 - 3x_3 + 3x_4 + x_5 &= -5 \\ x_4 - x_5 &= 4 \end{aligned}$$

A computer program would solve this system by back-substitution, rather than by computing the reduced echelon form. That is, the program would solve equation 3 for x_4 in terms of x_5 and substitute the expression for x_4 into equation 2, solve equation 2 for x_2, and then substitute the expressions for x_2 and x_4 into equation 1 and solve for x_1.

Our matrix format for the backward phase of row reduction, which produces the reduced echelon form, has the same number of arithmetic operations as back-substitution. But the discipline of the matrix format substantially reduces the likelihood of errors

during hand computations. The best strategy is to use only the *reduced* echelon form to solve a system! The *Study Guide* that accompanies this text offers several helpful suggestions for performing row operations accurately and rapidly.

Numerical Note

In general, the forward phase of row reduction takes much longer than the backward phase. An algorithm for solving a system is usually measured in flops (or floating point operations). A **flop** is one arithmetic operation $(+, -, *, /)$ on two real floating point numbers.[3] For an $n \times (n+1)$ matrix, the reduction to echelon form can take $2n^3/3 + n^2/2 - 7n/6$ flops (which is approximately $2n^3/3$ flops when n is moderately large—say, $n \geq 30$). In contrast, further reduction to reduced echelon form needs at most n^2 flops.

Existence and Uniqueness Questions

Although a nonreduced echelon form is a poor tool for solving a system, this form is just the right device for answering two fundamental questions posed in Section 1.1.

EXAMPLE 5 Determine the existence and uniqueness of the solutions to the system

$$\begin{aligned} 3x_2 - 6x_3 + 6x_4 + 4x_5 &= -5 \\ 3x_1 - 7x_2 + 8x_3 - 5x_4 + 8x_5 &= 9 \\ 3x_1 - 9x_2 + 12x_3 - 9x_4 + 6x_5 &= 15 \end{aligned}$$

SOLUTION The augmented matrix of this system was row reduced in Example 3 to

$$\begin{bmatrix} 3 & -9 & 12 & -9 & 6 & 15 \\ 0 & 2 & -4 & 4 & 2 & -6 \\ 0 & 0 & 0 & 0 & 1 & 4 \end{bmatrix} \tag{8}$$

The basic variables are x_1, x_2, and x_5; the free variables are x_3 and x_4. There is no equation such as $0 = 1$ that would indicate an inconsistent system, so we could use back-substitution to find a solution. But the *existence* of a solution is already clear in (8). Also, the solution is *not unique* because there are free variables. Each different choice of x_3 and x_4 determines a different solution. Thus the system has infinitely many solutions. ■

When a system is in echelon form and contains no equation of the form $0 = b$, with b nonzero, every nonzero equation contains a basic variable with a nonzero coefficient. Either the basic variables are completely determined (with no free variables) or at least one of the basic variables may be expressed in terms of one or more free variables. In the former case, there is a unique solution; in the latter case, there are infinitely many solutions (one for each choice of values for the free variables).

These remarks justify the following theorem.

[3] Traditionally, a *flop* was only a multiplication or division because addition and subtraction took much less time and could be ignored. The definition of *flop* given here is preferred now, as a result of advances in computer architecture. See Golub and Van Loan, *Matrix Computations*, 2nd ed. (Baltimore: The Johns Hopkins Press, 1989), pp. 19–20.

THEOREM 2

Existence and Uniqueness Theorem

A linear system is consistent if and only if the rightmost column of the augmented matrix is *not* a pivot column—that is, if and only if an echelon form of the augmented matrix has *no* row of the form

$$[\,0 \quad \cdots \quad 0 \quad b\,] \qquad \text{with } b \text{ nonzero}$$

If a linear system is consistent, then the solution set contains either (i) a unique solution, when there are no free variables, or (ii) infinitely many solutions, when there is at least one free variable.

The following procedure outlines how to find and describe all solutions of a linear system.

USING ROW REDUCTION TO SOLVE A LINEAR SYSTEM

1. Write the augmented matrix of the system.
2. Use the row reduction algorithm to obtain an equivalent augmented matrix in echelon form. Decide whether the system is consistent. If there is no solution, stop; otherwise, go to the next step.
3. Continue row reduction to obtain the reduced echelon form.
4. Write the system of equations corresponding to the matrix obtained in step 3.
5. Rewrite each nonzero equation from step 4 so that its one basic variable is expressed in terms of any free variables appearing in the equation.

Reasonable Answers

Remember that each augmented matrix corresponds to a system of equations. If you row reduce the augmented matrix $\begin{bmatrix} 1 & -2 & 1 & 2 \\ 1 & -1 & 2 & 5 \\ 0 & 1 & 1 & 3 \end{bmatrix}$ to get the matrix $\begin{bmatrix} 1 & 0 & 3 & 8 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$, the solution set is

$$\begin{cases} x_1 = 8 - 3x_3 \\ x_2 = 3 - x_3 \\ x_3 \text{ is free} \end{cases}$$

The system of equations corresponding to the original augmented matrix is

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

You can now check whether your solution is correct by substituting it into the original equations. Notice that you can just leave the free variables in the solution.

$$\begin{aligned}(8-3x_3) &- 2(3-x_3) + (x_3) = 8-3x_3-6+2x_3+x_3 = 2\\ (8-3x_3) &- (3-x_3) + 2(x_3) = 8-3x_3-3+x_3+2x_3 = 5\\ &(3-x_3) + (x_3) = 3-x_3+x_3 = 3\end{aligned}$$

You can now be confident you have a correct solution to the system of equations represented by the augmented matrix.

Practice Problems

1. Find the general solution of the linear system whose augmented matrix is

$$\begin{bmatrix}1 & -3 & -5 & 0\\ 0 & 1 & -1 & -1\end{bmatrix}$$

2. Find the general solution of the system

$$\begin{aligned}x_1 - 2x_2 - x_3 + 3x_4 &= 0\\ -2x_1 + 4x_2 + 5x_3 - 5x_4 &= 3\\ 3x_1 - 6x_2 - 6x_3 + 8x_4 &= 2\end{aligned}$$

3. Suppose a 4×7 coefficient matrix for a system of equations has 4 pivots. Is the system consistent? If the system is consistent, how many solutions are there?

1.2 Exercises

In Exercises 1 and 2, determine which matrices are in reduced echelon form and which others are only in echelon form.

1. a. $\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&1\end{bmatrix}$ b. $\begin{bmatrix}1&0&1&0\\0&0&1&0\\0&0&0&1\end{bmatrix}$

c. $\begin{bmatrix}1&0&0&0\\0&1&1&0\\0&0&0&0\\0&0&0&0\end{bmatrix}$ d. $\begin{bmatrix}1&1&0&1&1\\0&2&0&2&2\\0&0&0&3&3\\0&0&0&0&4\end{bmatrix}$

2. a. $\begin{bmatrix}1&1&0&1\\0&0&1&1\\0&0&0&0\end{bmatrix}$ b. $\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&1\end{bmatrix}$

c. $\begin{bmatrix}1&0&0&0\\1&0&0&0\\0&1&0&0\\0&0&1&1\end{bmatrix}$ d. $\begin{bmatrix}0&1&1&1&1\\0&0&2&2&2\\0&0&0&0&3\\0&0&0&0&0\end{bmatrix}$

Row reduce the matrices in Exercises 3 and 4 to reduced echelon form. Circle the pivot positions in the final matrix and in the original matrix, and list the pivot columns.

3. $\begin{bmatrix}1&2&3&4\\4&5&6&7\\6&7&8&9\end{bmatrix}$ **4.** $\begin{bmatrix}1&3&5&7\\3&5&7&9\\5&7&9&1\end{bmatrix}$

5. Describe the possible echelon forms of a nonzero 2×2 matrix. Use the symbols ■, $*$, and 0, as in the first part of Example 1.

6. Repeat Exercise 5 for a nonzero 3×2 matrix.

Find the general solutions of the systems whose augmented matrices are given in Exercises 7–14.

7. $\begin{bmatrix}1&2&3&4\\4&8&9&4\end{bmatrix}$ **8.** $\begin{bmatrix}1&4&0&7\\2&7&0&11\end{bmatrix}$

9. $\begin{bmatrix}0&1&-6&5\\1&-2&7&-4\end{bmatrix}$ **10.** $\begin{bmatrix}1&-2&-1&3\\3&-6&-2&2\end{bmatrix}$

11. $\begin{bmatrix}3&-4&2&0\\-9&12&-6&0\\-6&8&-4&0\end{bmatrix}$ **12.** $\begin{bmatrix}1&-7&0&6&5\\0&0&1&-2&-3\\-1&7&-4&2&7\end{bmatrix}$

13. $\begin{bmatrix}1&-3&0&-1&0&-2\\0&1&0&0&-4&1\\0&0&0&1&9&-4\\0&0&0&0&0&0\end{bmatrix}$

14. $\begin{bmatrix} 1 & 2 & -5 & -4 & 0 & -5 \\ 0 & 1 & -6 & -4 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

You may find it helpful to review the information in the Reasonable Answers box from this section before answering Exercises 15–18.

15. Write down the equations corresponding to the augmented matrix in Exercise 9 and verify your answer to Exercise 9 is correct by substituting the solutions you obtained back into the original equations.

16. Write down the equations corresponding to the augmented matrix in Exercise 10 and verify your answer to Exercise 10 is correct by substituting the solutions you obtained back into the original equations.

17. Write down the equations corresponding to the augmented matrix in Exercise 11 and verify your answer to Exercise 11 is correct by substituting the solutions you obtained back into the original equations.

18. Write down the equations corresponding to the augmented matrix in Exercise 12 and verify your answer to Exercise 12 is correct by substituting the solutions you obtained back into the original equations.

Exercises 19 and 20 use the notation of Example 1 for matrices in echelon form. Suppose each matrix represents the augmented matrix for a system of linear equations. In each case, determine if the system is consistent. If the system is consistent, determine if the solution is unique.

19. a. $\begin{bmatrix} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & \blacksquare & 0 \end{bmatrix}$

b. $\begin{bmatrix} 0 & \blacksquare & * & * & * \\ 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & 0 & \blacksquare \end{bmatrix}$

20. a. $\begin{bmatrix} \blacksquare & * & * \\ 0 & \blacksquare & * \\ 0 & 0 & 0 \end{bmatrix}$

b. $\begin{bmatrix} \blacksquare & * & * & * & * \\ 0 & 0 & \blacksquare & * & * \\ 0 & 0 & 0 & \blacksquare & * \end{bmatrix}$

In Exercises 21 and 22, determine the value(s) of h such that the matrix is the augmented matrix of a consistent linear system.

21. $\begin{bmatrix} 2 & 3 & h \\ 4 & 6 & 7 \end{bmatrix}$ **22.** $\begin{bmatrix} 1 & -4 & -3 \\ 6 & h & -9 \end{bmatrix}$

In Exercises 23 and 24, choose h and k such that the system has (a) no solution, (b) a unique solution, and (c) many solutions. Give separate answers for each part.

23. $x_1 + hx_2 = 2$
$4x_1 + 8x_2 = k$

24. $x_1 + 4x_2 = 5$
$2x_1 + hx_2 = k$

In Exercises 25–34, mark each statement True or False **(T/F)**. Justify each answer.[4]

25. **(T/F)** In some cases, a matrix may be row reduced to more than one matrix in reduced echelon form, using different sequences of row operations.

26. **(T/F)** The echelon form of a matrix is unique.

27. **(T/F)** The row reduction algorithm applies only to augmented matrices for a linear system.

28. **(T/F)** The pivot positions in a matrix depend on whether row interchanges are used in the row reduction process.

29. **(T/F)** A basic variable in a linear system is a variable that corresponds to a pivot column in the coefficient matrix.

30. **(T/F)** Reducing a matrix to echelon form is called the *forward phase* of the row reduction process.

31. **(T/F)** Finding a parametric description of the solution set of a linear system is the same as *solving* the system.

32. **(T/F)** Whenever a system has free variables, the solution set contains a unique solution.

33. **(T/F)** If one row in an echelon form of an augmented matrix is $[0 \quad 0 \quad 0 \quad 0 \quad 5]$, then the associated linear system is inconsistent.

34. **(T/F)** A general solution of a system is an explicit description of all solutions of the system.

35. Suppose a 3×5 *coefficient* matrix for a system has three pivot columns. Is the system consistent? Why or why not?

36. Suppose a system of linear equations has a 3×5 *augmented* matrix whose fifth column is a pivot column. Is the system consistent? Why (or why not)?

37. Suppose the coefficient matrix of a system of linear equations has a pivot position in every row. Explain why the system is consistent.

38. Suppose the coefficient matrix of a linear system of three equations in three variables has a pivot in each column. Explain why the system has a unique solution.

39. Restate the last sentence in Theorem 2 using the concept of pivot columns: "If a linear system is consistent, then the solution is unique if and only if _____."

40. What would you have to know about the pivot columns in an augmented matrix in order to know that the linear system is consistent and has a unique solution?

41. A system of linear equations with fewer equations than unknowns is sometimes called an *underdetermined system*.

[4] True/false questions of this type will appear in many sections. Methods for justifying your answers were described before the True or False exercises in Section 1.1.

Suppose that such a system happens to be consistent. Explain why there must be an infinite number of solutions.

42. Give an example of an inconsistent underdetermined system of two equations in three unknowns.

43. A system of linear equations with more equations than unknowns is sometimes called an *overdetermined system*. Can such a system be consistent? Illustrate your answer with a specific system of three equations in two unknowns.

44. Suppose an $n \times (n+1)$ matrix is row reduced to reduced echelon form. Approximately what fraction of the total number of operations (flops) is involved in the backward phase of the reduction when $n = 30$? when $n = 300$?

Suppose experimental data are represented by a set of points in the plane. An **interpolating polynomial** for the data is a polynomial whose graph passes through every point. In scientific work, such a polynomial can be used, for example, to estimate values between the known data points. Another use is to create curves for graphical images on a computer screen. One method for finding an interpolating polynomial is to solve a system of linear equations.

45. Find the interpolating polynomial $p(t) = a_0 + a_1 t + a_2 t^2$ for the data $(1, 11)$, $(2, 16)$, $(3, 19)$. That is, find a_0, a_1, and a_2 such that

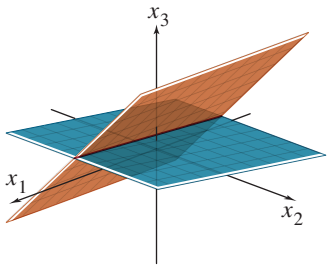
$$a_0 + a_1(1) + a_2(1)^2 = 11$$
$$a_0 + a_1(2) + a_2(2)^2 = 16$$
$$a_0 + a_1(3) + a_2(3)^2 = 19$$

T 46. In a wind tunnel experiment, the force on a projectile due to air resistance was measured at different velocities:

Velocity (100 ft/sec)	0	2	4	6	8	10
Force (100 lb)	0	2.90	14.8	39.6	74.3	119

Find an interpolating polynomial for these data and estimate the force on the projectile when the projectile is traveling at 750 ft/sec. Use $p(t) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 + a_5 t^5$. What happens if you try to use a polynomial of degree less than 5? (Try a cubic polynomial, for instance.)[5]

[5] Exercises marked with the symbol T are designed to be worked with the aid of a "Matrix program" (a computer program, such as MATLAB, Maple, Mathematica, MathCad, Octave, or Derive, or a programmable calculator with matrix capabilities, such as those manufactured by Texas Instruments or Hewlett-Packard).



The general solution of the system of equations is the line of intersection of the two planes.

Solutions to Practice Problems

1. The reduced echelon form of the augmented matrix and the corresponding system are

$$\begin{bmatrix} 1 & 0 & -8 & -3 \\ 0 & 1 & -1 & -1 \end{bmatrix} \quad \text{and} \quad \begin{aligned} x_1 - 8x_3 &= -3 \\ x_2 - x_3 &= -1 \end{aligned}$$

The basic variables are x_1 and x_2, and the general solution is

$$\begin{cases} x_1 = -3 + 8x_3 \\ x_2 = -1 + x_3 \\ x_3 \text{ is free} \end{cases}$$

Note: It is essential that the general solution describe each variable, with any parameters clearly identified. The following statement does *not* describe the solution:

$$\begin{cases} x_1 = -3 + 8x_3 \\ x_2 = -1 + x_3 \\ x_3 = 1 + x_2 \end{cases} \quad \text{Incorrect solution}$$

This description implies that x_2 and x_3 are *both* free, which certainly is not the case.

2. Row reduce the system's augmented matrix:

$$\begin{bmatrix} 1 & -2 & -1 & 3 & 0 \\ -2 & 4 & 5 & -5 & 3 \\ 3 & -6 & -6 & 8 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & -2 & -1 & 3 & 0 \\ 0 & 0 & 3 & 1 & 3 \\ 0 & 0 & -3 & -1 & 2 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & -2 & -1 & 3 & 0 \\ 0 & 0 & 3 & 1 & 3 \\ 0 & 0 & 0 & 0 & 5 \end{bmatrix}$$

Solutions to Practice Problems (Continued)

This echelon matrix shows that the system is *inconsistent*, because its rightmost column is a pivot column; the third row corresponds to the equation 0 = 5. There is no need to perform any more row operations. Note that the presence of the free variables in this problem is irrelevant because the system is inconsistent.

3. Since the coefficient matrix has four pivots, there is a pivot in every row of the coefficient matrix. This means that when the coefficient matrix is row reduced, it will *not* have a row of zeros, thus the corresponding row reduced augmented matrix can never have a row of the form $[0 \ 0 \ \cdots \ 0 \ b]$, where b is a nonzero number. By Theorem 2, the system is consistent. Moreover, since there are seven columns in the coefficient matrix and only four pivot columns, there will be three free variables resulting in infinitely many solutions.

1.3 Vector Equations

Important properties of linear systems can be described with the concept and notation of vectors. This section connects equations involving vectors to ordinary systems of equations. The term *vector* appears in a variety of mathematical and physical contexts, which we will discuss in Chapter 4, "Vector Spaces." Until then, *vector* will mean an *ordered list of numbers*. This simple idea enables us to get to interesting and important applications as quickly as possible.

Vectors in $\mathbb{R}^2$

A matrix with only one column is called a **column vector** or simply a **vector**. Examples of vectors with two entries are

$$\mathbf{u} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}, \qquad \mathbf{v} = \begin{bmatrix} .2 \\ .3 \end{bmatrix}, \qquad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

where w_1 and w_2 are any real numbers. The set of all vectors with two entries is denoted by $\mathbb{R}^2$ (read "r-two"). The $\mathbb{R}$ stands for the real numbers that appear as entries in the vectors, and the exponent 2 indicates that each vector contains two entries.[1]

Two vectors in $\mathbb{R}^2$ are **equal** if and only if their corresponding entries are equal. Thus $\begin{bmatrix} 4 \\ 7 \end{bmatrix}$ and $\begin{bmatrix} 7 \\ 4 \end{bmatrix}$ are *not* equal, because vectors in $\mathbb{R}^2$ are *ordered pairs* of real numbers.

Given two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^2$, their **sum** is the vector $\mathbf{u} + \mathbf{v}$ obtained by adding corresponding entries of $\mathbf{u}$ and $\mathbf{v}$. For example,

$$\begin{bmatrix} 1 \\ -2 \end{bmatrix} + \begin{bmatrix} 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 1+2 \\ -2+5 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$$

Given a vector $\mathbf{u}$ and a real number c, the **scalar multiple** of $\mathbf{u}$ by c is the vector $c\mathbf{u}$ obtained by multiplying each entry in $\mathbf{u}$ by c. For instance,

$$\text{if} \quad \mathbf{u} = \begin{bmatrix} 3 \\ -1 \end{bmatrix} \quad \text{and} \quad c = 5, \qquad \text{then} \quad c\mathbf{u} = 5\begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 15 \\ -5 \end{bmatrix}$$

[1] Most of the text concerns vectors and matrices that have only real entries. However, all definitions and theorems in Chapters 1–5, and in most of the rest of the text, remain valid if the entries are complex numbers. Complex vectors and matrices arise naturally, for example, in electrical engineering and physics.

The number c in $c\mathbf{u}$ is called a **scalar**; it is written in lightface type to distinguish it from the boldface vector $\mathbf{u}$.

The operations of scalar multiplication and vector addition can be combined, as in the following example.

EXAMPLE 1 Given $\mathbf{u} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 2 \\ -5 \end{bmatrix}$, find $4\mathbf{u}$, $(-3)\mathbf{v}$, and $4\mathbf{u} + (-3)\mathbf{v}$.

SOLUTION

$$4\mathbf{u} = \begin{bmatrix} 4 \\ -8 \end{bmatrix}, \qquad (-3)\mathbf{v} = \begin{bmatrix} -6 \\ 15 \end{bmatrix}$$

and

$$4\mathbf{u} + (-3)\mathbf{v} = \begin{bmatrix} 4 \\ -8 \end{bmatrix} + \begin{bmatrix} -6 \\ 15 \end{bmatrix} = \begin{bmatrix} -2 \\ 7 \end{bmatrix}$$

■

Sometimes, for convenience (and also to save space), this text may write a column vector such as $\begin{bmatrix} 3 \\ -1 \end{bmatrix}$ in the form $(3, -1)$. In this case, the parentheses and the comma distinguish the vector $(3, -1)$ from the 1×2 row matrix $\begin{bmatrix} 3 & -1 \end{bmatrix}$, written with brackets and no comma. Thus

$$\begin{bmatrix} 3 \\ -1 \end{bmatrix} \neq \begin{bmatrix} 3 & -1 \end{bmatrix}$$

because the matrices have different shapes, even though they have the same entries.

Geometric Descriptions of $\mathbb{R}^2$

Consider a rectangular coordinate system in the plane. Because each point in the plane is determined by an ordered pair of numbers, *we can identify a geometric point* (a, b) *with the column vector* $\begin{bmatrix} a \\ b \end{bmatrix}$. So we may regard $\mathbb{R}^2$ as the set of all points in the plane. See Figure 1.

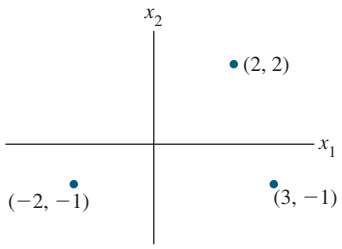


FIGURE 1 Vectors as points.

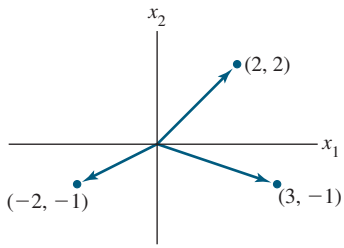


FIGURE 2 Vectors with arrows.

The geometric visualization of a vector such as $\begin{bmatrix} 3 \\ -1 \end{bmatrix}$ is often aided by including an arrow (directed line segment) from the origin $(0, 0)$ to the point $(3, -1)$, as in Figure 2. In this case, the individual points along the arrow itself have no special significance.[2]

The sum of two vectors has a useful geometric representation. The following rule can be verified by analytic geometry.

[2] In physics, arrows can represent forces and usually are free to move about in space. This interpretation of vectors will be discussed in Section 4.1.

Parallelogram Rule for Addition

If $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^2$ are represented as points in the plane, then $\mathbf{u} + \mathbf{v}$ corresponds to the fourth vertex of the parallelogram whose other vertices are $\mathbf{u}$, $\mathbf{0}$, and $\mathbf{v}$. See Figure 3.

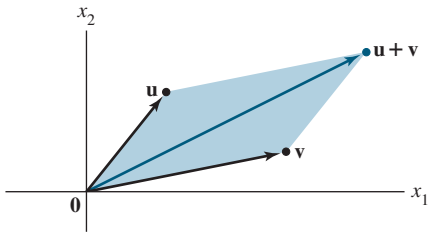


FIGURE 3 The parallelogram rule.

EXAMPLE 2 The vectors $\mathbf{u} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} -6 \\ 1 \end{bmatrix}$, and $\mathbf{u} + \mathbf{v} = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$ are displayed in Figure 4.

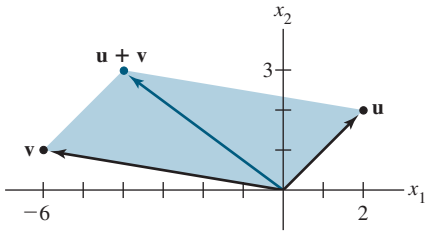
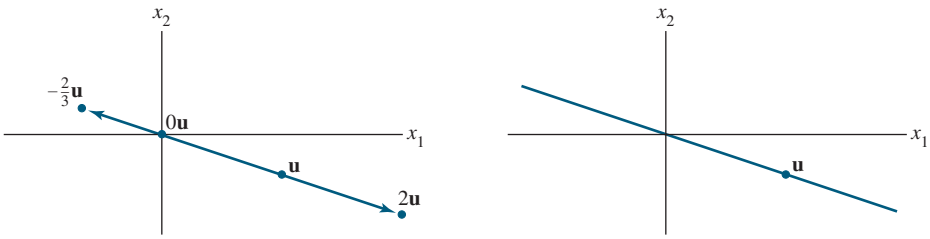


FIGURE 4

The next example illustrates the fact that the set of all scalar multiples of one fixed nonzero vector is a line through the origin, $(0, 0)$.

EXAMPLE 3 Let $\mathbf{u} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$. Display the vectors $\mathbf{u}$, $2\mathbf{u}$, and $-\frac{2}{3}\mathbf{u}$ on a graph.

SOLUTION See Figure 5, where $\mathbf{u}$, $2\mathbf{u} = \begin{bmatrix} 6 \\ -2 \end{bmatrix}$, and $-\frac{2}{3}\mathbf{u} = \begin{bmatrix} -2 \\ 2/3 \end{bmatrix}$ are displayed. The arrow for $2\mathbf{u}$ is twice as long as the arrow for $\mathbf{u}$, and the arrows point in the same direction. The arrow for $-\frac{2}{3}\mathbf{u}$ is two-thirds the length of the arrow for $\mathbf{u}$, and the arrows point in opposite directions. In general, the length of the arrow for $c\mathbf{u}$ is $|c|$ times the length of the arrow for $\mathbf{u}$. [Recall that the length of the line segment from $(0, 0)$ to (a, b) is $\sqrt{a^2 + b^2}$.



Typical multiples of $\mathbf{u}$ The set of all multiples of $\mathbf{u}$

FIGURE 5

Vectors in $\mathbb{R}^3$

Vectors in $\mathbb{R}^3$ are 3×1 column matrices with three entries. They are represented geometrically by points in a three-dimensional coordinate space, with arrows from the origin sometimes included for visual clarity. The vectors $\mathbf{a} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$ and $2\mathbf{a}$ are displayed in Figure 6.

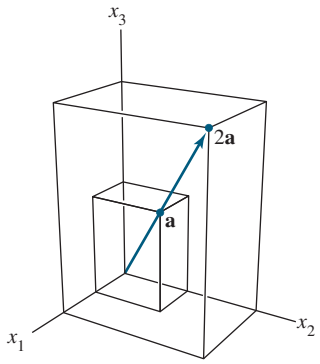


FIGURE 6
Scalar multiples.

Vectors in $\mathbb{R}^n$

If n is a positive integer, $\mathbb{R}^n$ (read "r-n") denotes the collection of all lists (or *ordered n-tuples*) of n real numbers, usually written as $n \times 1$ column matrices, such as

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

The vector whose entries are all zero is called the **zero vector** and is denoted by $\mathbf{0}$. (The number of entries in $\mathbf{0}$ will be clear from the context.)

Equality of vectors in $\mathbb{R}^n$ and the operations of scalar multiplication and vector addition in $\mathbb{R}^n$ are defined entry by entry just as in $\mathbb{R}^2$. These operations on vectors have the following properties, which can be verified directly from the corresponding properties for real numbers. See Practice Problem 1 and Exercises 41 and 42 at the end of this section.

Algebraic Properties of $\mathbb{R}^n$

For all $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in $\mathbb{R}^n$ and all scalars c and d:

(i) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$

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

(iii) $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$

(iv) $\mathbf{u} + (-\mathbf{u}) = -\mathbf{u} + \mathbf{u} = \mathbf{0}$, where $-\mathbf{u}$ denotes $(-1)\mathbf{u}$

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

(vi) $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$

(vii) $c(d\mathbf{u}) = (cd)\mathbf{u}$

(viii) $1\mathbf{u} = \mathbf{u}$

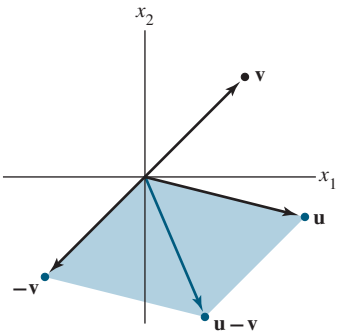


FIGURE 7
Vector subtraction.

For simplicity of notation, a vector such as $\mathbf{u} + (-1)\mathbf{v}$ is often written as $\mathbf{u} - \mathbf{v}$. Figure 7 shows $\mathbf{u} - \mathbf{v}$ as the sum of $\mathbf{u}$ and $-\mathbf{v}$.

Linear Combinations

Given vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p$ in $\mathbb{R}^n$ and given scalars $c_1, c_2, \ldots, c_p$, the vector $\mathbf{y}$ defined by

$$\mathbf{y} = c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p$$

is called a **linear combination** of $\mathbf{v}_1, \ldots, \mathbf{v}_p$ with **weights** $c_1, \ldots, c_p$. Algebraic Property (ii) above permits us to omit parentheses when forming such a linear combination. The weights in a linear combination can be any real numbers, including zero. For example, some linear combinations of vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are

$$\sqrt{3}\,\mathbf{v}_1 + \mathbf{v}_2, \quad \tfrac{1}{2}\mathbf{v}_1 \ (= \tfrac{1}{2}\mathbf{v}_1 + 0\mathbf{v}_2), \quad \text{and} \quad \mathbf{0} \ (= 0\mathbf{v}_1 + 0\mathbf{v}_2)$$

EXAMPLE 4 Figure 8 identifies selected linear combinations of $\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. (Note that sets of parallel grid lines are drawn through integer multiples of $\mathbf{v}_1$ and $\mathbf{v}_2$.) Estimate the linear combinations of $\mathbf{v}_1$ and $\mathbf{v}_2$ that generate the vectors $\mathbf{u}$ and $\mathbf{w}$.

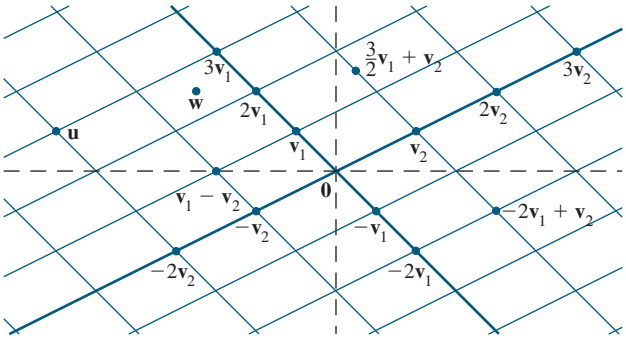


FIGURE 8 Linear combinations of $\mathbf{v}_1$ and $\mathbf{v}_2$.

SOLUTION The parallelogram rule shows that $\mathbf{u}$ is the sum of $3\mathbf{v}_1$ and $-2\mathbf{v}_2$; that is,

$$\mathbf{u} = 3\mathbf{v}_1 - 2\mathbf{v}_2$$

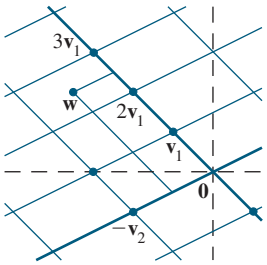


FIGURE 9

This expression for $\mathbf{u}$ can be interpreted as instructions for traveling from the origin to $\mathbf{u}$ along two straight paths. First, travel 3 units in the $\mathbf{v}_1$ direction to $3\mathbf{v}_1$, and then travel -2 units in the $\mathbf{v}_2$ direction (parallel to the line through $\mathbf{v}_2$ and $\mathbf{0}$). Next, although the vector $\mathbf{w}$ is not on a grid line, $\mathbf{w}$ appears to be about halfway between two pairs of grid lines, at the vertex of a parallelogram determined by $(5/2)\mathbf{v}_1$ and $(-1/2)\mathbf{v}_2$. (See Figure 9.) Thus a reasonable estimate for $\mathbf{w}$ is

$$\mathbf{w} = \tfrac{5}{2}\mathbf{v}_1 - \tfrac{1}{2}\mathbf{v}_2$$

The next example connects a problem about linear combinations to the fundamental existence question studied in Sections 1.1 and 1.2.

EXAMPLE 5 Let $\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}$. Determine whether $\mathbf{b}$ can be generated (or written) as a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$. That is, determine whether weights x_1 and x_2 exist such that

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{b} \tag{1}$$

If vector equation (1) has a solution, find it.

SOLUTION Use the definitions of scalar multiplication and vector addition to rewrite the vector equation

$$x_1\begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix} + x_2\begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}$$

$\mathbf{a}_1$ $\mathbf{a}_2$ $\mathbf{b}$

which is the same as

$$\begin{bmatrix} x_1 \\ -2x_1 \\ -5x_1 \end{bmatrix} + \begin{bmatrix} 2x_2 \\ 5x_2 \\ 6x_2 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}$$

and

$$\begin{bmatrix} x_1 + 2x_2 \\ -2x_1 + 5x_2 \\ -5x_1 + 6x_2 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix} \tag{2}$$

The vectors on the left and right sides of (2) are equal if and only if their corresponding entries are both equal. That is, x_1 and x_2 make the vector equation (1) true if and only if x_1 and x_2 satisfy the system

$$\begin{aligned} x_1 + 2x_2 &= 7 \\ -2x_1 + 5x_2 &= 4 \\ -5x_1 + 6x_2 &= -3 \end{aligned} \tag{3}$$

To solve this system, row reduce the augmented matrix of the system as follows:[3]

$$\begin{bmatrix} 1 & 2 & 7 \\ -2 & 5 & 4 \\ -5 & 6 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 7 \\ 0 & 9 & 18 \\ 0 & 16 & 32 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 7 \\ 0 & 1 & 2 \\ 0 & 16 & 32 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

The solution of (3) is $x_1 = 3$ and $x_2 = 2$. Hence $\mathbf{b}$ is a linear combination of $\mathbf{a}_1$ and $\mathbf{a}_2$, with weights $x_1 = 3$ and $x_2 = 2$. That is,

$$3\begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix} + 2\begin{bmatrix} 2 \\ 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ -3 \end{bmatrix}$$

■

Observe in Example 5 that the original vectors $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{b}$ are the columns of the augmented matrix that we row reduced:

$$\begin{bmatrix} 1 & 2 & 7 \\ -2 & 5 & 4 \\ -5 & 6 & -3 \end{bmatrix}$$
$$\uparrow \quad \uparrow \quad \uparrow$$
$$\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{b}$$

For brevity, write this matrix in a way that identifies its columns—namely

$$[\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{b}\,] \tag{4}$$

It is clear how to write this augmented matrix immediately from vector equation (1), without going through the intermediate steps of Example 5. Take the vectors in the order in which they appear in (1) and put them into the columns of a matrix as in (4).

The discussion above is easily modified to establish the following fundamental fact.

[3] The symbol ~ between matrices denotes row equivalence (Section 1.2).

A vector equation

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$$

has the same solution set as the linear system whose augmented matrix is

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n & \mathbf{b} \end{bmatrix} \tag{5}$$

In particular, $\mathbf{b}$ can be generated by a linear combination of $\mathbf{a}_1, \ldots, \mathbf{a}_n$ if and only if there exists a solution to the linear system corresponding to the matrix (5).

One of the key ideas in linear algebra is to study the set of all vectors that can be generated or written as a linear combination of a fixed set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ of vectors.

DEFINITION

If $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are in $\mathbb{R}^n$, then the set of all linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_p$ is denoted by $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ and is called the **subset of $\mathbb{R}^n$ spanned** (or **generated**) **by** $\mathbf{v}_1, \ldots, \mathbf{v}_p$. That is, $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is the collection of all vectors that can be written in the form

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p$$

with $c_1, \ldots, c_p$ scalars.

Asking whether a vector $\mathbf{b}$ is in $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ amounts to asking whether the vector equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_p\mathbf{v}_p = \mathbf{b}$$

has a solution, or, equivalently, asking whether the linear system with augmented matrix $[\,\mathbf{v}_1 \;\cdots\; \mathbf{v}_p \;\; \mathbf{b}\,]$ has a solution.

Note that $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ contains every scalar multiple of $\mathbf{v}_1$ (for example), since $c\mathbf{v}_1 = c\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_p$. In particular, the zero vector must be in $\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$.

A Geometric Description of Span{v} and Span{u, v}

Let $\mathbf{v}$ be a nonzero vector in $\mathbb{R}^3$. Then $\text{Span}\{\mathbf{v}\}$ is the set of all scalar multiples of $\mathbf{v}$, which is the set of points on the line in $\mathbb{R}^3$ through $\mathbf{v}$ and $\mathbf{0}$. See Figure 10.

If $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors in $\mathbb{R}^3$, with $\mathbf{v}$ not a multiple of $\mathbf{u}$, then $\text{Span}\{\mathbf{u}, \mathbf{v}\}$ is the plane in $\mathbb{R}^3$ that contains $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{0}$. In particular, $\text{Span}\{\mathbf{u}, \mathbf{v}\}$ contains the line in $\mathbb{R}^3$ through $\mathbf{u}$ and $\mathbf{0}$ and the line through $\mathbf{v}$ and $\mathbf{0}$. See Figure 11.

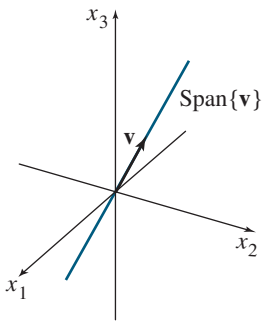


FIGURE 10 Span $\{\mathbf{v}\}$ as a line through the origin.

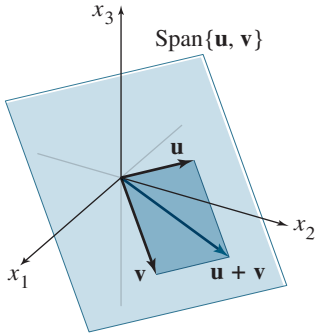


FIGURE 11 Span $\{\mathbf{u}, \mathbf{v}\}$ as a plane through the origin.

EXAMPLE 6 Let $\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} 5 \\ -13 \\ -3 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} -3 \\ 8 \\ 1 \end{bmatrix}$. Then Span $\{\mathbf{a}_1, \mathbf{a}_2\}$ is a plane through the origin in $\mathbb{R}^3$. Is $\mathbf{b}$ in that plane?

SOLUTION Does the equation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{b}$ have a solution? To answer this, row reduce the augmented matrix $[\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{b}\,]$:

$$\begin{bmatrix} 1 & 5 & -3 \\ -2 & -13 & 8 \\ 3 & -3 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 5 & -3 \\ 0 & -3 & 2 \\ 0 & -18 & 10 \end{bmatrix} \sim \begin{bmatrix} 1 & 5 & -3 \\ 0 & -3 & 2 \\ 0 & 0 & -2 \end{bmatrix}$$

The third equation is $0 = -2$, which shows that the system has no solution. The vector equation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 = \mathbf{b}$ has no solution, and so $\mathbf{b}$ is *not* in Span $\{\mathbf{a}_1, \mathbf{a}_2\}$. ■

Linear Combinations in Applications

The final example shows how scalar multiples and linear combinations can arise when a quantity such as "cost" is broken down into several categories. The basic principle for the example concerns the cost of producing several units of an item when the cost per unit is known:

$$\begin{Bmatrix} \text{number} \\ \text{of units} \end{Bmatrix} \cdot \begin{Bmatrix} \text{cost} \\ \text{per unit} \end{Bmatrix} = \begin{Bmatrix} \text{total} \\ \text{cost} \end{Bmatrix}$$

EXAMPLE 7 A company manufactures two products. For \$1.00 worth of product B, the company spends \$.45 on materials, \$.25 on labor, and \$.15 on overhead. For \$1.00 worth of product C, the company spends \$.40 on materials, \$.30 on labor, and \$.15 on overhead. Let

$$\mathbf{b} = \begin{bmatrix} .45 \\ .25 \\ .15 \end{bmatrix} \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} .40 \\ .30 \\ .15 \end{bmatrix}$$

Then $\mathbf{b}$ and $\mathbf{c}$ represent the "costs per dollar of income" for the two products.

a. What economic interpretation can be given to the vector $100\mathbf{b}$?

b. Suppose the company wishes to manufacture x_1 dollars worth of product B and x_2 dollars worth of product C. Give a vector that describes the various costs the company will have (for materials, labor, and overhead).

SOLUTION

a. Compute

$$100\mathbf{b} = 100\begin{bmatrix} .45 \\ .25 \\ .15 \end{bmatrix} = \begin{bmatrix} 45 \\ 25 \\ 15 \end{bmatrix}$$

The vector $100\mathbf{b}$ lists the various costs for producing \$100 worth of product B—namely \$45 for materials, \$25 for labor, and \$15 for overhead.

b. The costs of manufacturing x_1 dollars worth of B are given by the vector $x_1\mathbf{b}$, and the costs of manufacturing x_2 dollars worth of C are given by $x_2\mathbf{c}$. Hence the total costs for both products are given by the vector $x_1\mathbf{b} + x_2\mathbf{c}$. ■

Practice Problems

1. Prove that $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ for any $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$.

2. For what value(s) of h will $\mathbf{y}$ be in $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ if

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 5 \\ -4 \\ -7 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} -4 \\ 3 \\ h \end{bmatrix}$$

3. Let $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3$, $\mathbf{u}$, and $\mathbf{v}$ be vectors in $\mathbb{R}^n$. Suppose the vectors $\mathbf{u}$ and $\mathbf{v}$ are in Span $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$. Show that $\mathbf{u} + \mathbf{v}$ is also in Span $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$. [*Hint:* The solution requires the use of the definition of the span of a set of vectors. It is useful to review this definition before starting this exercise.]

1.3 Exercises

In Exercises 1 and 2, compute $\mathbf{u} + \mathbf{v}$ and $\mathbf{u} - 2\mathbf{v}$.

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

2. $\mathbf{u} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$

In Exercises 3 and 4, display the following vectors using arrows on an xy-graph: $\mathbf{u}$, $\mathbf{v}$, $-\mathbf{v}$, $-2\mathbf{v}$, $\mathbf{u} + \mathbf{v}$, $\mathbf{u} - \mathbf{v}$, and $\mathbf{u} - 2\mathbf{v}$. Notice that $\mathbf{u} - \mathbf{v}$ is the vertex of a parallelogram whose other vertices are $\mathbf{u}$, $\mathbf{0}$, and $-\mathbf{v}$.

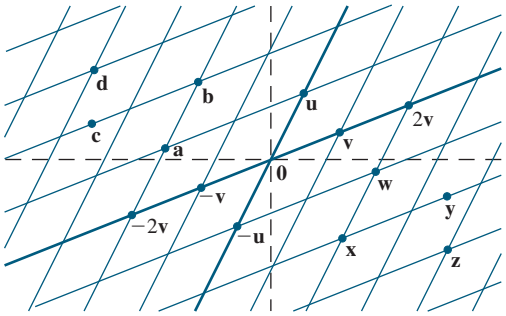
3. $\mathbf{u}$ and $\mathbf{v}$ as in Exercise 1 **4.** $\mathbf{u}$ and $\mathbf{v}$ as in Exercise 2

In Exercises 5 and 6, write a system of equations that is equivalent to the given vector equation.

5. $x_1 \begin{bmatrix} 4 \\ -3 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} -8 \\ 7 \\ 0 \end{bmatrix} = \begin{bmatrix} 9 \\ -6 \\ -5 \end{bmatrix}$

6. $x_1 \begin{bmatrix} -2 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 8 \\ 5 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ -6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Use the accompanying figure to write each vector listed in Exercises 7 and 8 as a linear combination of $\mathbf{u}$ and $\mathbf{v}$. Is every vector in $\mathbb{R}^2$ a linear combination of $\mathbf{u}$ and $\mathbf{v}$?



7. Vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$

8. Vectors $\mathbf{w}$, $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$

In Exercises 9 and 10, write a vector equation that is equivalent to the given system of equations.

9.
$$\begin{aligned} x_2 + 5x_3 &= 0 \\ 4x_1 + 6x_2 - x_3 &= 0 \\ -x_1 + 3x_2 - 8x_3 &= 0 \end{aligned}$$

10.
$$\begin{aligned} 4x_1 + x_2 + 3x_3 &= 9 \\ x_1 - 7x_2 - 2x_3 &= 2 \\ 8x_1 + 6x_2 - 5x_3 &= 15 \end{aligned}$$

In Exercises 11 and 12, determine if $\mathbf{b}$ is a linear combination of $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$.

11. $\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \mathbf{a}_3 = \begin{bmatrix} 5 \\ -6 \\ 8 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix}$

12. $\mathbf{a}_1 = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 0 \\ 5 \\ 5 \end{bmatrix}, \mathbf{a}_3 = \begin{bmatrix} 2 \\ 0 \\ 8 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} -5 \\ 11 \\ -7 \end{bmatrix}$

In Exercises 13 and 14, determine if $\mathbf{b}$ is a linear combination of the vectors formed from the columns of the matrix A.

13. $A = \begin{bmatrix} 1 & -4 & 2 \\ 0 & 3 & 5 \\ -2 & 8 & -4 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 3 \\ -7 \\ -3 \end{bmatrix}$

14. $A = \begin{bmatrix} 1 & -2 & -6 \\ 0 & 3 & 7 \\ 1 & -2 & 5 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 11 \\ -5 \\ 9 \end{bmatrix}$

In Exercises 15 and 16, list five vectors in $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$. For each vector, show the weights on $\mathbf{v}_1$ and $\mathbf{v}_2$ used to generate the vector and list the three entries of the vector. Do not make a sketch.

15. $\mathbf{v}_1 = \begin{bmatrix} 7 \\ 1 \\ -6 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -5 \\ 3 \\ 0 \end{bmatrix}$

16. $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -2 \\ 0 \\ 3 \end{bmatrix}$

17. Let $\mathbf{a}_1 = \begin{bmatrix} 1 \\ 4 \\ -2 \end{bmatrix}$, $\mathbf{a}_2 = \begin{bmatrix} -2 \\ -3 \\ 7 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} 4 \\ 1 \\ h \end{bmatrix}$. For what value(s) of h is $\mathbf{b}$ in the plane spanned by $\mathbf{a}_1$ and $\mathbf{a}_2$?

18. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ -4 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -5 \\ 1 \\ 7 \end{bmatrix}$, and $\mathbf{y} = \begin{bmatrix} h \\ -1 \\ -5 \end{bmatrix}$. For what value(s) of h is $\mathbf{y}$ in the plane generated by $\mathbf{v}_1$ and $\mathbf{v}_2$?

19. Give a geometric description of Span $\{\mathbf{v}_1, \mathbf{v}_2\}$ for the vectors $\mathbf{v}_1 = \begin{bmatrix} 8 \\ 2 \\ -6 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 12 \\ 3 \\ -9 \end{bmatrix}$.

20. Give a geometric description of Span $\{\mathbf{v}_1, \mathbf{v}_2\}$ for the vectors in Exercise 16.

21. Let $\mathbf{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Show that $\begin{bmatrix} h \\ k \end{bmatrix}$ is in Span $\{\mathbf{u}, \mathbf{v}\}$ for all h and k.

22. Construct a 3×3 matrix A, with nonzero entries, and a vector $\mathbf{b}$ in $\mathbb{R}^3$ such that $\mathbf{b}$ is *not* in the set spanned by the columns of A.

In Exercises 23–32, mark each statement True or False **(T/F)**. Justify each answer.

23. **(T/F)** Another notation for the vector $\begin{bmatrix} -4 \\ 3 \end{bmatrix}$ is $[-4 \;\; 3]$.

24. **(T/F)** Any list of five real numbers is a vector in $\mathbb{R}^5$.

25. **(T/F)** The points in the plane corresponding to $\begin{bmatrix} -2 \\ 5 \end{bmatrix}$ and $\begin{bmatrix} -5 \\ 2 \end{bmatrix}$ lie on a line through the origin.

26. **(T/F)** The vector $\mathbf{u}$ results when a vector $\mathbf{u} - \mathbf{v}$ is added to the vector $\mathbf{v}$.

27. **(T/F)** An example of a linear combination of vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ is the vector $\frac{1}{2}\mathbf{v}_1$.

28. **(T/F)** The weights $c_1, \ldots, c_p$ in a linear combination $c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p$ cannot all be zero.

29. **(T/F)** The solution set of the linear system whose augmented matrix is $[\, \mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \mathbf{a}_3 \;\; \mathbf{b} \,]$ is the same as the solution set of the equation $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3 = \mathbf{b}$.

30. **(T/F)** When $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors, Span $\{\mathbf{u}, \mathbf{v}\}$ contains the line through $\mathbf{u}$ and the origin.

31. **(T/F)** The set Span $\{\mathbf{u}, \mathbf{v}\}$ is always visualized as a plane through the origin.

32. **(T/F)** Asking whether the linear system corresponding to an augmented matrix $[\, \mathbf{a}_1 \;\; \mathbf{a}_2 \;\; \mathbf{a}_3 \;\; \mathbf{b} \,]$ has a solution amounts to asking whether $\mathbf{b}$ is in Span $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$.

33. Let $A = \begin{bmatrix} 1 & 0 & -4 \\ 0 & 3 & -2 \\ -2 & 6 & 3 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 4 \\ 1 \\ -4 \end{bmatrix}$. Denote the columns of A by $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$, and let $W = \text{Span}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$.

a. Is $\mathbf{b}$ in $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$? How many vectors are in $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$?

b. Is $\mathbf{b}$ in W? How many vectors are in W?

c. Show that $\mathbf{a}_1$ is in W. [*Hint:* Row operations are unnecessary.]

34. Let $A = \begin{bmatrix} 2 & 0 & 6 \\ -1 & 8 & 5 \\ 1 & -2 & 1 \end{bmatrix}$, let $\mathbf{b} = \begin{bmatrix} 10 \\ 3 \\ 3 \end{bmatrix}$, and let W be the set of all linear combinations of the columns of A.

a. Is $\mathbf{b}$ in W?

b. Show that the third column of A is in W.

35. A mining company has two mines. One day's operation at mine 1 produces ore that contains 20 metric tons of copper and 550 kilograms of silver, while one day's operation at mine 2 produces ore that contains 30 metric tons of copper and 500 kilograms of silver. Let $\mathbf{v}_1 = \begin{bmatrix} 20 \\ 550 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 30 \\ 500 \end{bmatrix}$. Then $\mathbf{v}_1$ and $\mathbf{v}_2$ represent the "output per day" of mine 1 and mine 2, respectively.

a. What physical interpretation can be given to the vector $5\mathbf{v}_1$?

b. Suppose the company operates mine 1 for x_1 days and mine 2 for x_2 days. Write a vector equation whose solution gives the number of days each mine should operate in order to produce 150 tons of copper and 2825 kilograms of silver. Do not solve the equation.

T c. Solve the equation in (b).

36. A steam plant burns two types of coal: anthracite (A) and bituminous (B). For each ton of A burned, the plant produces 27.6 million Btu of heat, 3100 grams (g) of sulfur dioxide, and 250 g of particulate matter (solid-particle pollutants). For each ton of B burned, the plant produces 30.2 million Btu, 6400 g of sulfur dioxide, and 360 g of particulate matter.

a. How much heat does the steam plant produce when it burns x_1 tons of A and x_2 tons of B?

b. Suppose the output of the steam plant is described by a vector that lists the amounts of heat, sulfur dioxide, and particulate matter. Express this output as a linear combination of two vectors, assuming that the plant burns x_1 tons of A and x_2 tons of B.

T c. Over a certain time period, the steam plant produced 162 million Btu of heat, 23,610 g of sulfur dioxide, and 1623 g of particulate matter. Determine how many tons of each type of coal the steam plant must have burned. Include a vector equation as part of your solution.

37. Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be points in $\mathbb{R}^3$ and suppose that for $j = 1, \ldots, k$ an object with mass m_j is located at point $\mathbf{v}_j$. Physicists call such objects *point masses*. The total mass of the system of point masses is

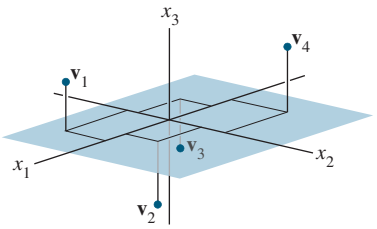
$$m = m_1 + \cdots + m_k$$

The *center of mass* (or *center of gravity*) of the system is

$$\overline{\mathbf{v}} = \frac{1}{m}[m_1\mathbf{v}_1 + \cdots + m_k\mathbf{v}_k]$$

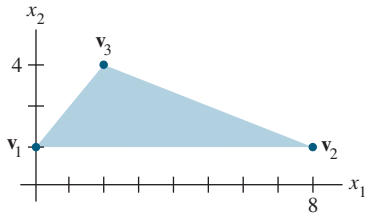
Compute the center of gravity of the system consisting of the following point masses (see the figure):

Point	Mass
$\mathbf{v}_1 = (5, -4, 3)$	2 g
$\mathbf{v}_2 = (4, 3, -2)$	5 g
$\mathbf{v}_3 = (-4, -3, -1)$	2 g
$\mathbf{v}_4 = (-9, 8, 6)$	1 g



38. Let $\mathbf{v}$ be the center of mass of a system of point masses located at $\mathbf{v}_1, \ldots, \mathbf{v}_k$ as in Exercise 37. Is $\mathbf{v}$ in Span $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$? Explain.

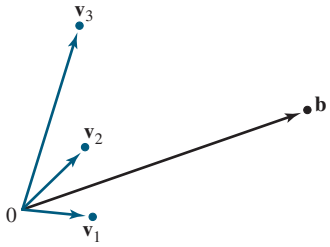
39. A thin triangular plate of uniform density and thickness has vertices at $\mathbf{v}_1 = (0, 1)$, $\mathbf{v}_2 = (8, 1)$, and $\mathbf{v}_3 = (2, 4)$, as in the figure below, and the mass of the plate is 3 g.



a. Find the (x, y)-coordinates of the center of mass of the plate. This "balance point" of the plate coincides with the center of mass of a system consisting of three 1-gram point masses located at the vertices of the plate.

b. Determine how to distribute an additional mass of 6 g at the three vertices of the plate to move the balance point of the plate to $(2, 2)$. [*Hint:* Let w_1, w_2, and w_3 denote the masses added at the three vertices, so that $w_1 + w_2 + w_3 = 6$.]

40. Consider the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$, and $\mathbf{b}$ in $\mathbb{R}^2$, shown in the figure. Does the equation $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{b}$ have a solution? Is the solution unique? Use the figure to explain your answers.



41. Use the vectors $\mathbf{u} = (u_1, \ldots, u_n)$, $\mathbf{v} = (v_1, \ldots, v_n)$, and $\mathbf{w} = (w_1, \ldots, w_n)$ to verify the following algebraic properties of $\mathbb{R}^n$.

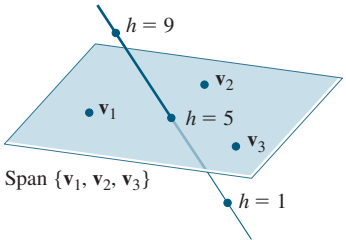
a. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$

b. $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$ for each scalar c

42. Use the vector $\mathbf{u} = (u_1, \ldots, u_n)$ to verify the following algebraic properties of $\mathbb{R}^n$.

a. $\mathbf{u} + (-\mathbf{u}) = (-\mathbf{u}) + \mathbf{u} = 0$

b. $c(d\mathbf{u}) = (cd)\mathbf{u}$ for all scalars c and d



The points $\begin{bmatrix} -4 \\ 3 \\ h \end{bmatrix}$ lie on a line that intersects the plane when $h = 5$.

Solutions to Practice Problems

1. Take arbitrary vectors $\mathbf{u} = (u_1, \ldots, u_n)$ and $\mathbf{v} = (v_1, \ldots, v_n)$ in $\mathbb{R}^n$, and compute

$$\begin{aligned} \mathbf{u} + \mathbf{v} &= (u_1 + v_1, \ldots, u_n + v_n) && \text{Definition of vector addition} \\ &= (v_1 + u_1, \ldots, v_n + u_n) && \text{Commutativity of addition in } \mathbb{R} \\ &= \mathbf{v} + \mathbf{u} && \text{Definition of vector addition} \end{aligned}$$

2. The vector $\mathbf{y}$ belongs to Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ if and only if there exist scalars x_1, x_2, x_3 such that

$$x_1\begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix} + x_2\begin{bmatrix} 5 \\ -4 \\ -7 \end{bmatrix} + x_3\begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -4 \\ 3 \\ h \end{bmatrix}$$

This vector equation is equivalent to a system of three linear equations in three unknowns. If you row reduce the augmented matrix for this system, you find that

$$\begin{bmatrix} 1 & 5 & -3 & -4 \\ -1 & -4 & 1 & 3 \\ -2 & -7 & 0 & h \end{bmatrix} \sim \begin{bmatrix} 1 & 5 & -3 & -4 \\ 0 & 1 & -2 & -1 \\ 0 & 3 & -6 & h-8 \end{bmatrix} \sim \begin{bmatrix} 1 & 5 & -3 & -4 \\ 0 & 1 & -2 & -1 \\ 0 & 0 & 0 & h-5 \end{bmatrix}$$

The system is consistent if and only if there is no pivot in the fourth column. That is, $h - 5$ must be 0. So $\mathbf{y}$ is in Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ if and only if $h = 5$.

Remember: The presence of a free variable in a system does not guarantee that the system is consistent.

3. Since the vectors $\mathbf{u}$ and $\mathbf{v}$ are in Span $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$, there exist scalars c_1, c_2, c_3 and d_1, d_2, d_3 such that

$$\mathbf{u} = c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + c_3\mathbf{w}_3 \quad \text{and} \quad \mathbf{v} = d_1\mathbf{w}_1 + d_2\mathbf{w}_2 + d_3\mathbf{w}_3.$$

Notice

$$\begin{aligned} \mathbf{u} + \mathbf{v} &= c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + c_3\mathbf{w}_3 + d_1\mathbf{w}_1 + d_2\mathbf{w}_2 + d_3\mathbf{w}_3 \\ &= (c_1 + d_1)\mathbf{w}_1 + (c_2 + d_2)\mathbf{w}_2 + (c_3 + d_3)\mathbf{w}_3 \end{aligned}$$

Since $c_1 + d_1, c_2 + d_2$, and $c_3 + d_3$ are also scalars, the vector $\mathbf{u} + \mathbf{v}$ is in Span $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\}$.

1.4 The Matrix Equation $A\mathbf{x} = \mathbf{b}$

A fundamental idea in linear algebra is to view a linear combination of vectors as the product of a matrix and a vector. The following definition permits us to rephrase some of the concepts of Section 1.3 in new ways.

DEFINITION

If A is an $m \times n$ matrix, with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$, and if $\mathbf{x}$ is in $\mathbb{R}^n$, then the **product of A and $\mathbf{x}$**, denoted by $A\mathbf{x}$, is **the linear combination of the columns of A using the corresponding entries in $\mathbf{x}$ as weights**; that is,

$$A\mathbf{x} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$$

Note that $A\mathbf{x}$ is defined only if the number of columns of A equals the number of entries in $\mathbf{x}$.

EXAMPLE 1

a. $$\begin{bmatrix} 1 & 2 & -1 \\ 0 & -5 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \\ 7 \end{bmatrix} = 4\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3\begin{bmatrix} 2 \\ -5 \end{bmatrix} + 7\begin{bmatrix} -1 \\ 3 \end{bmatrix}$$
$$= \begin{bmatrix} 4 \\ 0 \end{bmatrix} + \begin{bmatrix} 6 \\ -15 \end{bmatrix} + \begin{bmatrix} -7 \\ 21 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$$

b. $$\begin{bmatrix} 2 & -3 \\ 8 & 0 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 7 \end{bmatrix} = 4\begin{bmatrix} 2 \\ 8 \\ -5 \end{bmatrix} + 7\begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 32 \\ -20 \end{bmatrix} + \begin{bmatrix} -21 \\ 0 \\ 14 \end{bmatrix} = \begin{bmatrix} -13 \\ 32 \\ -6 \end{bmatrix}$$ ■

EXAMPLE 2 For $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ in $\mathbb{R}^m$, write the linear combination $3\mathbf{v}_1 - 5\mathbf{v}_2 + 7\mathbf{v}_3$ as a matrix times a vector.

SOLUTION Place $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ into the columns of a matrix A and place the weights 3, -5, and 7 into a vector $\mathbf{x}$. That is,

$$3\mathbf{v}_1 - 5\mathbf{v}_2 + 7\mathbf{v}_3 = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} \begin{bmatrix} 3 \\ -5 \\ 7 \end{bmatrix} = A\mathbf{x}$$

■

Section 1.3 showed how to write a system of linear equations as a vector equation involving a linear combination of vectors. For example, the system

$$\begin{aligned} x_1 + 2x_2 - x_3 &= 4 \\ -5x_2 + 3x_3 &= 1 \end{aligned} \tag{1}$$

is equivalent to

$$x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ -5 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix} \tag{2}$$

As in Example 2, the linear combination on the left side is a matrix times a vector, so that (2) becomes

$$\begin{bmatrix} 1 & 2 & -1 \\ 0 & -5 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \end{bmatrix} \tag{3}$$

Equation (3) has the form $A\mathbf{x} = \mathbf{b}$. Such an equation is called a **matrix equation**, to distinguish it from a vector equation such as is shown in (2).

Notice how the matrix in (3) is just the matrix of coefficients of the system (1). Similar calculations show that any system of linear equations, or any vector equation such as (2), can be written as an equivalent matrix equation in the form $A\mathbf{x} = \mathbf{b}$. This simple observation will be used repeatedly throughout the text.

Here is the formal result.

THEOREM 3

If A is an $m \times n$ matrix, with columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$, and if $\mathbf{b}$ is in $\mathbb{R}^m$, the matrix equation

$$A\mathbf{x} = \mathbf{b} \tag{4}$$

has the same solution set as the vector equation

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b} \tag{5}$$

which, in turn, has the same solution set as the system of linear equations whose augmented matrix is

$$\begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n & \mathbf{b} \end{bmatrix} \tag{6}$$

Theorem 3 provides a powerful tool for gaining insight into problems in linear algebra, because a system of linear equations may now be viewed in three different but equivalent ways: as a matrix equation, as a vector equation, or as a system of linear equations. Whenever you construct a mathematical model of a problem in real life, you are free to choose whichever viewpoint is most natural. Then you may switch from one formulation of a problem to another whenever it is convenient. In any case, the matrix equation (4), the vector equation (5), and the system of equations are all solved in the same way—by row reducing the augmented matrix (6). Other methods of solution will be discussed later.

Existence of Solutions

The definition of $A\mathbf{x}$ leads directly to the following useful fact.

The equation $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is a linear combination of the columns of A.

Section 1.3 considered the existence question, "Is $\mathbf{b}$ in Span $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$?" Equivalently, "Is $A\mathbf{x} = \mathbf{b}$ consistent?" A harder existence problem is to determine whether the equation $A\mathbf{x} = \mathbf{b}$ is consistent *for all* possible $\mathbf{b}$.

EXAMPLE 3 Let $A = \begin{bmatrix} 1 & 3 & 4 \\ -4 & 2 & -6 \\ -3 & -2 & -7 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$. Is the equation $A\mathbf{x} = \mathbf{b}$ consistent for all possible b_1, b_2, b_3?

SOLUTION Row reduce the augmented matrix for $A\mathbf{x} = \mathbf{b}$:

$$\begin{bmatrix} 1 & 3 & 4 & b_1 \\ -4 & 2 & -6 & b_2 \\ -3 & -2 & -7 & b_3 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 & 4 & b_1 \\ 0 & 14 & 10 & b_2 + 4b_1 \\ 0 & 7 & 5 & b_3 + 3b_1 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 3 & 4 & b_1 \\ 0 & 14 & 10 & b_2 + 4b_1 \\ 0 & 0 & 0 & b_3 + 3b_1 - \frac{1}{2}(b_2 + 4b_1) \end{bmatrix}$$

The third entry in column 4 equals $b_1 - \frac{1}{2}b_2 + b_3$. The equation $A\mathbf{x} = \mathbf{b}$ is *not* consistent for every $\mathbf{b}$ because some choices of $\mathbf{b}$ can make $b_1 - \frac{1}{2}b_2 + b_3$ nonzero. ■

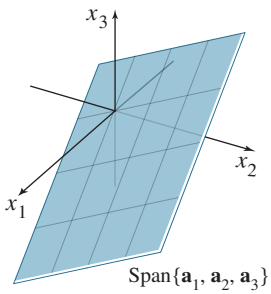


FIGURE 1
The columns of $A = [\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3\,]$ span a plane through $\mathbf{0}$.

The reduced matrix in Example 3 provides a description of all $\mathbf{b}$ for which the equation $A\mathbf{x} = \mathbf{b}$ *is* consistent: The entries in $\mathbf{b}$ must satisfy

$$b_1 - \tfrac{1}{2}b_2 + b_3 = 0$$

This is the equation of a plane through the origin in $\mathbb{R}^3$. The plane is the set of all linear combinations of the three columns of A. See Figure 1.

The equation $A\mathbf{x} = \mathbf{b}$ in Example 3 fails to be consistent for all $\mathbf{b}$ because the echelon form of A has a row of zeros. If A had a pivot in all three rows, we would not care about the calculations in the augmented column because in this case an echelon form of the augmented matrix could not have a row such as $[\,0 \quad 0 \quad 0 \quad 1\,]$.

In the next theorem, the sentence "The columns of A span $\mathbb{R}^m$" means that *every* $\mathbf{b}$ in $\mathbb{R}^m$ is a linear combination of the columns of A. In general, a set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ in $\mathbb{R}^m$ **spans** (or **generates**) $\mathbb{R}^m$ if every vector in $\mathbb{R}^m$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_p$—that is, if Span $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\} = \mathbb{R}^m$.

THEOREM 4

Let A be an $m \times n$ matrix. Then the following statements are logically equivalent. That is, for a particular A, either they are all true statements or they are all false.

a. For each $\mathbf{b}$ in $\mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a solution.
b. Each $\mathbf{b}$ in $\mathbb{R}^m$ is a linear combination of the columns of A.
c. The columns of A span $\mathbb{R}^m$.
d. A has a pivot position in every row.

Theorem 4 is one of the most useful theorems in this chapter. Statements (a), (b), and (c) are equivalent because of the definition of $A\mathbf{x}$ and what it means for a set of vectors to span $\mathbb{R}^m$. The discussion after Example 3 suggests why (a) and (d) are equivalent; a proof is given at the end of the section. The exercises will provide examples of how Theorem 4 is used.

Warning: Theorem 4 is about a *coefficient matrix*, not an augmented matrix. If an augmented matrix $[\,A \quad \mathbf{b}\,]$ has a pivot position in every row, then the equation $A\mathbf{x} = \mathbf{b}$ may or may not be consistent.

Computation of *Ax*

The calculations in Example 1 were based on the definition of the product of a matrix A and a vector $\mathbf{x}$. The following simple example will lead to a more efficient method for calculating the entries in $A\mathbf{x}$ when working problems by hand.

EXAMPLE 4 Compute $A\mathbf{x}$, where $A = \begin{bmatrix} 2 & 3 & 4 \\ -1 & 5 & -3 \\ 6 & -2 & 8 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$.

SOLUTION From the definition,

$$\begin{bmatrix} 2 & 3 & 4 \\ -1 & 5 & -3 \\ 6 & -2 & 8 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_1\begin{bmatrix} 2 \\ -1 \\ 6 \end{bmatrix} + x_2\begin{bmatrix} 3 \\ 5 \\ -2 \end{bmatrix} + x_3\begin{bmatrix} 4 \\ -3 \\ 8 \end{bmatrix}$$

$$= \begin{bmatrix} 2x_1 \\ -x_1 \\ 6x_1 \end{bmatrix} + \begin{bmatrix} 3x_2 \\ 5x_2 \\ -2x_2 \end{bmatrix} + \begin{bmatrix} 4x_3 \\ -3x_3 \\ 8x_3 \end{bmatrix} \tag{7}$$

$$= \begin{bmatrix} 2x_1 + 3x_2 + 4x_3 \\ -x_1 + 5x_2 - 3x_3 \\ 6x_1 - 2x_2 + 8x_3 \end{bmatrix}$$

The first entry in the product $A\mathbf{x}$ is a sum of products (sometimes called a *dot product*), using the first row of A and the entries in $\mathbf{x}$. That is,

$$\begin{bmatrix} 2 & 3 & 4 \\ & & \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2x_1 + 3x_2 + 4x_3 \\ \end{bmatrix}$$

This matrix shows how to compute the first entry in $A\mathbf{x}$ directly, without writing down all the calculations shown in (7). Similarly, the second entry in $A\mathbf{x}$ can be calculated at once by multiplying the entries in the second row of A by the corresponding entries in $\mathbf{x}$ and then summing the resulting products:

$$\begin{bmatrix} -1 & 5 & -3 \\ & & \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -x_1 + 5x_2 - 3x_3 \\ \end{bmatrix}$$

Likewise, the third entry in $A\mathbf{x}$ can be calculated from the third row of A and the entries in $\mathbf{x}$. ■

Row–Vector Rule for Computing *Ax*

If the product $A\mathbf{x}$ is defined, then the ith entry in $A\mathbf{x}$ is the sum of the products of corresponding entries from row i of A and from the vector $\mathbf{x}$.

EXAMPLE 5

a. $\begin{bmatrix} 1 & 2 & -1 \\ 0 & -5 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \\ 7 \end{bmatrix} = \begin{bmatrix} 1 \cdot 4 + 2 \cdot 3 + (-1) \cdot 7 \\ 0 \cdot 4 + (-5) \cdot 3 + 3 \cdot 7 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix}$

b. $\begin{bmatrix} 2 & -3 \\ 8 & 0 \\ -5 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 7 \end{bmatrix} = \begin{bmatrix} 2 \cdot 4 + (-3) \cdot 7 \\ 8 \cdot 4 + 0 \cdot 7 \\ (-5) \cdot 4 + 2 \cdot 7 \end{bmatrix} = \begin{bmatrix} -13 \\ 32 \\ -6 \end{bmatrix}$

c. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r \\ s \\ t \end{bmatrix} = \begin{bmatrix} 1 \cdot r + 0 \cdot s + 0 \cdot t \\ 0 \cdot r + 1 \cdot s + 0 \cdot t \\ 0 \cdot r + 0 \cdot s + 1 \cdot t \end{bmatrix} = \begin{bmatrix} r \\ s \\ t \end{bmatrix}$ ■

By definition, the matrix in Example 5(c) with 1's on the diagonal and 0's elsewhere is called an **identity matrix** and is denoted by I. The calculation in part (c) shows that $I\mathbf{x} = \mathbf{x}$ for every $\mathbf{x}$ in $\mathbb{R}^3$. There is an analogous $n \times n$ identity matrix, sometimes written as I_n. As in part (c), $I_n\mathbf{x} = \mathbf{x}$ for every $\mathbf{x}$ in $\mathbb{R}^n$.

Properties of the Matrix–Vector Product $A\mathbf{x}$

The facts in the next theorem are important and will be used throughout the text. The proof relies on the definition of $A\mathbf{x}$ and the algebraic properties of $\mathbb{R}^n$.

THEOREM 5

If A is an $m \times n$ matrix, $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^n$, and c is a scalar, then:

a. $A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v}$;

b. $A(c\mathbf{u}) = c(A\mathbf{u})$.

PROOF For simplicity, take $n = 3$, $A = [\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3\,]$, and $\mathbf{u}$, $\mathbf{v}$ in $\mathbb{R}^3$. (The proof of the general case is similar.) For $i = 1, 2, 3$, let u_i and v_i be the ith entries in $\mathbf{u}$ and $\mathbf{v}$, respectively. To prove statement (a), compute $A(\mathbf{u} + \mathbf{v})$ as a linear combination of the columns of A using the entries in $\mathbf{u} + \mathbf{v}$ as weights.

$$A(\mathbf{u} + \mathbf{v}) = [\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3\,] \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ u_3 + v_3 \end{bmatrix}$$

$$= (u_1 + v_1)\mathbf{a}_1 + (u_2 + v_2)\mathbf{a}_2 + (u_3 + v_3)\mathbf{a}_3$$

(Entries in $\mathbf{u} + \mathbf{v}$; Columns of A)

$$= (u_1\mathbf{a}_1 + u_2\mathbf{a}_2 + u_3\mathbf{a}_3) + (v_1\mathbf{a}_1 + v_2\mathbf{a}_2 + v_3\mathbf{a}_3)$$

$$= A\mathbf{u} + A\mathbf{v}$$

To prove statement (b), compute $A(c\mathbf{u})$ as a linear combination of the columns of A using the entries in $c\mathbf{u}$ as weights.

$$A(c\mathbf{u}) = [\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3\,] \begin{bmatrix} cu_1 \\ cu_2 \\ cu_3 \end{bmatrix} = (cu_1)\mathbf{a}_1 + (cu_2)\mathbf{a}_2 + (cu_3)\mathbf{a}_3$$

$$= c(u_1\mathbf{a}_1) + c(u_2\mathbf{a}_2) + c(u_3\mathbf{a}_3)$$

$$= c(u_1\mathbf{a}_1 + u_2\mathbf{a}_2 + u_3\mathbf{a}_3)$$

$$= c(A\mathbf{u})$$

■

Numerical Note

To optimize a computer algorithm to compute $A\mathbf{x}$, the sequence of calculations should involve data stored in contiguous memory locations. The most widely used professional algorithms for matrix computations are written in Fortran, a language that stores a matrix as a set of columns. Such algorithms compute $A\mathbf{x}$ as a linear combination of the columns of A. In contrast, if a program is written in the popular language C, which stores matrices by rows, $A\mathbf{x}$ should be computed via the alternative rule that uses the rows of A.

PROOF OF THEOREM 4 As was pointed out after Theorem 4, statements (a), (b), and (c) are logically equivalent. So, it suffices to show (for an arbitrary matrix A) that (a) and (d) are either both true or both false. This will tie all four statements together.

Let U be an echelon form of A. Given $\mathbf{b}$ in $\mathbb{R}^m$, we can row reduce the augmented matrix $[\,A \quad \mathbf{b}\,]$ to an augmented matrix $[\,U \quad \mathbf{d}\,]$ for some $\mathbf{d}$ in $\mathbb{R}^m$:

$$[\,A \quad \mathbf{b}\,] \sim \cdots \sim [\,U \quad \mathbf{d}\,]$$

If statement (d) is true, then each row of U contains a pivot position and there can be no pivot in the augmented column. So $A\mathbf{x} = \mathbf{b}$ has a solution for any $\mathbf{b}$, and (a) is true. If (d) is false, the last row of U is all zeros. Let $\mathbf{d}$ be any vector with a 1 in its last entry. Then $[\,U \quad \mathbf{d}\,]$ represents an *inconsistent* system. Since row operations are reversible, $[\,U \quad \mathbf{d}\,]$ can be transformed into the form $[\,A \quad \mathbf{b}\,]$. The new system $A\mathbf{x} = \mathbf{b}$ is also inconsistent, and (a) is false. ■

Practice Problems

1. Let $A = \begin{bmatrix} 1 & 5 & -2 & 0 \\ -3 & 1 & 9 & -5 \\ 4 & -8 & -1 & 7 \end{bmatrix}$, $\mathbf{p} = \begin{bmatrix} 3 \\ -2 \\ 0 \\ -4 \end{bmatrix}$, and $\mathbf{b} = \begin{bmatrix} -7 \\ 9 \\ 0 \end{bmatrix}$. It can be shown that $\mathbf{p}$ is a solution of $A\mathbf{x} = \mathbf{b}$. Use this fact to exhibit $\mathbf{b}$ as a specific linear combination of the columns of A.

2. Let $A = \begin{bmatrix} 2 & 5 \\ 3 & 1 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} -3 \\ 5 \end{bmatrix}$. Verify Theorem 5(a) in this case by computing $A(\mathbf{u} + \mathbf{v})$ and $A\mathbf{u} + A\mathbf{v}$.

3. Construct a 3×3 matrix A and vectors $\mathbf{b}$ and $\mathbf{c}$ in $\mathbb{R}^3$ so that $A\mathbf{x} = \mathbf{b}$ has a solution, but $A\mathbf{x} = \mathbf{c}$ does not.

1.4 Exercises

Compute the products in Exercises 1–4 using (a) the definition, as in Example 1, and (b) the row–vector rule for computing $A\mathbf{x}$. If a product is undefined, explain why.

1. $\begin{bmatrix} -4 & 2 \\ 1 & 6 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ 1 \\ 7 \end{bmatrix}$

2. $\begin{bmatrix} 2 \\ 6 \\ -1 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$

3. $\begin{bmatrix} 6 & 5 \\ -4 & -3 \\ 7 & 6 \end{bmatrix}\begin{bmatrix} 1 \\ -3 \end{bmatrix}$

4. $\begin{bmatrix} 8 & 3 & 1 \\ 5 & 1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

In Exercises 5–8, use the definition of $A\mathbf{x}$ to write the matrix equation as a vector equation, or vice versa.

5. $\begin{bmatrix} 7 & 2 & -9 & 3 \\ -4 & -5 & 7 & -2 \end{bmatrix}\begin{bmatrix} 6 \\ -9 \\ 1 \\ -8 \end{bmatrix} = \begin{bmatrix} -9 \\ 44 \end{bmatrix}$

6. $\begin{bmatrix} 7 & -3 \\ 2 & 1 \\ 9 & -6 \\ -3 & 2 \end{bmatrix}\begin{bmatrix} -2 \\ -5 \end{bmatrix} = \begin{bmatrix} 1 \\ -9 \\ 12 \\ -4 \end{bmatrix}$

7. $x_1\begin{bmatrix}4\\-1\\7\\-4\end{bmatrix} + x_2\begin{bmatrix}-5\\3\\-5\\1\end{bmatrix} + x_3\begin{bmatrix}7\\-8\\0\\2\end{bmatrix} = \begin{bmatrix}6\\-8\\0\\-7\end{bmatrix}$

8. $z_1\begin{bmatrix}4\\-2\end{bmatrix} + z_2\begin{bmatrix}-4\\5\end{bmatrix} + z_3\begin{bmatrix}-5\\4\end{bmatrix} + z_4\begin{bmatrix}3\\0\end{bmatrix} = \begin{bmatrix}4\\13\end{bmatrix}$

In Exercises 9 and 10, write the system first as a vector equation and then as a matrix equation.

9.
$$\begin{aligned} 4x_1 + x_2 - 7x_3 &= 8\\ x_2 + 6x_3 &= 0\end{aligned}$$

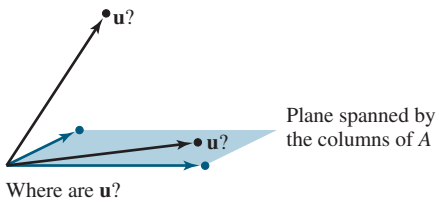
10.
$$\begin{aligned} 8x_1 - x_2 &= 4\\ 5x_1 + 4x_2 &= 1\\ x_1 - 3x_2 &= 2\end{aligned}$$

Given A and $\mathbf{b}$ in Exercises 11 and 12, write the augmented matrix for the linear system that corresponds to the matrix equation $A\mathbf{x} = \mathbf{b}$. Then solve the system and write the solution as a vector.

11. $A = \begin{bmatrix}1&2&4\\0&1&5\\-2&-4&-3\end{bmatrix}$, $\mathbf{b} = \begin{bmatrix}-2\\2\\9\end{bmatrix}$

12. $A = \begin{bmatrix}1&2&1\\-3&-1&2\\0&5&3\end{bmatrix}$, $\mathbf{b} = \begin{bmatrix}0\\1\\-1\end{bmatrix}$

13. Let $\mathbf{u} = \begin{bmatrix}0\\4\\4\end{bmatrix}$ and $A = \begin{bmatrix}3&-5\\-2&6\\1&1\end{bmatrix}$. Is $\mathbf{u}$ in the plane in $\mathbb{R}^3$ spanned by the columns of A? (See the figure.) Why or why not?



14. Let $\mathbf{u} = \begin{bmatrix}2\\-3\\2\end{bmatrix}$ and $A = \begin{bmatrix}5&8&7\\0&1&-1\\1&3&0\end{bmatrix}$. Is $\mathbf{u}$ in the subset of $\mathbb{R}^3$ spanned by the columns of A? Why or why not?

15. Let $A = \begin{bmatrix}3&-4\\-6&8\end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix}b_1\\b_2\end{bmatrix}$. Show that the equation $A\mathbf{x} = \mathbf{b}$ does not have a solution for all possible $\mathbf{b}$, and describe the set of all $\mathbf{b}$ for which $A\mathbf{x} = \mathbf{b}$ *does* have a solution.

16. Repeat Exercise 15: $A = \begin{bmatrix}1&-3&-4\\-3&2&6\\5&-1&-8\end{bmatrix}$, $\mathbf{b} = \begin{bmatrix}b_1\\b_2\\b_3\end{bmatrix}$.

Exercises 17–20 refer to the matrices A and B below. Make appropriate calculations that justify your answers and mention an appropriate theorem.

$$A = \begin{bmatrix}1&3&0&3\\-1&-1&-1&1\\0&-4&2&-8\\2&0&3&-1\end{bmatrix} \qquad B = \begin{bmatrix}1&3&-2&2\\0&1&1&-5\\1&2&-3&7\\-2&-8&2&-1\end{bmatrix}$$

17. How many rows of A contain a pivot position? Does the equation $A\mathbf{x} = \mathbf{b}$ have a solution for each $\mathbf{b}$ in $\mathbb{R}^4$?

18. Do the columns of B span $\mathbb{R}^4$? Does the equation $B\mathbf{x} = \mathbf{y}$ have a solution for each $\mathbf{y}$ in $\mathbb{R}^4$?

19. Can each vector in $\mathbb{R}^4$ be written as a linear combination of the columns of the matrix A above? Do the columns of A span $\mathbb{R}^4$?

20. Can every vector in $\mathbb{R}^4$ be written as a linear combination of the columns of the matrix B above? Do the columns of B span $\mathbb{R}^3$?

21. Let $\mathbf{v}_1 = \begin{bmatrix}1\\0\\-1\\0\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}0\\-1\\0\\1\end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix}1\\0\\0\\-1\end{bmatrix}$.
Does $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ span $\mathbb{R}^4$? Why or why not?

22. Let $\mathbf{v}_1 = \begin{bmatrix}0\\0\\-2\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}0\\-3\\8\end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix}4\\-1\\-5\end{bmatrix}$.
Does $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ span $\mathbb{R}^3$? Why or why not?

In Exercises 23–34, mark each statement True or False **(T/F)**. Justify each answer.

23. (T/F) The equation $A\mathbf{x} = \mathbf{b}$ is referred to as a *vector equation*.

24. (T/F) Every matrix equation $A\mathbf{x} = \mathbf{b}$ corresponds to a vector equation with the same solution set.

25. (T/F) If the equation $A\mathbf{x} = \mathbf{b}$ is inconsistent, then $\mathbf{b}$ is not in the set spanned by the columns of A.

26. (T/F) A vector $\mathbf{b}$ is a linear combination of the columns of a matrix A if and only if the equation $A\mathbf{x} = \mathbf{b}$ has at least one solution.

27. (T/F) The equation $A\mathbf{x} = \mathbf{b}$ is consistent if the augmented matrix $[\,A \quad \mathbf{b}\,]$ has a pivot position in every row.

28. (T/F) If A is an $m \times n$ matrix whose columns do not span $\mathbb{R}^m$, then the equation $A\mathbf{x} = \mathbf{b}$ is inconsistent for some $\mathbf{b}$ in $\mathbb{R}^m$.

29. (T/F) The first entry in the product $A\mathbf{x}$ is a sum of products.

30. (T/F) Any linear combination vectors can always be written in the form $A\mathbf{x}$ for a suitable matrix A and vector $\mathbf{x}$.

31. (T/F) If the columns of an $m \times n$ matrix A span $\mathbb{R}^m$, then the equation $A\mathbf{x} = \mathbf{b}$ is consistent for each $\mathbf{b}$ in $\mathbb{R}^m$.

32. (T/F) The solution set of a linear system whose augmented matrix is $[\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3 \quad \mathbf{b}\,]$ is the same as the solution set of $A\mathbf{x} = \mathbf{b}$, if $A = [\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3\,]$.

33. **(T/F)** If A is an $m \times n$ matrix and if the equation $A\mathbf{x} = \mathbf{b}$ is inconsistent for some $\mathbf{b}$ in $\mathbb{R}^m$, then A cannot have a pivot position in every row.

34. **(T/F)** If the augmented matrix $[\,A \quad \mathbf{b}\,]$ has a pivot position in every row, then the equation $A\mathbf{x} = \mathbf{b}$ is inconsistent.

35. Note that $\begin{bmatrix} 3 & -4 & 2 \\ 6 & -3 & 4 \\ -8 & 9 & -5 \end{bmatrix}\begin{bmatrix} -4 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} -2 \\ -9 \\ 8 \end{bmatrix}$. Use this fact (and no row operations) to find scalars c_1, c_2, c_3 such that $\begin{bmatrix} -2 \\ -9 \\ 8 \end{bmatrix} = c_1\begin{bmatrix} 3 \\ 6 \\ -8 \end{bmatrix} + c_2\begin{bmatrix} -4 \\ -3 \\ 9 \end{bmatrix} + c_3\begin{bmatrix} 2 \\ 4 \\ -5 \end{bmatrix}$.

36. Let $\mathbf{u} = \begin{bmatrix} 7 \\ 2 \\ 5 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 3 \\ 1 \\ 3 \end{bmatrix}$, and $\mathbf{w} = \begin{bmatrix} 6 \\ 1 \\ 0 \end{bmatrix}$. It can be shown that $3\mathbf{u} - 5\mathbf{v} - \mathbf{w} = \mathbf{0}$. Use this fact (and no row operations) to find x_1 and x_2 that satisfy the equation $\begin{bmatrix} 7 & 3 \\ 2 & 1 \\ 5 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 1 \\ 0 \end{bmatrix}$.

37. Let $\mathbf{q}_1$, $\mathbf{q}_2$, $\mathbf{q}_3$, and $\mathbf{v}$ represent vectors in $\mathbb{R}^5$, and let x_1, x_2, and x_3 denote scalars. Write the following vector equation as a matrix equation. Identify any symbols you choose to use.

$$x_1\mathbf{q}_1 + x_2\mathbf{q}_2 + x_3\mathbf{q}_3 = \mathbf{v}$$

38. Rewrite the (numerical) matrix equation below in symbolic form as a vector equation, using symbols $\mathbf{v}_1, \mathbf{v}_2, \ldots$ for the vectors and $c_1, c_2, \ldots$ for scalars. Define what each symbol represents, using the data given in the matrix equation.

$$\begin{bmatrix} -3 & 5 & -4 & 9 & 7 \\ 5 & 8 & 1 & -2 & -4 \end{bmatrix}\begin{bmatrix} -3 \\ 2 \\ 4 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ -1 \end{bmatrix}$$

39. Construct a 3×3 matrix, not in echelon form, whose columns span $\mathbb{R}^3$. Show that the matrix you construct has the desired property.

40. Construct a 3×3 matrix, not in echelon form, whose columns do *not* span $\mathbb{R}^3$. Show that the matrix you construct has the desired property.

STUDY GUIDE offers additional resources for mastering the concept of span.

41. Let A be a 3×2 matrix. Explain why the equation $A\mathbf{x} = \mathbf{b}$ cannot be consistent for all $\mathbf{b}$ in $\mathbb{R}^3$. Generalize your argument to the case of an arbitrary A with more rows than columns.

42. Could a set of three vectors in $\mathbb{R}^4$ span all of $\mathbb{R}^4$? Explain. What about n vectors in $\mathbb{R}^m$ when n is less than m?

43. Suppose A is a 4×3 matrix and $\mathbf{b}$ is a vector in $\mathbb{R}^4$ with the property that $A\mathbf{x} = \mathbf{b}$ has a unique solution. What can you say about the reduced echelon form of A? Justify your answer.

44. Suppose A is a 3×3 matrix and $\mathbf{b}$ is a vector in $\mathbb{R}^3$ with the property that $A\mathbf{x} = \mathbf{b}$ has a unique solution. Explain why the columns of A must span $\mathbb{R}^3$.

45. Let A be a 3×4 matrix, let $\mathbf{y}_1$ and $\mathbf{y}_2$ be vectors in $\mathbb{R}^3$, and let $\mathbf{w} = \mathbf{y}_1 + \mathbf{y}_2$. Suppose $\mathbf{y}_1 = A\mathbf{x}_1$ and $\mathbf{y}_2 = A\mathbf{x}_2$ for some vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ in $\mathbb{R}^4$. What fact allows you to conclude that the system $A\mathbf{x} = \mathbf{w}$ is consistent? (*Note:* $\mathbf{x}_1$ and $\mathbf{x}_2$ denote vectors, not scalar entries in vectors.)

46. Let A be a 5×3 matrix, let $\mathbf{y}$ be a vector in $\mathbb{R}^3$, and let $\mathbf{z}$ be a vector in $\mathbb{R}^5$. Suppose $A\mathbf{y} = \mathbf{z}$. What fact allows you to conclude that the system $A\mathbf{x} = 4\mathbf{z}$ is consistent?

T In Exercises 47–50, determine if the columns of the matrix span $\mathbb{R}^4$.

47. $\begin{bmatrix} 7 & 2 & -5 & 8 \\ -5 & -3 & 4 & -9 \\ 6 & 10 & -2 & 7 \\ -7 & 9 & 2 & 15 \end{bmatrix}$

48. $\begin{bmatrix} 5 & -7 & -4 & 9 \\ 6 & -8 & -7 & 5 \\ 4 & -4 & -9 & -9 \\ -9 & 11 & 16 & 7 \end{bmatrix}$

49. $\begin{bmatrix} 12 & -7 & 11 & -9 & 5 \\ -9 & 4 & -8 & 7 & -3 \\ -6 & 11 & -7 & 3 & -9 \\ 4 & -6 & 10 & -5 & 12 \end{bmatrix}$

50. $\begin{bmatrix} 8 & 11 & -6 & -7 & 13 \\ -7 & -8 & 5 & 6 & -9 \\ 11 & 7 & -7 & -9 & -6 \\ -3 & 4 & 1 & 8 & 7 \end{bmatrix}$

T **51.** Find a column of the matrix in Exercise 49 that can be deleted and yet have the remaining matrix columns still span $\mathbb{R}^4$.

T **52.** Find a column of the matrix in Exercise 50 that can be deleted and yet have the remaining matrix columns still span $\mathbb{R}^4$. Can you delete more than one column?

Solutions to Practice Problems

1. The matrix equation

$$\begin{bmatrix} 1 & 5 & -2 & 0 \\ -3 & 1 & 9 & -5 \\ 4 & -8 & -1 & 7 \end{bmatrix}\begin{bmatrix} 3 \\ -2 \\ 0 \\ -4 \end{bmatrix} = \begin{bmatrix} -7 \\ 9 \\ 0 \end{bmatrix}$$

is equivalent to the vector equation

$$3\begin{bmatrix}1\\-3\\4\end{bmatrix}-2\begin{bmatrix}5\\1\\-8\end{bmatrix}+0\begin{bmatrix}-2\\9\\-1\end{bmatrix}-4\begin{bmatrix}0\\-5\\7\end{bmatrix}=\begin{bmatrix}-7\\9\\0\end{bmatrix},$$

which expresses $\mathbf{b}$ as a linear combination of the columns of A.

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

$$A(\mathbf{u}+\mathbf{v})=\begin{bmatrix}2&5\\3&1\end{bmatrix}\begin{bmatrix}1\\4\end{bmatrix}=\begin{bmatrix}2+20\\3+4\end{bmatrix}=\begin{bmatrix}22\\7\end{bmatrix}$$

$$A\mathbf{u}+A\mathbf{v}=\begin{bmatrix}2&5\\3&1\end{bmatrix}\begin{bmatrix}4\\-1\end{bmatrix}+\begin{bmatrix}2&5\\3&1\end{bmatrix}\begin{bmatrix}-3\\5\end{bmatrix}$$

$$=\begin{bmatrix}3\\11\end{bmatrix}+\begin{bmatrix}19\\-4\end{bmatrix}=\begin{bmatrix}22\\7\end{bmatrix}$$

Remark: There are, in fact, infinitely many correct solutions to Practice Problem 3. When creating matrices to satisfy specified criteria, it is often useful to create matrices that are straightforward, such as those already in reduced echelon form. Here is one possible solution:

3. Let

$$A=\begin{bmatrix}1&0&1\\0&1&1\\0&0&0\end{bmatrix},\ \mathbf{b}=\begin{bmatrix}3\\2\\0\end{bmatrix},\text{ and }\mathbf{c}=\begin{bmatrix}3\\2\\1\end{bmatrix}.$$

Notice the reduced echelon form of the augmented matrix corresponding to $A\mathbf{x}=\mathbf{b}$ is

$$\begin{bmatrix}1&0&1&3\\0&1&1&2\\0&0&0&0\end{bmatrix},$$

which corresponds to a consistent system, and hence $A\mathbf{x}=\mathbf{b}$ has solutions. The reduced echelon form of the augmented matrix corresponding to $A\mathbf{x}=\mathbf{c}$ is

$$\begin{bmatrix}1&0&1&3\\0&1&1&2\\0&0&0&1\end{bmatrix},$$

which corresponds to an inconsistent system, and hence $A\mathbf{x}=\mathbf{c}$ does not have any solutions.

1.5 Solution Sets of Linear Systems

Solution sets of linear systems are important objects of study in linear algebra. They will appear later in several different contexts. This section uses vector notation to give explicit and geometric descriptions of such solution sets.

Homogeneous Linear Systems

A system of linear equations is said to be **homogeneous** if it can be written in the form $A\mathbf{x} = \mathbf{0}$, where A is an $m \times n$ matrix and $\mathbf{0}$ is the zero vector in $\mathbb{R}^m$. Such a system $A\mathbf{x} = \mathbf{0}$ *always* has at least one solution, namely $\mathbf{x} = \mathbf{0}$ (the zero vector in $\mathbb{R}^n$). This zero solution is usually called the **trivial solution**. For a given equation $A\mathbf{x} = \mathbf{0}$, the important question is whether there exists a **nontrivial solution**, that is, a nonzero vector $\mathbf{x}$ that satisfies $A\mathbf{x} = \mathbf{0}$. The Existence and Uniqueness Theorem in Section 1.2 (Theorem 2) leads immediately to the following fact.

The homogeneous equation $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution if and only if the equation has at least one free variable.

EXAMPLE 1 Determine if the following homogeneous system has a nontrivial solution. Then describe the solution set.

$$\begin{aligned} 3x_1 + 5x_2 - 4x_3 &= 0 \\ -3x_1 - 2x_2 + 4x_3 &= 0 \\ 6x_1 + x_2 - 8x_3 &= 0 \end{aligned}$$

SOLUTION Let A be the matrix of coefficients of the system and row reduce the augmented matrix $[\,A \quad \mathbf{0}\,]$ to echelon form:

$$\begin{bmatrix} 3 & 5 & -4 & 0 \\ -3 & -2 & 4 & 0 \\ 6 & 1 & -8 & 0 \end{bmatrix} \sim \begin{bmatrix} 3 & 5 & -4 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & -9 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 3 & 5 & -4 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Since x_3 is a free variable, $A\mathbf{x} = \mathbf{0}$ has nontrivial solutions (one for each nonzero choice of x_3). To describe the solution set, continue the row reduction of $[\,A \quad \mathbf{0}\,]$ to *reduced* echelon form:

$$\begin{bmatrix} 1 & 0 & -\frac{4}{3} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{aligned} x_1 - \tfrac{4}{3}x_3 &= 0 \\ x_2 &= 0 \\ 0 &= 0 \end{aligned}$$

Solve for the basic variables x_1 and x_2 and obtain $x_1 = \frac{4}{3}x_3$, $x_2 = 0$, with x_3 free. As a vector, the general solution of $A\mathbf{x} = \mathbf{0}$ has the form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{4}{3}x_3 \\ 0 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix} = x_3\mathbf{v}, \quad \text{where } \mathbf{v} = \begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix}$$

Here x_3 is factored out of the expression for the general solution vector. This shows that every solution of $A\mathbf{x} = \mathbf{0}$ in this case is a scalar multiple of $\mathbf{v}$. The trivial solution is obtained by choosing $x_3 = 0$. Geometrically, the solution set is a line through $\mathbf{0}$ in $\mathbb{R}^3$. See Figure 1. ■

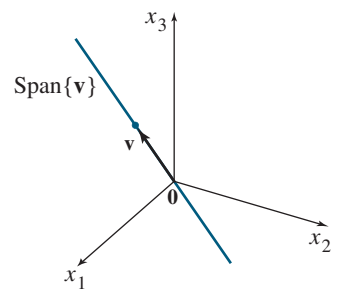


FIGURE 1

Notice that a nontrivial solution $\mathbf{x}$ can have some zero entries so long as not all of its entries are zero.

EXAMPLE 2 A single linear equation can be treated as a very simple system of equations. Describe all solutions of the homogeneous "system"

$$10x_1 - 3x_2 - 2x_3 = 0 \tag{1}$$

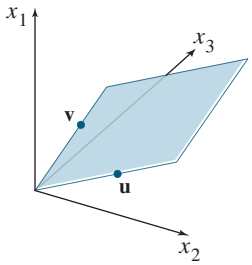


FIGURE 2

SOLUTION There is no need for matrix notation. Solve for the basic variable x_1 in terms of the free variables. The general solution is $x_1 = .3x_2 + .2x_3$, with x_2 and x_3 free. As a vector, the general solution is

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} .3x_2 + .2x_3 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} .3x_2 \\ x_2 \\ 0 \end{bmatrix} + \begin{bmatrix} .2x_3 \\ 0 \\ x_3 \end{bmatrix}$$

$$= x_2 \begin{bmatrix} .3 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} .2 \\ 0 \\ 1 \end{bmatrix} \quad (\text{with } x_2, x_3 \text{ free}) \qquad (2)$$

$\mathbf{u}$ $\mathbf{v}$

This calculation shows that every solution of (1) is a linear combination of the vectors $\mathbf{u}$ and $\mathbf{v}$, shown in (2). That is, the solution set is Span $\{\mathbf{u}, \mathbf{v}\}$. Since neither $\mathbf{u}$ nor $\mathbf{v}$ is a scalar multiple of the other, the solution set is a plane through the origin. See Figure 2. ■

Examples 1 and 2, along with the exercises, illustrate the fact that the solution set of a homogeneous equation $A\mathbf{x} = \mathbf{0}$ can always be expressed explicitly as Span $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ for suitable vectors $\mathbf{v}_1, \ldots, \mathbf{v}_p$. If the only solution is the zero vector, then the solution set is Span $\{\mathbf{0}\}$. If the equation $A\mathbf{x} = \mathbf{0}$ has only one free variable, the solution set is a line through the origin, as in Figure 1. A plane through the origin, as in Figure 2, provides a good mental image for the solution set of $A\mathbf{x} = \mathbf{0}$ when there are two or more free variables. Note, however, that a similar figure can be used to visualize Span $\{\mathbf{u}, \mathbf{v}\}$ even when $\mathbf{u}$ and $\mathbf{v}$ do not arise as solutions of $A\mathbf{x} = \mathbf{0}$. See Figure 11 in Section 1.3.

Parametric Vector Form

The original equation (1) for the plane in Example 2 is an *implicit* description of the plane. Solving this equation amounts to finding an *explicit* description of the plane as the set spanned by $\mathbf{u}$ and $\mathbf{v}$. Equation (2) is called a **parametric vector equation** of the plane. Sometimes such an equation is written as

$$\mathbf{x} = s\mathbf{u} + t\mathbf{v} \quad (s, t \text{ in } \mathbb{R})$$

to emphasize that the parameters vary over all real numbers. In Example 1, the equation $\mathbf{x} = x_3\mathbf{v}$ (with x_3 free), or $\mathbf{x} = t\mathbf{v}$ (with t in $\mathbb{R}$), is a parametric vector equation of a line. Whenever a solution set is described explicitly with vectors as in Examples 1 and 2, we say that the solution is in **parametric vector form**.

Solutions of Nonhomogeneous Systems

When a nonhomogeneous linear system has many solutions, the general solution can be written in parametric vector form as one vector plus an arbitrary linear combination of vectors that satisfy the corresponding homogeneous system.

EXAMPLE 3 Describe all solutions of $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 3 & 5 & -4 \\ -3 & -2 & 4 \\ 6 & 1 & -8 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 7 \\ -1 \\ -4 \end{bmatrix}$$

SOLUTION Here A is the matrix of coefficients from Example 1. Row operations on $[\,A \quad \mathbf{b}\,]$ produce

$$\begin{bmatrix} 3 & 5 & -4 & 7 \\ -3 & -2 & 4 & -1 \\ 6 & 1 & -8 & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & -\frac{4}{3} & -1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \begin{aligned} x_1 \quad & -\tfrac{4}{3}x_3 = -1 \\ x_2 \quad & = 2 \\ 0 \quad & = 0 \end{aligned}$$

Thus $x_1 = -1 + \frac{4}{3}x_3$, $x_2 = 2$, and x_3 is free. As a vector, the general solution of $A\mathbf{x} = \mathbf{b}$ has the form

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 + \frac{4}{3}x_3 \\ 2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{4}{3}x_3 \\ 0 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} \frac{4}{3} \\ 0 \\ 1 \end{bmatrix}$$

$\mathbf{p}$ $\mathbf{v}$

The equation $\mathbf{x} = \mathbf{p} + x_3\mathbf{v}$, or, writing t as a general parameter,

$$\mathbf{x} = \mathbf{p} + t\mathbf{v} \quad (t \text{ in } \mathbb{R}) \tag{3}$$

describes the solution set of $A\mathbf{x} = \mathbf{b}$ in parametric vector form. Recall from Example 1 that the solution set of $A\mathbf{x} = \mathbf{0}$ has the parametric vector equation

$$\mathbf{x} = t\mathbf{v} \quad (t \text{ in } \mathbb{R}) \tag{4}$$

[with the same $\mathbf{v}$ that appears in (3)]. Thus the solutions of $A\mathbf{x} = \mathbf{b}$ are obtained by adding the vector $\mathbf{p}$ to the solutions of $A\mathbf{x} = \mathbf{0}$. The vector $\mathbf{p}$ itself is just one particular solution of $A\mathbf{x} = \mathbf{b}$ [corresponding to $t = 0$ in (3)]. ■

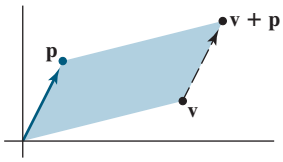


FIGURE 3
Adding $\mathbf{p}$ to $\mathbf{v}$ translates $\mathbf{v}$ to $\mathbf{v} + \mathbf{p}$.

To describe the solution set of $A\mathbf{x} = \mathbf{b}$ geometrically, we can think of vector addition as a *translation*. Given $\mathbf{v}$ and $\mathbf{p}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$, the effect of adding $\mathbf{p}$ to $\mathbf{v}$ is to *move* $\mathbf{v}$ in a direction parallel to the line through $\mathbf{p}$ and $\mathbf{0}$. We say that $\mathbf{v}$ is **translated by p** to $\mathbf{v} + \mathbf{p}$. See Figure 3. If each point on a line L in $\mathbb{R}^2$ or $\mathbb{R}^3$ is translated by a vector $\mathbf{p}$, the result is a line parallel to L. See Figure 4.

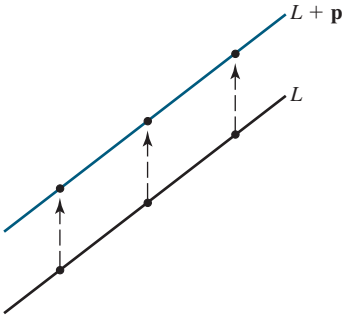


FIGURE 4
Translated line.

Suppose L is the line through $\mathbf{0}$ and $\mathbf{v}$, described by equation (4). Adding $\mathbf{p}$ to each point on L produces the translated line described by equation (3). Note that $\mathbf{p}$ is on the line in equation (3). We call (3) **the equation of the line through p parallel to v**. Thus *the solution set of* $A\mathbf{x} = \mathbf{b}$ *is a line through* $\mathbf{p}$ *parallel to the solution set of* $A\mathbf{x} = \mathbf{0}$. Figure 5 illustrates this case.

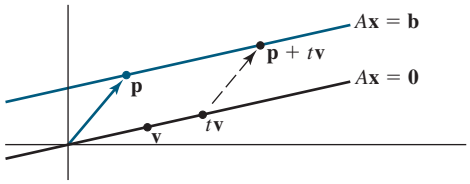


FIGURE 5 Parallel solution sets of $A\mathbf{x} = \mathbf{b}$ and $A\mathbf{x} = \mathbf{0}$.

The relation between the solution sets of $A\mathbf{x} = \mathbf{b}$ and $A\mathbf{x} = \mathbf{0}$ shown in Figure 5 generalizes to any *consistent* equation $A\mathbf{x} = \mathbf{b}$, although the solution set will be larger than a line when there are several free variables. The following theorem gives the precise statement. See Exercise 37 at the end of this section for a proof.

THEOREM 6

Suppose the equation $A\mathbf{x} = \mathbf{b}$ is consistent for some given $\mathbf{b}$, and let $\mathbf{p}$ be a solution. Then the solution set of $A\mathbf{x} = \mathbf{b}$ is the set of all vectors of the form $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$, where $\mathbf{v}_h$ is any solution of the homogeneous equation $A\mathbf{x} = \mathbf{0}$.

Theorem 6 says that if $A\mathbf{x} = \mathbf{b}$ has a solution, then the solution set is obtained by translating the solution set of $A\mathbf{x} = \mathbf{0}$, using any particular solution $\mathbf{p}$ of $A\mathbf{x} = \mathbf{b}$ for the translation. Figure 6 illustrates the case in which there are two free variables. Even when $n > 3$, our mental image of the solution set of a consistent system $A\mathbf{x} = \mathbf{b}$ (with $\mathbf{b} \neq \mathbf{0}$) is either a single nonzero point or a line or plane not passing through the origin.

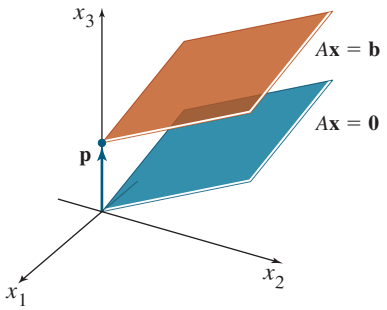


FIGURE 6 Parallel solution sets of $A\mathbf{x} = \mathbf{b}$ and $A\mathbf{x} = \mathbf{0}$.

Warning: Theorem 6 and Figure 6 apply only to an equation $A\mathbf{x} = \mathbf{b}$ that has at least one nonzero solution $\mathbf{p}$. When $A\mathbf{x} = \mathbf{b}$ has no solution, the solution set is empty.

The following algorithm outlines the calculations shown in Examples 1, 2, and 3.

WRITING A SOLUTION SET (OF A CONSISTENT SYSTEM) IN PARAMETRIC VECTOR FORM

1. Row reduce the augmented matrix to reduced echelon form.
2. Express each basic variable in terms of any free variables appearing in an equation.
3. Write a typical solution $\mathbf{x}$ as a vector whose entries depend on the free variables, if any.
4. Decompose $\mathbf{x}$ into a linear combination of vectors (with numeric entries) using the free variables as parameters.

Reasonable Answers

To verify that the solutions you found are indeed solutions to the homogeneous equation $A\mathbf{x} = \mathbf{0}$, simply multiply the matrix by each vector in your solution and check that the result is the zero vector. For example, if $A = \begin{bmatrix} 1 & -2 & 1 & 2 \\ 1 & -1 & 2 & 5 \\ 0 & 1 & 1 & 3 \end{bmatrix}$, and you found the homogeneous solutions to

Reasonable Answers (Continued)

be $x_3\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix}+x_4\begin{bmatrix}-8\\-3\\0\\1\end{bmatrix}$, check $\begin{bmatrix}1&-2&1&2\\1&-1&2&5\\0&1&1&3\end{bmatrix}\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$ and $\begin{bmatrix}1&-2&1&2\\1&-1&2&5\\0&1&1&3\end{bmatrix}\begin{bmatrix}-8\\-3\\0\\1\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$. Then $A\left(x_3\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix}+x_4\begin{bmatrix}-8\\-3\\0\\1\end{bmatrix}\right)$ $=x_3A\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix}+x_4A\begin{bmatrix}-8\\-3\\0\\1\end{bmatrix}$, which is equal to $x_3\begin{bmatrix}0\\0\\0\end{bmatrix}+x_4\begin{bmatrix}0\\0\\0\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$, as desired.

If you are solving $A\mathbf{x}=\mathbf{b}$, then you can again verify that you have correct solutions by multiplying the matrix by each vector in your solutions. The product of A with the first vector (the one that is *not* part of the solution to the homogeneous equation) should be $\mathbf{b}$. The product of A with the remaining vectors (the ones that are part of the solution to the homogeneous equation) should of course be $\mathbf{0}$.

For example, to verify that $\begin{bmatrix}2\\1\\1\\2\end{bmatrix}+x_3\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix}+x_4\begin{bmatrix}-8\\-3\\0\\1\end{bmatrix}$ are solutions to $A\mathbf{x}=\begin{bmatrix}5\\13\\8\end{bmatrix}$, check $\begin{bmatrix}1&-2&1&2\\1&-1&2&5\\0&1&1&3\end{bmatrix}\begin{bmatrix}2\\1\\1\\2\end{bmatrix}=\begin{bmatrix}5\\13\\8\end{bmatrix}$, and use the calculations from above. Notice $A\left(\begin{bmatrix}2\\1\\1\\2\end{bmatrix}+x_3\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix}+x_4\begin{bmatrix}-8\\-3\\0\\1\end{bmatrix}\right)$ $=A\begin{bmatrix}2\\1\\1\\2\end{bmatrix}+x_3A\begin{bmatrix}-3\\-1\\1\\0\end{bmatrix}+x_4A\begin{bmatrix}-8\\-3\\0\\1\end{bmatrix}$, which is equal to $\begin{bmatrix}5\\13\\8\end{bmatrix}+x_3\begin{bmatrix}0\\0\\0\end{bmatrix}$ $+x_4\begin{bmatrix}0\\0\\0\end{bmatrix}=\begin{bmatrix}5\\13\\8\end{bmatrix}$, as desired.

Practice Problems

1. Each of the following equations determines a plane in $\mathbb{R}^3$. Do the two planes intersect? If so, describe their intersection.

$$\begin{aligned} x_1 + 4x_2 - 5x_3 &= 0 \\ 2x_1 - x_2 + 8x_3 &= 9 \end{aligned}$$

2. Write the general solution of $10x_1 - 3x_2 - 2x_3 = 7$ in parametric vector form, and relate the solution set to the one found in Example 2.

3. Prove the first part of Theorem 6: Suppose that $\mathbf{p}$ is a solution of $A\mathbf{x} = \mathbf{b}$, so that $A\mathbf{p} = \mathbf{b}$. Let $\mathbf{v}_h$ be any solution to the homogeneous equation $A\mathbf{x} = \mathbf{0}$, and let $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$. Show that $\mathbf{w}$ is a solution to $A\mathbf{x} = \mathbf{b}$.

1.5 Exercises

In Exercises 1–4, determine if the system has a nontrivial solution. Try to use as few row operations as possible.

1.
$$\begin{aligned} 2x_1 - 5x_2 + 8x_3 &= 0 \\ -2x_1 - 7x_2 + x_3 &= 0 \\ 4x_1 + 2x_2 + 7x_3 &= 0 \end{aligned}$$

2.
$$\begin{aligned} x_1 - 3x_2 + 7x_3 &= 0 \\ -2x_1 + x_2 - 4x_3 &= 0 \\ x_1 + 2x_2 + 9x_3 &= 0 \end{aligned}$$

3.
$$\begin{aligned} -3x_1 + 5x_2 - 7x_3 &= 0 \\ -6x_1 + 7x_2 + x_3 &= 0 \end{aligned}$$

4.
$$\begin{aligned} -5x_1 + 7x_2 + 9x_3 &= 0 \\ x_1 - 2x_2 + 6x_3 &= 0 \end{aligned}$$

In Exercises 5 and 6, follow the method of Examples 1 and 2 to write the solution set of the given homogeneous system in parametric vector form.

5.
$$\begin{aligned} x_1 + 3x_2 + x_3 &= 0 \\ -4x_1 - 9x_2 + 2x_3 &= 0 \\ - 3x_2 - 6x_3 &= 0 \end{aligned}$$

6.
$$\begin{aligned} x_1 + 3x_2 - 5x_3 &= 0 \\ x_1 + 4x_2 - 8x_3 &= 0 \\ -3x_1 - 7x_2 + 9x_3 &= 0 \end{aligned}$$

In Exercises 7–12, describe all solutions of $A\mathbf{x} = \mathbf{0}$ in parametric vector form, where A is row equivalent to the given matrix.

7. $\begin{bmatrix} 1 & 3 & -3 & 7 \\ 0 & 1 & -4 & 5 \end{bmatrix}$

8. $\begin{bmatrix} 1 & -2 & -9 & 5 \\ 0 & 1 & 2 & -6 \end{bmatrix}$

9. $\begin{bmatrix} 2 & -8 & 6 \\ -1 & 4 & -3 \end{bmatrix}$

10. $\begin{bmatrix} 1 & 3 & 0 & -4 \\ 2 & 6 & 0 & -8 \end{bmatrix}$

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

12. $\begin{bmatrix} 1 & 5 & 2 & -6 & 9 & 0 \\ 0 & 0 & 1 & -7 & 4 & -8 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

You may find it helpful to review the information in the Reasonable Answers box from this section before answering Exercises 13–16.

13. Verify that the solutions you found to Exercise 9 are indeed homogeneous solutions.

14. Verify that the solutions you found to Exercise 10 are indeed homogeneous solutions.

15. Verify that the solutions you found to Exercise 11 are indeed homogeneous solutions.

16. Verify that the solutions you found to Exercise 12 are indeed homogeneous solutions.

17. Suppose the solution set of a certain system of linear equations can be described as $x_1 = 5 + 4x_3$, $x_2 = -2 - 7x_3$, with x_3 free. Use vectors to describe this set as a line in $\mathbb{R}^3$.

18. Suppose the solution set of a certain system of linear equations can be described as $x_1 = 3x_4$, $x_2 = 8 + x_4$, $x_3 = 2 - 5x_4$, with x_4 free. Use vectors to describe this set as a line in $\mathbb{R}^4$.

19. Follow the method of Example 3 to describe the solutions of the following system in parametric vector form. Also, give a geometric description of the solution set and compare it to that in Exercise 5.

$$\begin{aligned} x_1 + 3x_2 + x_3 &= 1 \\ -4x_1 - 9x_2 + 2x_3 &= -1 \\ - 3x_2 - 6x_3 &= -3 \end{aligned}$$

20. As in Exercise 19, describe the solutions of the following system in parametric vector form, and provide a geometric comparison with the solution set in Exercise 6.

$$\begin{aligned} x_1 + 3x_2 - 5x_3 &= 4 \\ x_1 + 4x_2 - 8x_3 &= 7 \\ -3x_1 - 7x_2 + 9x_3 &= -6 \end{aligned}$$

21. Describe and compare the solution sets of $x_1 + 9x_2 - 4x_3 = 0$ and $x_1 + 9x_2 - 4x_3 = -2$.

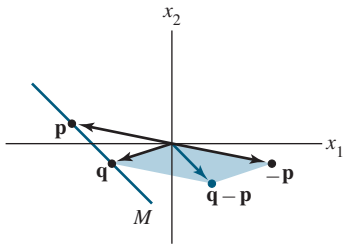
22. Describe and compare the solution sets of $x_1 - 3x_2 + 5x_3 = 0$ and $x_1 - 3x_2 + 5x_3 = 4$.

In Exercises 23 and 24, find the parametric equation of the line through **a** parallel to **b**.

23. $\mathbf{a} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} -5 \\ 3 \end{bmatrix}$ **24.** $\mathbf{a} = \begin{bmatrix} 5 \\ -2 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} -4 \\ 9 \end{bmatrix}$

In Exercises 25 and 26, find a parametric equation of the line M through **p** and **q**. [*Hint:* M is parallel to the vector $\mathbf{q} - \mathbf{p}$. See the figure below.]

25. $\mathbf{p} = \begin{bmatrix} 2 \\ -5 \end{bmatrix}, \mathbf{q} = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$ **26.** $\mathbf{p} = \begin{bmatrix} -6 \\ 3 \end{bmatrix}, \mathbf{q} = \begin{bmatrix} 0 \\ -4 \end{bmatrix}$



The line through **p** and **q**.

In Exercises 27–36, mark each statement True or False **(T/F)**. Justify each answer.

27. **(T/F)** A homogeneous equation is always consistent.

28. **(T/F)** If **x** is a nontrivial solution of $A\mathbf{x} = \mathbf{0}$, then every entry in **x** is nonzero.

29. **(T/F)** The equation $A\mathbf{x} = \mathbf{0}$ gives an explicit description of its solution set.

30. **(T/F)** The equation $\mathbf{x} = x_2\mathbf{u} + x_3\mathbf{v}$, with x_2 and x_3 free (and neither **u** nor **v** a multiple of the other), describes a plane through the origin.

31. **(T/F)** The homogeneous equation $A\mathbf{x} = \mathbf{0}$ has the trivial solution if and only if the equation has at least one free variable.

32. **(T/F)** The equation $A\mathbf{x} = \mathbf{b}$ is homogeneous if the zero vector is a solution.

33. **(T/F)** The equation $\mathbf{x} = \mathbf{p} + t\mathbf{v}$ describes a line through **v** parallel to **p**.

34. **(T/F)** The effect of adding **p** to a vector is to move the vector in a direction parallel to **p**.

35. **(T/F)** The solution set of $A\mathbf{x} = \mathbf{b}$ is the set of all vectors of the form $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$, where $\mathbf{v}_h$ is any solution of the equation $A\mathbf{x} = \mathbf{0}$.

36. **(T/F)** The solution set of $A\mathbf{x} = \mathbf{b}$ is obtained by translating the solution set of $A\mathbf{x} = \mathbf{0}$.

37. Prove the second part of Theorem 6: Let **w** be any solution of $A\mathbf{x} = \mathbf{b}$, and define $\mathbf{v}_h = \mathbf{w} - \mathbf{p}$. Show that $\mathbf{v}_h$ is a solution of $A\mathbf{x} = \mathbf{0}$. This shows that every solution of $A\mathbf{x} = \mathbf{b}$ has the form $\mathbf{w} = \mathbf{p} + \mathbf{v}_h$, with **p** a particular solution of $A\mathbf{x} = \mathbf{b}$ and $\mathbf{v}_h$ a solution of $A\mathbf{x} = \mathbf{0}$.

38. Suppose $A\mathbf{x} = \mathbf{b}$ has a solution. Explain why the solution is unique precisely when $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

39. Suppose A is the 3×3 *zero* matrix (with all zero entries). Describe the solution set of the equation $A\mathbf{x} = \mathbf{0}$.

40. If $\mathbf{b} \neq \mathbf{0}$, can the solution set of $A\mathbf{x} = \mathbf{b}$ be a plane through the origin? Explain.

In Exercises 41–44, (a) does the equation $A\mathbf{x} = \mathbf{0}$ have a nontrivial solution and (b) does the equation $A\mathbf{x} = \mathbf{b}$ have at least one solution for every possible **b**?

41. A is a 3×3 matrix with three pivot positions.

42. A is a 3×3 matrix with two pivot positions.

43. A is a 3×2 matrix with two pivot positions.

44. A is a 2×4 matrix with two pivot positions.

45. Given $A = \begin{bmatrix} -2 & -6 \\ 7 & 21 \\ -3 & -9 \end{bmatrix}$, find one nontrivial solution of $A\mathbf{x} = \mathbf{0}$ by inspection. [*Hint:* Think of the equation $A\mathbf{x} = \mathbf{0}$ written as a vector equation.]

46. Given $A = \begin{bmatrix} 4 & -6 \\ -8 & 12 \\ 6 & -9 \end{bmatrix}$, find one nontrivial solution of $A\mathbf{x} = \mathbf{0}$ by inspection.

47. Construct a 3×3 nonzero matrix A such that the vector $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is a solution of $A\mathbf{x} = \mathbf{0}$.

48. Construct a 3×3 nonzero matrix A such that the vector $\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ is a solution of $A\mathbf{x} = \mathbf{0}$.

49. Construct a 2×2 matrix A such that the solution set of the equation $A\mathbf{x} = \mathbf{0}$ is the line in $\mathbb{R}^2$ through $(4, 1)$ and the origin. Then, find a vector **b** in $\mathbb{R}^2$ such that the solution set of $A\mathbf{x} = \mathbf{b}$ is *not* a line in $\mathbb{R}^2$ parallel to the solution set of $A\mathbf{x} = \mathbf{0}$. Why does this *not* contradict Theorem 6?

50. Suppose A is a 3×3 matrix and **y** is a vector in $\mathbb{R}^3$ such that the equation $A\mathbf{x} = \mathbf{y}$ does *not* have a solution. Does there exist a vector **z** in $\mathbb{R}^3$ such that the equation $A\mathbf{x} = \mathbf{z}$ has a unique solution? Discuss.

51. Let A be an $m \times n$ matrix and let **u** be a vector in $\mathbb{R}^n$ that satisfies the equation $A\mathbf{x} = \mathbf{0}$. Show that for any scalar c, the vector $c\mathbf{u}$ also satisfies $A\mathbf{x} = \mathbf{0}$. [That is, show that $A(c\mathbf{u}) = \mathbf{0}$.]

52. Let A be an $m \times n$ matrix, and let **u** and **v** be vectors in $\mathbb{R}^n$ with the property that $A\mathbf{u} = \mathbf{0}$ and $A\mathbf{v} = \mathbf{0}$. Explain why $A(\mathbf{u} + \mathbf{v})$ must be the zero vector. Then explain why $A(c\mathbf{u} + d\mathbf{v}) = \mathbf{0}$ for each pair of scalars c and d.

1.7 Linear Independence

The homogeneous equations in Section 1.5 can be studied from a different perspective by writing them as vector equations. In this way, the focus shifts from the unknown solutions of $A\mathbf{x} = \mathbf{0}$ to the vectors that appear in the vector equations.

For instance, consider the equation

$$x_1\begin{bmatrix}1\\2\\3\end{bmatrix} + x_2\begin{bmatrix}4\\5\\6\end{bmatrix} + x_3\begin{bmatrix}2\\1\\0\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix} \tag{1}$$

This equation has a trivial solution, of course, where $x_1 = x_2 = x_3 = 0$. As in Section 1.5, the main issue is whether the trivial solution is the *only one*.

DEFINITION

An indexed set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ in $\mathbb{R}^n$ is said to be **linearly independent** if the vector equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_p\mathbf{v}_p = \mathbf{0}$$

has only the trivial solution. The set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is said to be **linearly dependent** if there exist weights $c_1, \ldots, c_p$, not all zero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0} \tag{2}$$

Equation (2) is called a **linear dependence relation** among $\mathbf{v}_1, \ldots, \mathbf{v}_p$ when the weights are not all zero. An indexed set is linearly dependent if and only if it is not linearly independent. For brevity, we may say that $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are linearly dependent when we mean that $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a linearly dependent set. We use analogous terminology for linearly independent sets.

EXAMPLE 1 Let $\mathbf{v}_1 = \begin{bmatrix}1\\2\\3\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}4\\5\\6\end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix}2\\1\\0\end{bmatrix}$.

a. Determine if the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly independent.

b. If possible, find a linear dependence relation among $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$.

SOLUTION

a. We must determine if there is a nontrivial solution of equation (1) above. Row operations on the associated augmented matrix show that

$$\begin{bmatrix}1 & 4 & 2 & 0\\2 & 5 & 1 & 0\\3 & 6 & 0 & 0\end{bmatrix} \sim \begin{bmatrix}1 & 4 & 2 & 0\\0 & -3 & -3 & 0\\0 & 0 & 0 & 0\end{bmatrix}$$

Clearly, x_1 and x_2 are basic variables, and x_3 is free. Each nonzero value of x_3 determines a nontrivial solution of (1). Hence $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are linearly dependent (and not linearly independent).

b. To find a linear dependence relation among $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$, completely row reduce the augmented matrix and write the new system:

$$\begin{bmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{aligned} x_1 \quad - 2x_3 &= 0 \\ x_2 + \; x_3 &= 0 \\ 0 &= 0 \end{aligned}$$

Thus $x_1 = 2x_3$, $x_2 = -x_3$, and x_3 is free. Choose any nonzero value for x_3—say, $x_3 = 5$. Then $x_1 = 10$ and $x_2 = -5$. Substitute these values into equation (1) and obtain

$$10\mathbf{v}_1 - 5\mathbf{v}_2 + 5\mathbf{v}_3 = \mathbf{0}$$

This is one (out of infinitely many) possible linear dependence relations among $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. ■

Linear Independence of Matrix Columns

Suppose that we begin with a matrix $A = [\,\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_n\,]$ instead of a set of vectors. The matrix equation $A\mathbf{x} = \mathbf{0}$ can be written as

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{0}$$

Each linear dependence relation among the columns of A corresponds to a nontrivial solution of $A\mathbf{x} = \mathbf{0}$. Thus we have the following important fact.

The columns of a matrix A are linearly independent if and only if the equation $A\mathbf{x} = \mathbf{0}$ has *only* the trivial solution. (3)

EXAMPLE 2 Determine if the columns of the matrix $A = \begin{bmatrix} 0 & 1 & 4 \\ 1 & 2 & -1 \\ 5 & 8 & 0 \end{bmatrix}$ are linearly independent.

SOLUTION To study $A\mathbf{x} = \mathbf{0}$, row reduce the augmented matrix:

$$\begin{bmatrix} 0 & 1 & 4 & 0 \\ 1 & 2 & -1 & 0 \\ 5 & 8 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 4 & 0 \\ 0 & -2 & 5 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 4 & 0 \\ 0 & 0 & 13 & 0 \end{bmatrix}$$

At this point, it is clear that there are three basic variables and no free variables. So the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, and the columns of A are linearly independent. ■

Sets of One or Two Vectors

A set containing only one vector—say, $\mathbf{v}$—is linearly independent if and only if $\mathbf{v}$ is not the zero vector. This is because the vector equation $x_1\mathbf{v} = \mathbf{0}$ has only the trivial solution when $\mathbf{v} \neq \mathbf{0}$. The zero vector is linearly dependent because $x_1\mathbf{0} = \mathbf{0}$ has many nontrivial solutions.

The next example will explain the nature of a linearly dependent set of two vectors.

EXAMPLE 3 Determine if the following sets of vectors are linearly independent.

a. $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$ b. $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$

SOLUTION

a. Notice that $\mathbf{v}_2$ is a multiple of $\mathbf{v}_1$, namely $\mathbf{v}_2 = 2\mathbf{v}_1$. Hence $-2\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{0}$, which shows that $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly dependent.

b. The vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ are certainly *not* multiples of one another. Could they be linearly dependent? Suppose c and d satisfy

$$c\mathbf{v}_1 + d\mathbf{v}_2 = \mathbf{0}$$

If $c \neq 0$, then we can solve for $\mathbf{v}_1$ in terms of $\mathbf{v}_2$, namely $\mathbf{v}_1 = (-d/c)\mathbf{v}_2$. This result is impossible because $\mathbf{v}_1$ is *not* a multiple of $\mathbf{v}_2$. So c must be zero. Similarly, d must also be zero. Thus $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a linearly independent set. ■

The arguments in Example 3 show that you can always decide *by inspection* when a set of two vectors is linearly dependent. Row operations are unnecessary. Simply check whether at least one of the vectors is a scalar times the other. (The test applies only to sets of *two* vectors.)

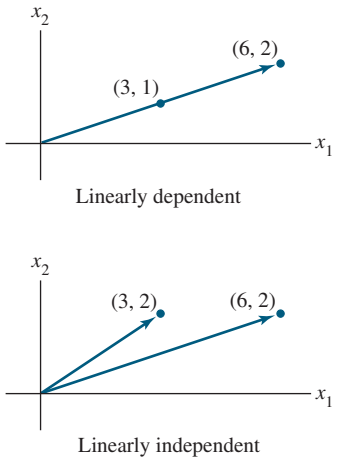


FIGURE 1

A set of two vectors $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly dependent if at least one of the vectors is a multiple of the other. The set is linearly independent if and only if neither of the vectors is a multiple of the other.

In geometric terms, two vectors are linearly dependent if and only if they lie on the same line through the origin. Figure 1 shows the vectors from Example 3.

Sets of Two or More Vectors

The proof of the next theorem is similar to the solution of Example 3. Details are given at the end of this section.

THEOREM 7

Characterization of Linearly Dependent Sets

An indexed set $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ of two or more vectors is linearly dependent if and only if at least one of the vectors in S is a linear combination of the others. In fact, if S is linearly dependent and $\mathbf{v}_1 \neq \mathbf{0}$, then some $\mathbf{v}_j$ (with $j > 1$) is a linear combination of the preceding vectors, $\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}$.

Warning: Theorem 7 does *not* say that *every* vector in a linearly dependent set is a linear combination of the preceding vectors. A vector in a linearly dependent set may fail to be a linear combination of the other vectors. See Practice Problem 1(c).

EXAMPLE 4 Let $\mathbf{u} = \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix}$. Describe the set spanned by $\mathbf{u}$ and $\mathbf{v}$, and explain why a vector $\mathbf{w}$ is in Span $\{\mathbf{u}, \mathbf{v}\}$ if and only if $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly dependent.

SOLUTION The vectors $\mathbf{u}$ and $\mathbf{v}$ are linearly independent because neither vector is a multiple of the other, and so they span a plane in $\mathbb{R}^3$. (See Section 1.3.) In fact, Span $\{\mathbf{u}, \mathbf{v}\}$ is the x_1x_2-plane (with $x_3 = 0$). If $\mathbf{w}$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$, then $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly dependent, by Theorem 7. Conversely, suppose that $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly dependent. By Theorem 7, some vector in $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is a linear combination of the preceding vectors (since $\mathbf{u} \neq \mathbf{0}$). That vector must be $\mathbf{w}$, since $\mathbf{v}$ is not a multiple of $\mathbf{u}$. So $\mathbf{w}$ is in Span $\{\mathbf{u}, \mathbf{v}\}$. See Figure 2. ■

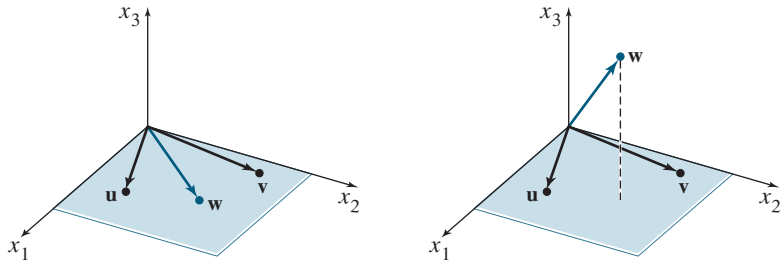


FIGURE 2 Linear dependence in $\mathbb{R}^3$.

Example 4 generalizes to any set $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ in $\mathbb{R}^3$ with $\mathbf{u}$ and $\mathbf{v}$ linearly independent. The set $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ will be linearly dependent if and only if $\mathbf{w}$ is in the plane spanned by $\mathbf{u}$ and $\mathbf{v}$.

The next two theorems describe special cases in which the linear dependence of a set is automatic. Moreover, Theorem 8 will be a key result for work in later chapters.

THEOREM 8

If a set contains more vectors than there are entries in each vector, then the set is linearly dependent. That is, any set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ in $\mathbb{R}^n$ is linearly dependent if $p > n$.

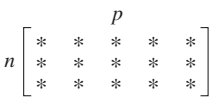


FIGURE 3
If $p > n$, the columns are linearly dependent.

PROOF Let $A = [\,\mathbf{v}_1 \;\cdots\; \mathbf{v}_p\,]$. Then A is $n \times p$, and the equation $A\mathbf{x} = \mathbf{0}$ corresponds to a system of n equations in p unknowns. If $p > n$, there are more variables than equations, so there must be a free variable. Hence $A\mathbf{x} = \mathbf{0}$ has a nontrivial solution, and the columns of A are linearly dependent. See Figure 3 for a matrix version of this theorem. ■

Warning: Theorem 8 says nothing about the case in which the number of vectors in the set does *not* exceed the number of entries in each vector.

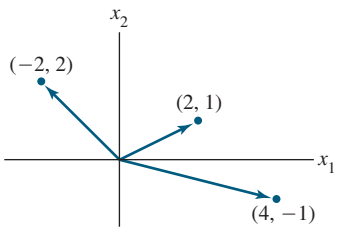


FIGURE 4
A linearly dependent set in $\mathbb{R}^2$.

EXAMPLE 5 The vectors $\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ -1 \end{bmatrix}, \begin{bmatrix} -2 \\ 2 \end{bmatrix}$ are linearly dependent by Theorem 8, because there are three vectors in the set and there are only two entries in each vector. Notice, however, that none of the vectors is a multiple of one of the other vectors. See Figure 4. ■

THEOREM 9

If a set $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ in $\mathbb{R}^n$ contains the zero vector, then the set is linearly dependent.

PROOF By renumbering the vectors, we may suppose $\mathbf{v}_1 = \mathbf{0}$. Then the equation $1\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_p = \mathbf{0}$ shows that S is linearly dependent. ■

EXAMPLE 6 Determine by inspection if the given set is linearly dependent.

a. $\begin{bmatrix} 1 \\ 7 \\ 6 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 9 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 5 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \\ 8 \end{bmatrix}$ b. $\begin{bmatrix} 2 \\ 3 \\ 5 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 8 \end{bmatrix}$ c. $\begin{bmatrix} -2 \\ 4 \\ 6 \\ 10 \end{bmatrix}, \begin{bmatrix} 3 \\ -6 \\ -9 \\ 15 \end{bmatrix}$

SOLUTION

a. The set contains four vectors, each of which has only three entries. So the set is linearly dependent by Theorem 8.

b. Theorem 8 does not apply here because the number of vectors does not exceed the number of entries in each vector. Since the zero vector is in the set, the set is linearly dependent by Theorem 9.

c. Compare the corresponding entries of the two vectors. The second vector seems to be $-3/2$ times the first vector. This relation holds for the first three pairs of entries, but fails for the fourth pair. Thus neither of the vectors is a multiple of the other, and hence they are linearly independent. ■

In general, you should read a section thoroughly *several* times to absorb an important concept such as linear independence. The notes in the *Study Guide* for this section will help you learn to form mental images of key ideas in linear algebra. For instance, the following proof is worth reading carefully because it shows how the definition of linear independence can be *used*.

PROOF OF THEOREM 7 (Characterization of Linearly Dependent Sets)
If some $\mathbf{v}_j$ in S equals a linear combination of the other vectors, then $\mathbf{v}_j$ can be subtracted from both sides of the equation, producing a linear dependence relation with a nonzero weight (-1) on $\mathbf{v}_j$. [For instance, if $\mathbf{v}_1 = c_2\mathbf{v}_2 + c_3\mathbf{v}_3$, then $\mathbf{0} = (-1)\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + 0\mathbf{v}_4 + \cdots + 0\mathbf{v}_p$.] Thus S is linearly dependent.

Conversely, suppose S is linearly dependent. If $\mathbf{v}_1$ is zero, then it is a (trivial) linear combination of the other vectors in S. Otherwise, $\mathbf{v}_1 \neq \mathbf{0}$, and there exist weights $c_1, \ldots, c_p$, not all zero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0}$$

Let j be the largest subscript for which $c_j \neq 0$. If $j = 1$, then $c_1\mathbf{v}_1 = \mathbf{0}$, which is impossible because $\mathbf{v}_1 \neq \mathbf{0}$. So $j > 1$, and

$$\begin{aligned} c_1\mathbf{v}_1 + \cdots + c_j\mathbf{v}_j + 0\mathbf{v}_{j+1} + \cdots + 0\mathbf{v}_p &= \mathbf{0} \\ c_j\mathbf{v}_j &= -c_1\mathbf{v}_1 - \cdots - c_{j-1}\mathbf{v}_{j-1} \\ \mathbf{v}_j &= \left(-\frac{c_1}{c_j}\right)\mathbf{v}_1 + \cdots + \left(-\frac{c_{j-1}}{c_j}\right)\mathbf{v}_{j-1} \end{aligned}$$ ■

Practice Problems

1. Let $\mathbf{u} = \begin{bmatrix} 3 \\ 2 \\ -4 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} -6 \\ 1 \\ 7 \end{bmatrix}$, $\mathbf{w} = \begin{bmatrix} 0 \\ -5 \\ 2 \end{bmatrix}$, and $\mathbf{z} = \begin{bmatrix} 3 \\ 7 \\ -5 \end{bmatrix}$.

a. Are the sets $\{\mathbf{u}, \mathbf{v}\}$, $\{\mathbf{u}, \mathbf{w}\}$, $\{\mathbf{u}, \mathbf{z}\}$, $\{\mathbf{v}, \mathbf{w}\}$, $\{\mathbf{v}, \mathbf{z}\}$, and $\{\mathbf{w}, \mathbf{z}\}$ each linearly independent? Why or why not?

b. Does the answer to Part (a) imply that $\{\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z}\}$ is linearly independent?

c. To determine if $\{\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z}\}$ is linearly dependent, is it wise to check if, say, $\mathbf{w}$ is a linear combination of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{z}$?

d. Is $\{\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z}\}$ linearly dependent?

2. Suppose that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a linearly dependent set of vectors in $\mathbb{R}^n$ and $\mathbf{v}_4$ is a vector in $\mathbb{R}^n$. Show that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is also a linearly dependent set.

1.7 Exercises

In Exercises 1–4, determine if the vectors are linearly independent. Justify each answer.

1. $\begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 7 \\ 2 \\ -6 \end{bmatrix}, \begin{bmatrix} -2 \\ -1 \\ 6 \end{bmatrix}$

2. $\begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 5 \\ -8 \end{bmatrix}, \begin{bmatrix} -3 \\ 4 \\ 1 \end{bmatrix}$

3. $\begin{bmatrix} 1 \\ -3 \end{bmatrix}, \begin{bmatrix} -3 \\ 6 \end{bmatrix}$

4. $\begin{bmatrix} -1 \\ 4 \end{bmatrix}, \begin{bmatrix} -2 \\ 8 \end{bmatrix}$

In Exercises 5–8, determine if the columns of the matrix form a linearly independent set. Justify each answer.

5. $\begin{bmatrix} 0 & -8 & 5 \\ 3 & -7 & 4 \\ -1 & 5 & -4 \\ 1 & -3 & 2 \end{bmatrix}$

6. $\begin{bmatrix} -4 & -3 & 0 \\ 0 & -1 & 4 \\ 1 & 0 & 3 \\ 5 & 4 & 6 \end{bmatrix}$

7. $\begin{bmatrix} 1 & 4 & -3 & 0 \\ -2 & -7 & 5 & 1 \\ -4 & -5 & 7 & 5 \end{bmatrix}$

8. $\begin{bmatrix} 1 & -3 & 3 & -2 \\ -3 & 7 & -1 & 2 \\ 0 & 1 & -4 & 3 \end{bmatrix}$

In Exercises 9 and 10, (a) for what values of h is $\mathbf{v}_3$ in Span $\{\mathbf{v}_1, \mathbf{v}_2\}$, and (b) for what values of h is $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ linearly *dependent*? Justify each answer.

9. $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -3 \\ 2 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -3 \\ 10 \\ -6 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 2 \\ -7 \\ h \end{bmatrix}$

10. $\mathbf{v}_1 = \begin{bmatrix} 1 \\ -5 \\ -3 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -2 \\ 10 \\ 6 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 2 \\ -10 \\ h \end{bmatrix}$

In Exercises 11–14, find the value(s) of h for which the vectors are linearly *dependent*. Justify each answer.

11. $\begin{bmatrix} 1 \\ -1 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ -5 \\ 7 \end{bmatrix}, \begin{bmatrix} -1 \\ 5 \\ h \end{bmatrix}$

12. $\begin{bmatrix} 2 \\ -4 \\ 1 \end{bmatrix}, \begin{bmatrix} -6 \\ 7 \\ -3 \end{bmatrix}, \begin{bmatrix} 8 \\ h \\ 4 \end{bmatrix}$

13. $\begin{bmatrix} 1 \\ 5 \\ -3 \end{bmatrix}, \begin{bmatrix} -2 \\ -9 \\ 6 \end{bmatrix}, \begin{bmatrix} 3 \\ h \\ -9 \end{bmatrix}$

14. $\begin{bmatrix} 1 \\ -3 \\ 4 \end{bmatrix}, \begin{bmatrix} -6 \\ 8 \\ 7 \end{bmatrix}, \begin{bmatrix} 4 \\ -2 \\ h \end{bmatrix}$

Determine by inspection whether the vectors in Exercises 15–20 are linearly *independent*. Justify each answer.

15. $\begin{bmatrix} 5 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 8 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 7 \end{bmatrix}$

16. $\begin{bmatrix} 4 \\ -2 \\ 6 \end{bmatrix}, \begin{bmatrix} 6 \\ -3 \\ 9 \end{bmatrix}$

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

18. $\begin{bmatrix} 4 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \begin{bmatrix} 8 \\ 1 \end{bmatrix}$

19. $\begin{bmatrix} -8 \\ 12 \\ -4 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ -1 \end{bmatrix}$

20. $\begin{bmatrix} 1 \\ 4 \\ -7 \end{bmatrix}, \begin{bmatrix} -2 \\ 5 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

In Exercises 21–28, mark each statement True or False **(T/F)**. Justify each answer on the basis of a careful reading of the text.

21. **(T/F)** The columns of a matrix A are linearly independent if the equation $A\mathbf{x} = \mathbf{0}$ has the trivial solution.

22. **(T/F)** Two vectors are linearly dependent if and only if they lie on a line through the origin.

23. **(T/F)** If S is a linearly dependent set, then each vector is a linear combination of the other vectors in S.

24. **(T/F)** If a set contains fewer vectors than there are entries in the vectors, then the set is linearly independent.

25. **(T/F)** The columns of any 4×5 matrix are linearly dependent.

26. **(T/F)** If $\mathbf{x}$ and $\mathbf{y}$ are linearly independent, and if $\mathbf{z}$ is in Span $\{\mathbf{x}, \mathbf{y}\}$, then $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ is linearly dependent.

27. **(T/F)** If $\mathbf{x}$ and $\mathbf{y}$ are linearly independent, and if $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ is linearly dependent, then $\mathbf{z}$ is in Span $\{\mathbf{x}, \mathbf{y}\}$.

28. **(T/F)** If a set in $\mathbb{R}^n$ is linearly dependent, then the set contains more vectors than there are entries in each vector.

In Exercises 29–32, describe the possible echelon forms of the matrix. Use the notation of Example 1 in Section 1.2.

29. A is a 3×3 matrix with linearly independent columns.

30. A is a 2×2 matrix with linearly dependent columns.

31. A is a 4×2 matrix, $A = [\mathbf{a}_1 \quad \mathbf{a}_2]$, and $\mathbf{a}_2$ is not a multiple of $\mathbf{a}_1$.

32. A is a 4×3 matrix, $A = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3]$, such that $\{\mathbf{a}_1, \mathbf{a}_2\}$ is linearly independent and $\mathbf{a}_3$ is not in Span $\{\mathbf{a}_1, \mathbf{a}_2\}$.

33. How many pivot columns must a 7×5 matrix have if its columns are linearly independent? Why?

34. How many pivot columns must a 5×7 matrix have if its columns span $\mathbb{R}^5$? Why?

35. Construct 3×2 matrices A and B such that $A\mathbf{x} = \mathbf{0}$ has only the trivial solution and $B\mathbf{x} = \mathbf{0}$ has a nontrivial solution.

36. a. Fill in the blank in the following statement: "If A is an $m \times n$ matrix, then the columns of A are linearly independent if and only if A has _____ pivot columns."

b. Explain why the statement in (a) is true.

Exercises 37 and 38 should be solved *without performing row operations*. [*Hint:* Write $A\mathbf{x} = \mathbf{0}$ as a vector equation.]

37. Given $A = \begin{bmatrix} 2 & 3 & 5 \\ -5 & 1 & -4 \\ -3 & -1 & -4 \\ 1 & 0 & 1 \end{bmatrix}$, observe that the third column is the sum of the first two columns. Find a nontrivial solution of $A\mathbf{x} = \mathbf{0}$.

38. Given $A = \begin{bmatrix} 5 & 1 & 8 \\ -9 & 5 & 6 \\ 6 & -5 & -9 \end{bmatrix}$, observe that the first column plus three times the second column equals the third column. Find a nontrivial solution of $A\mathbf{x} = \mathbf{0}$.

Each statement in Exercises 39–44 is either true (in all cases) or false (for at least one example). If false, construct a specific example to show that the statement is not always true. Such an example is called a *counterexample* to the statement. If a statement is true, give a justification. (One specific example cannot explain why a statement is always true. You will have to do more work here than in Exercises 21–28.)

39. (T/F-C) If $\mathbf{v}_1, \ldots, \mathbf{v}_4$ are in $\mathbb{R}^4$ and $\mathbf{v}_3 = 2\mathbf{v}_1 + \mathbf{v}_2$, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is linearly dependent.

40. (T/F-C) If $\mathbf{v}_1, \ldots, \mathbf{v}_4$ are in $\mathbb{R}^4$ and $\mathbf{v}_3 = \mathbf{0}$, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is linearly dependent.

41. (T/F-C) If $\mathbf{v}_1$ and $\mathbf{v}_2$ are in $\mathbb{R}^4$ and $\mathbf{v}_2$ is not a scalar multiple of $\mathbf{v}_1$, then $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly independent.

42. (T/F-C) If $\mathbf{v}_1, \ldots, \mathbf{v}_4$ are in $\mathbb{R}^4$ and $\mathbf{v}_3$ is *not* a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4$, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is linearly independent.

43. (T/F-C) If $\mathbf{v}_1, \ldots, \mathbf{v}_4$ are in $\mathbb{R}^4$ and $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly dependent, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ is also linearly dependent.

44. (T/F-C) If $\mathbf{v}_1, \ldots, \mathbf{v}_4$ are linearly independent vectors in $\mathbb{R}^4$, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is also linearly independent. [*Hint:* Think about $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 + 0 \cdot \mathbf{v}_4 = \mathbf{0}$.]

45. Suppose A is an $m \times n$ matrix with the property that for all $\mathbf{b}$ in $\mathbb{R}^m$ the equation $A\mathbf{x} = \mathbf{b}$ has at most one solution. Use the definition of linear independence to explain why the columns of A must be linearly independent.

46. Suppose an $m \times n$ matrix A has n pivot columns. Explain why for each $\mathbf{b}$ in $\mathbb{R}^m$ the equation $A\mathbf{x} = \mathbf{b}$ has at most one solution. [*Hint:* Explain why $A\mathbf{x} = \mathbf{b}$ cannot have infinitely many solutions.]

T In Exercises 47 and 48, use as many columns of A as possible to construct a matrix B with the property that the equation $B\mathbf{x} = \mathbf{0}$ has only the trivial solution. Solve $B\mathbf{x} = \mathbf{0}$ to verify your work.

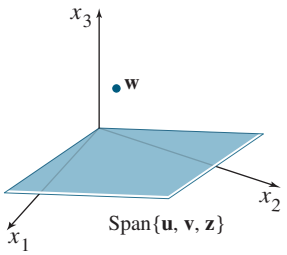
47. $A = \begin{bmatrix} 8 & -3 & 0 & -7 & 2 \\ -9 & 4 & 5 & 11 & -7 \\ 6 & -2 & 2 & -4 & 4 \\ 5 & -1 & 7 & 0 & 10 \end{bmatrix}$

48. $A = \begin{bmatrix} 12 & 10 & -6 & -3 & 7 & 10 \\ -7 & -6 & 4 & 7 & -9 & 5 \\ 9 & 9 & -9 & -5 & 5 & -1 \\ -4 & -3 & 1 & 6 & -8 & 9 \\ 8 & 7 & -5 & -9 & 11 & -8 \end{bmatrix}$

T 49. With A and B as in Exercise 47 select a column $\mathbf{v}$ of A that was not used in the construction of B and determine if $\mathbf{v}$ is in the set spanned by the columns of B. (Describe your calculations.)

T 50. Repeat Exercise 49 with the matrices A and B from Exercise 48. Then give an explanation for what you discover, assuming that B was constructed as specified.

STUDY GUIDE offers additional resources for mastering the concept of linear independence.



Solutions to Practice Problems

1. a. Yes. In each case, neither vector is a multiple of the other. Thus each set is linearly independent.

b. No. The observation in Part (a), by itself, says nothing about the linear independence of $\{\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{z}\}$.

c. No. When testing for linear independence, it is usually a poor idea to check if one selected vector is a linear combination of the others. It may happen that the selected vector is not a linear combination of the others and yet the whole set of vectors is linearly dependent. In this practice problem, $\mathbf{w}$ is not a linear combination of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{z}$.

d. Yes, by Theorem 8. There are more vectors (four) than entries (three) in them.

2. Applying the definition of linearly dependent to $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ implies that there exist scalars c_1, c_2, and c_3, not all zero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 = \mathbf{0}.$$

Adding $0\,\mathbf{v}_4 = \mathbf{0}$ to both sides of this equation results in

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + 0\,\mathbf{v}_4 = \mathbf{0}.$$

Since c_1, c_2, c_3 and 0 are not *all* zero, the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ satisfies the definition of a linearly dependent set.

1.8 Introduction to Linear Transformations

The difference between a matrix equation $A\mathbf{x} = \mathbf{b}$ and the associated vector equation $x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{b}$ is merely a matter of notation. However, a matrix equation $A\mathbf{x} = \mathbf{b}$ can arise in linear algebra (and in applications such as computer graphics and signal processing) in a way that is not directly connected with linear combinations of vectors. This happens when we think of the matrix A as an object that "acts" on a vector $\mathbf{x}$ by multiplication to produce a new vector called $A\mathbf{x}$.

For instance, the equations

$$\underset{A}{\begin{bmatrix} 4 & -3 & 1 & 3 \\ 2 & 0 & 5 & 1 \end{bmatrix}} \underset{\mathbf{x}}{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}} = \underset{\mathbf{b}}{\begin{bmatrix} 5 \\ 8 \end{bmatrix}} \quad \text{and} \quad \underset{A}{\begin{bmatrix} 4 & -3 & 1 & 3 \\ 2 & 0 & 5 & 1 \end{bmatrix}} \underset{\mathbf{u}}{\begin{bmatrix} 1 \\ 4 \\ -1 \\ 3 \end{bmatrix}} = \underset{\mathbf{0}}{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}$$

say that multiplication by A transforms $\mathbf{x}$ into $\mathbf{b}$ and transforms $\mathbf{u}$ into the zero vector. See Figure 1.

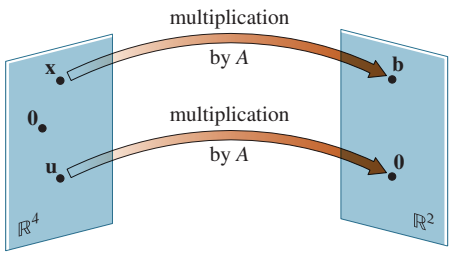


FIGURE 1 Transforming vectors via matrix multiplication.

From this new point of view, solving the equation $A\mathbf{x} = \mathbf{b}$ amounts to finding all vectors $\mathbf{x}$ in $\mathbb{R}^4$ that are transformed into the vector $\mathbf{b}$ in $\mathbb{R}^2$ under the "action" of multiplication by A.

The correspondence from $\mathbf{x}$ to $A\mathbf{x}$ is a *function* from one set of vectors to another. This concept generalizes the common notion of a function as a rule that transforms one real number into another.

A **transformation** (or **function** or **mapping**) T from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a rule that assigns to each vector $\mathbf{x}$ in $\mathbb{R}^n$ a vector $T(\mathbf{x})$ in $\mathbb{R}^m$. The set $\mathbb{R}^n$ is called the **domain** of T, and $\mathbb{R}^m$

is called the **codomain** of T. The notation $T : \mathbb{R}^n \to \mathbb{R}^m$ indicates that the domain of T is $\mathbb{R}^n$ and the codomain is $\mathbb{R}^m$. For $\mathbf{x}$ in $\mathbb{R}^n$, the vector $T(\mathbf{x})$ in $\mathbb{R}^m$ is called the **image** of $\mathbf{x}$ (under the action of T). The set of all images $T(\mathbf{x})$ is called the **range** of T. See Figure 2.

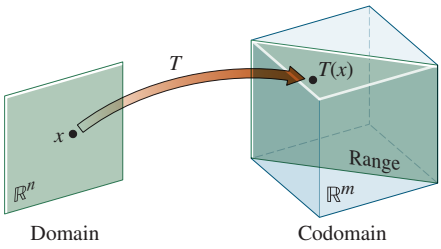
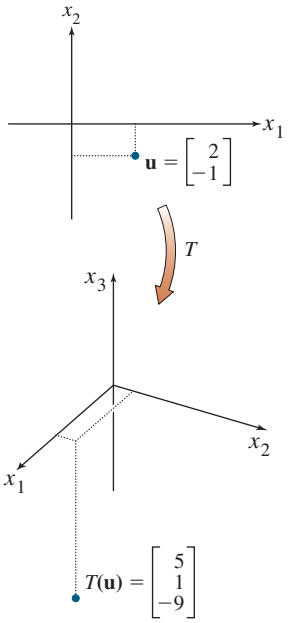


FIGURE 2 Domain, codomain, and range of $T : \mathbb{R}^n \to \mathbb{R}^m$.

The new terminology in this section is important because a dynamic view of matrix–vector multiplication is the key to understanding several ideas in linear algebra and to building mathematical models of physical systems that evolve over time. Such *dynamical systems* will be discussed in Sections 1.10, 4.8, and throughout Chapter 5.

Matrix Transformations

The rest of this section focuses on mappings associated with matrix multiplication. For each $\mathbf{x}$ in $\mathbb{R}^n$, $T(\mathbf{x})$ is computed as $A\mathbf{x}$, where A is an $m \times n$ matrix. For simplicity, we sometimes denote such a *matrix transformation* by $\mathbf{x} \mapsto A\mathbf{x}$. Observe that the domain of T is $\mathbb{R}^n$ when A has n columns and the codomain of T is $\mathbb{R}^m$ when each column of A has m entries. The range of T is the set of all linear combinations of the columns of A, because each image $T(\mathbf{x})$ is of the form $A\mathbf{x}$.



EXAMPLE 1 Let $A = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix}$, $\mathbf{c} = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}$, and define a transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ by $T(\mathbf{x}) = A\mathbf{x}$, so that

$$T(\mathbf{x}) = A\mathbf{x} = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - 3x_2 \\ 3x_1 + 5x_2 \\ -x_1 + 7x_2 \end{bmatrix}$$

a. Find $T(\mathbf{u})$, the image of $\mathbf{u}$ under the transformation T.

b. Find an $\mathbf{x}$ in $\mathbb{R}^2$ whose image under T is $\mathbf{b}$.

c. Is there more than one $\mathbf{x}$ whose image under T is $\mathbf{b}$?

d. Determine if $\mathbf{c}$ is in the range of the transformation T.

SOLUTION

a. Compute

$$T(\mathbf{u}) = A\mathbf{u} = \begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ -9 \end{bmatrix}$$

b. Solve $T(\mathbf{x}) = \mathbf{b}$ for $\mathbf{x}$. That is, solve $A\mathbf{x} = \mathbf{b}$, or

$$\begin{bmatrix} 1 & -3 \\ 3 & 5 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ -5 \end{bmatrix} \tag{1}$$

Using the method discussed in Section 1.4, row reduce the augmented matrix:

$$\begin{bmatrix} 1 & -3 & 3 \\ 3 & 5 & 2 \\ -1 & 7 & -5 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 14 & -7 \\ 0 & 4 & -2 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 1 & -.5 \\ 0 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1.5 \\ 0 & 1 & -.5 \\ 0 & 0 & 0 \end{bmatrix} \quad (2)$$

Hence $x_1 = 1.5, x_2 = -.5$, and $\mathbf{x} = \begin{bmatrix} 1.5 \\ -.5 \end{bmatrix}$. The image of this $\mathbf{x}$ under T is the given vector $\mathbf{b}$.

c. Any $\mathbf{x}$ whose image under T is $\mathbf{b}$ must satisfy equation (1). From (2), it is clear that equation (1) has a unique solution. So there is exactly one $\mathbf{x}$ whose image is $\mathbf{b}$.

d. The vector $\mathbf{c}$ is in the range of T if $\mathbf{c}$ is the image of some $\mathbf{x}$ in $\mathbb{R}^2$, that is, if $\mathbf{c} = T(\mathbf{x})$ for some $\mathbf{x}$. This is just another way of asking if the system $A\mathbf{x} = \mathbf{c}$ is consistent. To find the answer, row reduce the augmented matrix:

$$\begin{bmatrix} 1 & -3 & 3 \\ 3 & 5 & 2 \\ -1 & 7 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 14 & -7 \\ 0 & 4 & 8 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 1 & 2 \\ 0 & 14 & -7 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & -35 \end{bmatrix}$$

The third equation, $0 = -35$, shows that the system is inconsistent. So $\mathbf{c}$ is *not* in the range of T. ■

The question in Example 1(c) is a *uniqueness* problem for a system of linear equations, translated here into the language of matrix transformations: Is $\mathbf{b}$ the image of a *unique* $\mathbf{x}$ in $\mathbb{R}^n$? Similarly, Example 1(d) is an *existence* problem: Does there *exist* an $\mathbf{x}$ whose image is $\mathbf{c}$?

The next two matrix transformations can be viewed geometrically. They reinforce the dynamic view of a matrix as something that transforms vectors into other vectors. Section 2.7 contains other interesting examples connected with computer graphics.

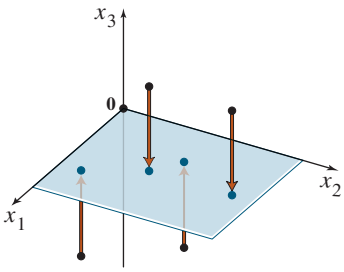


FIGURE 3
A projection transformation.

EXAMPLE 2 If $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, then the transformation $\mathbf{x} \mapsto A\mathbf{x}$ *projects* points in $\mathbb{R}^3$ onto the x_1x_2-plane because

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \mapsto \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ 0 \end{bmatrix}$$

See Figure 3. ■



sheep



sheared sheep

EXAMPLE 3 Let $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$. The transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(\mathbf{x}) = A\mathbf{x}$ is called a **shear transformation**. It can be shown that if T acts on each point in the 2×2 square shown in Figure 4, then the set of images forms the sheared parallelogram. The key idea is to show that T maps line segments onto line segments (as shown in Exercise 35) and then to check that the corners of the square map onto the vertices of the parallelogram. For instance, the image of the point $\mathbf{u} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$ is $T(\mathbf{u}) = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$, and the image of $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$ is $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$. T deforms the square as if the top of the square were pushed to the right while the base is held fixed. Shear transformations appear in physics, geology, and crystallography. ■

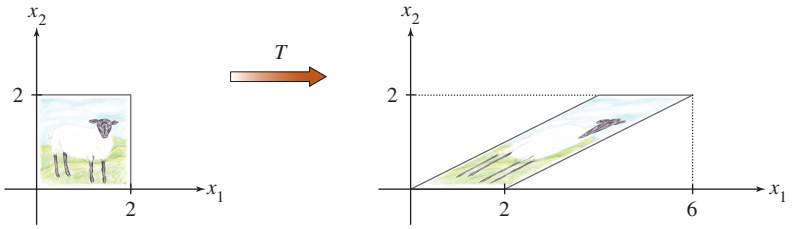


FIGURE 4 A shear transformation.

Linear Transformations

Theorem 5 in Section 1.4 shows that if A is $m \times n$, then the transformation $\mathbf{x} \mapsto A\mathbf{x}$ has the properties

$$A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} \quad \text{and} \quad A(c\mathbf{u}) = cA\mathbf{u}$$

for all $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^n$ and all scalars c. These properties, written in function notation, identify the most important class of transformations in linear algebra.

DEFINITION

A transformation (or mapping) T is **linear** if

(i) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v}$ in the domain of T;

(ii) $T(c\mathbf{u}) = cT(\mathbf{u})$ for all scalars c and all $\mathbf{u}$ in the domain of T.

Every matrix transformation is a linear transformation. Important examples of linear transformations that are not matrix transformations will be discussed in Chapters 4 and 5.

Linear transformations *preserve the operations of vector addition and scalar multiplication*. Property (i) says that the result $T(\mathbf{u} + \mathbf{v})$ of first adding $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ and then applying T is the same as first applying T to $\mathbf{u}$ and to $\mathbf{v}$ and then adding $T(\mathbf{u})$ and $T(\mathbf{v})$ in $\mathbb{R}^m$. These two properties lead easily to the following useful facts.

If T is a linear transformation, then

$$T(\mathbf{0}) = \mathbf{0} \tag{3}$$

and

$$T(c\mathbf{u} + d\mathbf{v}) = cT(\mathbf{u}) + dT(\mathbf{v}) \tag{4}$$

for all vectors $\mathbf{u}, \mathbf{v}$ in the domain of T and all scalars c, d.

Property (3) follows from condition (ii) in the definition, because $T(\mathbf{0}) = T(0\mathbf{u}) = 0T(\mathbf{u}) = \mathbf{0}$. Property (4) requires both (i) and (ii):

$$T(c\mathbf{u} + d\mathbf{v}) = T(c\mathbf{u}) + T(d\mathbf{v}) = cT(\mathbf{u}) + dT(\mathbf{v})$$

Observe that *if a transformation satisfies* (4) *for all* $\mathbf{u}$, $\mathbf{v}$ *and* c, d, *it must be linear*. (Set $c = d = 1$ for preservation of addition, and set $d = 0$ for preservation of scalar multiplication.) Repeated application of (4) produces a useful generalization:

$$T(c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p) = c_1T(\mathbf{v}_1) + \cdots + c_pT(\mathbf{v}_p) \tag{5}$$

In engineering and physics, (5) is referred to as a *superposition principle*. Think of $\mathbf{v}_1, \ldots, \mathbf{v}_p$ as signals that go into a system and $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_p)$ as the responses of that system to the signals. The system satisfies the superposition principle if whenever an input is expressed as a linear combination of such signals, the system's response is *the same* linear combination of the responses to the individual signals. We will return to this idea in Chapter 4.

EXAMPLE 4 Given a scalar r, define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(\mathbf{x}) = r\mathbf{x}$. T is called a **contraction** when $0 \le r \le 1$ and a **dilation** when $r > 1$. Let $r = 3$, and show that T is a linear transformation.

SOLUTION Let $\mathbf{u}$, $\mathbf{v}$ be in $\mathbb{R}^2$ and let c, d be scalars. Then

$$\begin{aligned} T(c\mathbf{u} + d\mathbf{v}) &= 3(c\mathbf{u} + d\mathbf{v}) && \text{Definition of } T \\ &= 3c\mathbf{u} + 3d\mathbf{v} \\ &= c(3\mathbf{u}) + d(3\mathbf{v}) && \text{Vector arithmetic} \\ &= cT(\mathbf{u}) + dT(\mathbf{v}) \end{aligned}$$

Thus T is a linear transformation because it satisfies (4). See Figure 5. ■

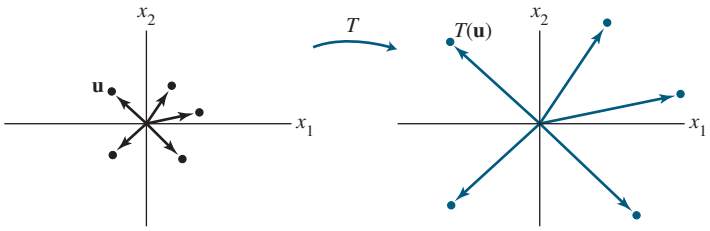


FIGURE 5 A dilation transformation.

EXAMPLE 5 Define a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$T(\mathbf{x}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix}$$

Find the images under T of $\mathbf{u} = \begin{bmatrix} 4 \\ 1 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, and $\mathbf{u} + \mathbf{v} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$.

SOLUTION

$$T(\mathbf{u}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 4 \end{bmatrix}, \qquad T(\mathbf{v}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -3 \\ 2 \end{bmatrix},$$

$$T(\mathbf{u} + \mathbf{v}) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 6 \\ 4 \end{bmatrix} = \begin{bmatrix} -4 \\ 6 \end{bmatrix}$$

Note that $T(\mathbf{u} + \mathbf{v})$ is obviously equal to $T(\mathbf{u}) + T(\mathbf{v})$. It appears from Figure 6 that T rotates $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{u} + \mathbf{v}$ counterclockwise about the origin through 90°. In fact, T transforms the entire parallelogram determined by $\mathbf{u}$ and $\mathbf{v}$ into the one determined by $T(\mathbf{u})$ and $T(\mathbf{v})$. (See Exercise 36.) ■

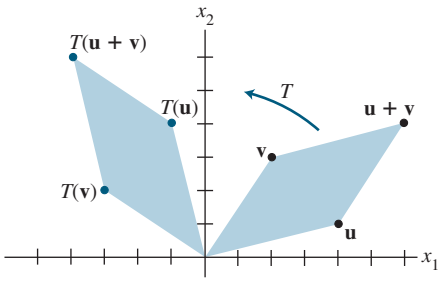


FIGURE 6 A rotation transformation.

The final example is not geometrical; instead, it shows how a linear mapping can transform one type of data into another.

EXAMPLE 6 A company manufactures two products, B and C. Using data from Example 7 in Section 1.3, we construct a "unit cost" matrix, $U = [\,\mathbf{b} \quad \mathbf{c}\,]$, whose columns describe the "costs per dollar of output" for the products:

$$U = \begin{array}{c} \begin{matrix} \text{Product} \\ \text{B} \quad\quad \text{C} \end{matrix} \\ \begin{bmatrix} .45 & .40 \\ .25 & .30 \\ .15 & .15 \end{bmatrix} \end{array} \begin{array}{l} \\ \text{Materials} \\ \text{Labor} \\ \text{Overhead} \end{array}$$

Let $\mathbf{x} = (x_1, x_2)$ be a "production" vector, corresponding to x_1 dollars of product B and x_2 dollars of product C, and define $T : \mathbb{R}^2 \to \mathbb{R}^3$ by

$$T(\mathbf{x}) = U\mathbf{x} = x_1 \begin{bmatrix} .45 \\ .25 \\ .15 \end{bmatrix} + x_2 \begin{bmatrix} .40 \\ .30 \\ .15 \end{bmatrix} = \begin{bmatrix} \text{Total cost of materials} \\ \text{Total cost of labor} \\ \text{Total cost of overhead} \end{bmatrix}$$

The mapping T transforms a list of production quantities (measured in dollars) into a list of total costs. The linearity of this mapping is reflected in two ways:

1. If production is increased by a factor of, say, 4, from $\mathbf{x}$ to $4\mathbf{x}$, then the costs will increase by the same factor, from $T(\mathbf{x})$ to $4T(\mathbf{x})$.
2. If $\mathbf{x}$ and $\mathbf{y}$ are production vectors, then the total cost vector associated with the combined production $\mathbf{x} + \mathbf{y}$ is precisely the sum of the cost vectors $T(\mathbf{x})$ and $T(\mathbf{y})$. ■

Practice Problems

1. Suppose $T : \mathbb{R}^5 \to \mathbb{R}^2$ and $T(\mathbf{x}) = A\mathbf{x}$ for some matrix A and for each $\mathbf{x}$ in $\mathbb{R}^5$. How many rows and columns does A have?
2. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Give a geometric description of the transformation $\mathbf{x} \mapsto A\mathbf{x}$.
3. The line segment from $\mathbf{0}$ to a vector $\mathbf{u}$ is the set of points of the form $t\mathbf{u}$, where $0 \le t \le 1$. Show that a linear transformation T maps this segment into the segment between $\mathbf{0}$ and $T(\mathbf{u})$.

1.8 Exercises

1. Let $A = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$, and define $T : \mathbb{R}^2 \to \mathbb{R}^2$ by $T(\mathbf{x}) = A\mathbf{x}$. Find the images under T of $\mathbf{u} = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$.

2. Let $A = \begin{bmatrix} .5 & 0 & 0 \\ 0 & .5 & 0 \\ 0 & 0 & .5 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 1 \\ 0 \\ -4 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$. Define $T : \mathbb{R}^3 \to \mathbb{R}^3$ by $T(\mathbf{x}) = A\mathbf{x}$. Find $T(\mathbf{u})$ and $T(\mathbf{v})$.

In Exercises 3–6, with T defined by $T(\mathbf{x}) = A\mathbf{x}$, find a vector $\mathbf{x}$ whose image under T is $\mathbf{b}$, and determine whether $\mathbf{x}$ is unique.

3. $A = \begin{bmatrix} 1 & 0 & -2 \\ -2 & 1 & 6 \\ 3 & -2 & -5 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} -1 \\ 7 \\ -3 \end{bmatrix}$

4. $A = \begin{bmatrix} 1 & -3 & 2 \\ 0 & 1 & -4 \\ 3 & -5 & -9 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 6 \\ -7 \\ -9 \end{bmatrix}$

5. $A = \begin{bmatrix} 1 & -5 & -7 \\ -3 & 7 & 5 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$

6. $A = \begin{bmatrix} 1 & -2 & 1 \\ 3 & -4 & 5 \\ 0 & 1 & 1 \\ -3 & 5 & -4 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 1 \\ 9 \\ 3 \\ -6 \end{bmatrix}$

7. Let A be a 4×6 matrix. What must a and b be in order to define $T : \mathbb{R}^a \to \mathbb{R}^b$ by $T(\mathbf{x}) = A\mathbf{x}$?

8. How many rows and columns must a matrix A have in order to define a mapping from $\mathbb{R}^3$ into $\mathbb{R}^6$ by the rule $T(\mathbf{x}) = A\mathbf{x}$?

For Exercises 9 and 10, find all $\mathbf{x}$ in $\mathbb{R}^4$ that are mapped into the zero vector by the transformation $\mathbf{x} \mapsto A\mathbf{x}$ for the given matrix A.

9. $A = \begin{bmatrix} 1 & -4 & 7 & -5 \\ 0 & 1 & -4 & 3 \\ 2 & -6 & 6 & -4 \end{bmatrix}$

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

11. Let $\mathbf{b} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$, and let A be the matrix in Exercise 9. Is $\mathbf{b}$ in the range of the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$? Why or why not?

12. Let $\mathbf{b} = \begin{bmatrix} -1 \\ 3 \\ -1 \\ 4 \end{bmatrix}$, and let A be the matrix in Exercise 10. Is $\mathbf{b}$ in the range of the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$? Why or why not?

In Exercises 13–16, use a rectangular coordinate system to plot $\mathbf{u} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} -2 \\ 4 \end{bmatrix}$, and their images under the given transformation T. (Make a separate and reasonably large sketch for each exercise.) Describe geometrically what T does to each vector $\mathbf{x}$ in $\mathbb{R}^2$.

13. $T(\mathbf{x}) = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

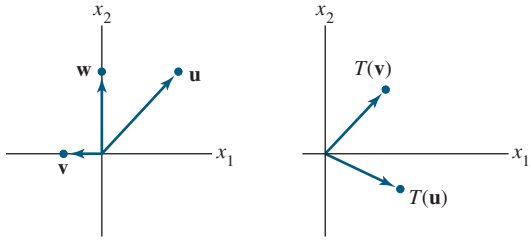
14. $T(\mathbf{x}) = \begin{bmatrix} .5 & 0 \\ 0 & .5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

15. $T(\mathbf{x}) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

16. $T(\mathbf{x}) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

17. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation that maps $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ into $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$ and maps $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ into $\begin{bmatrix} 1 \\ -5 \end{bmatrix}$. Use the fact that T is linear to find the images under T of $5\mathbf{u}$, $4\mathbf{v}$, and $5\mathbf{u} + 4\mathbf{v}$.

18. The figure shows vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$, along with the images $T(\mathbf{u})$ and $T(\mathbf{v})$ under the action of a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$. Copy this figure carefully, and draw the image $T(\mathbf{w})$ as accurately as possible. [*Hint:* First, write $\mathbf{w}$ as a linear combination of $\mathbf{u}$ and $\mathbf{v}$.]



19. Let $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $\mathbf{y}_1 = \begin{bmatrix} 2 \\ 5 \end{bmatrix}$, and $\mathbf{y}_2 = \begin{bmatrix} -1 \\ 6 \end{bmatrix}$, and let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation that maps $\mathbf{e}_1$ into $\mathbf{y}_1$ and maps $\mathbf{e}_2$ into $\mathbf{y}_2$. Find the images of $\begin{bmatrix} 5 \\ -3 \end{bmatrix}$ and $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$.

20. Let $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} -3 \\ 5 \end{bmatrix}$, and $\mathbf{v}_1 = \begin{bmatrix} 2 \\ -9 \end{bmatrix}$, and let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation that maps $\mathbf{x}$ into $x_1\mathbf{v}_1 + x_2\mathbf{v}_2$. Find a matrix A such that $T(\mathbf{x})$ is $A\mathbf{x}$ for each $\mathbf{x}$.

In Exercises 21–30, mark each statement True or False **(T/F)**. Justify each answer.

21. **(T/F)** A linear transformation is a special type of function.

22. **(T/F)** Every matrix transformation is a linear transformation.

23. **(T/F)** If A is a 3×5 matrix and T is a transformation defined by $T(\mathbf{x}) = A\mathbf{x}$, then the domain of T is $\mathbb{R}^3$.

24. **(T/F)** The codomain of the transformation $\mathbf{x} \mapsto A\mathbf{x}$ is the set of all linear combinations of the columns of A.

25. **(T/F)** If A is an $m \times n$ matrix, then the range of the transformation $\mathbf{x} \mapsto A\mathbf{x}$ is $\mathbb{R}^m$.

26. **(T/F)** If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation and if $\mathbf{c}$ is in $\mathbb{R}^m$, then a uniqueness question is "Is $\mathbf{c}$ in the range of T?"

27. **(T/F)** Every linear transformation is a matrix transformation.

28. **(T/F)** A linear transformation preserves the operations of vector addition and scalar multiplication.

29. **(T/F)** A transformation T is linear if and only if $T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2)$ for all $\mathbf{v}_1$ and $\mathbf{v}_2$ in the domain of T and for all scalars c_1 and c_2.

30. **(T/F)** The superposition principle is a physical description of a linear transformation.

31. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation that reflects each point through the x_1-axis. (See Practice Problem 2.) Make two sketches similar to Figure 6 that illustrate properties (i) and (ii) of a linear transformation.

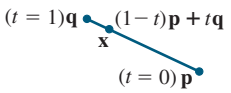
32. Suppose vectors $\mathbf{v}_1, \ldots, \mathbf{v}_p$ span $\mathbb{R}^n$, and let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. Suppose $T(\mathbf{v}_i) = \mathbf{0}$ for $i = 1, \ldots, p$. Show that T is the zero transformation. That is, show that if $\mathbf{x}$ is any vector in $\mathbb{R}^n$, then $T(\mathbf{x}) = \mathbf{0}$.

33. Given $\mathbf{v} \neq \mathbf{0}$ and $\mathbf{p}$ in $\mathbb{R}^n$, the line through $\mathbf{p}$ in the direction of $\mathbf{v}$ has the parametric equation $\mathbf{x} = \mathbf{p} + t\mathbf{v}$. Show that a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ maps this line onto another line or onto a single point (a *degenerate line*).

34. Let $\mathbf{u}$ and $\mathbf{v}$ be linearly independent vectors in $\mathbb{R}^3$, and let P be the plane through $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{0}$. The parametric equation of P is $\mathbf{x} = s\mathbf{u} + t\mathbf{v}$ (with s, t in $\mathbb{R}$). Show that a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ maps P onto a plane through $\mathbf{0}$, or onto a line through $\mathbf{0}$, or onto just the origin in $\mathbb{R}^3$. What must be true about $T(\mathbf{u})$ and $T(\mathbf{v})$ in order for the image of the plane P to be a plane?

35. a. Show that the line through vectors $\mathbf{p}$ and $\mathbf{q}$ in $\mathbb{R}^n$ may be written in the parametric form $\mathbf{x} = (1 - t)\mathbf{p} + t\mathbf{q}$. (Refer to the figure with Exercises 25 and 26 in Section 1.5.)

b. The line segment from $\mathbf{p}$ to $\mathbf{q}$ is the set of points of the form $(1 - t)\mathbf{p} + t\mathbf{q}$ for $0 \leq t \leq 1$ (as shown in the figure below). Show that a linear transformation T maps this line segment onto a line segment or onto a single point.



36. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$. It can be shown that the set P of all points in the parallelogram determined by $\mathbf{u}$ and $\mathbf{v}$ has the form $a\mathbf{u} + b\mathbf{v}$, for $0 \leq a \leq 1, 0 \leq b \leq 1$. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Explain why the image of a point in P under the transformation T lies in the parallelogram determined by $T(\mathbf{u})$ and $T(\mathbf{v})$.

37. Define $f : \mathbb{R} \to \mathbb{R}$ by $f(x) = mx + b$.

a. Show that f is a linear transformation when $b = 0$.

b. Find a property of a linear transformation that is violated when $b \neq 0$.

c. Why is f called a linear function?

38. An *affine transformation* $T : \mathbb{R}^n \to \mathbb{R}^m$ has the form $T(x) = A\mathbf{x} + \mathbf{b}$, with A an $m \times n$ matrix and $\mathbf{b}$ in $\mathbb{R}^m$. Show that T is *not* a linear transformation when $\mathbf{b} \neq \mathbf{0}$. (Affine transformations are important in computer graphics.)

39. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and let $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ be a linearly dependent set in $\mathbb{R}^n$. Explain why the set $\{T(\mathbf{v}_1), T(\mathbf{v}_2), T(\mathbf{v}_3)\}$ is linearly dependent.

In Exercises 40–44, column vectors are written as rows, such as $\mathbf{x} = (x_1, x_2)$, and $T(\mathbf{x})$ is written as $T(x_1, x_2)$.

40. Show that the transformation T defined by $T(x_1, x_2) = (4x_1 - 2x_2, 3|x_2|)$ is not linear.

41. Show that the transformation T defined by $T(x_1, x_2) = (2x_1 - 3x_2, x_1 + 4, 5x_2)$ is not linear.

42. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Show that if T maps two linearly independent vectors onto a linearly dependent set, then the equation $T(\mathbf{x}) = \mathbf{0}$ has a nontrivial solution. [*Hint:* Suppose $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are linearly independent and yet $T(\mathbf{u})$ and $T(\mathbf{v})$ are linearly dependent. Then $c_1T(\mathbf{u}) + c_2T(\mathbf{v}) = \mathbf{0}$ for some weights c_1 and c_2, not both zero. Use this equation.]

43. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the transformation that reflects each vector $\mathbf{x} = (x_1, x_2, x_3)$ through the plane $x_3 = 0$ onto $T(\mathbf{x}) = (x_1, x_2, -x_3)$. Show that T is a linear transformation. [See Example 4 for ideas.]

44. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the transformation that projects each vector $\mathbf{x} = (x_1, x_2, x_3)$ onto the plane $x_2 = 0$, so $T(\mathbf{x}) = (x_1, 0, x_3)$. Show that T is a linear transformation.

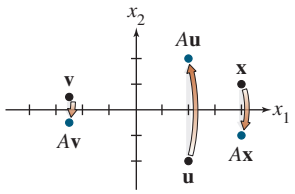
T In Exercises 45 and 46, the given matrix determines a linear transformation T. Find all $\mathbf{x}$ such that $T(\mathbf{x}) = \mathbf{0}$.

45. $\begin{bmatrix} 4 & -2 & 5 & -5 \\ -9 & 7 & -8 & 0 \\ -6 & 4 & 5 & 3 \\ 5 & -3 & 8 & -4 \end{bmatrix}$ **46.** $\begin{bmatrix} -9 & -4 & -9 & 4 \\ 5 & -8 & -7 & 6 \\ 7 & 11 & 16 & -9 \\ 9 & -7 & -4 & 5 \end{bmatrix}$

T **47.** Let $\mathbf{b} = \begin{bmatrix} 7 \\ 5 \\ 9 \\ 7 \end{bmatrix}$ and let A be the matrix in Exercise 45. Is $\mathbf{b}$ in the range of the transformation $\mathbf{x} \mapsto A\mathbf{x}$? If so, find an $\mathbf{x}$ whose image under the transformation is $\mathbf{b}$.

T **48.** Let $\mathbf{b} = \begin{bmatrix} -7 \\ -7 \\ 13 \\ -5 \end{bmatrix}$ and let A be the matrix in Exercise 46. Is $\mathbf{b}$ in the range of the transformation $\mathbf{x} \mapsto A\mathbf{x}$? If so, find an $\mathbf{x}$ whose image under the transformation is $\mathbf{b}$.

STUDY GUIDE offers additional resources for mastering linear transformations.



The transformation $\mathbf{x} \mapsto A\mathbf{x}$.

Solutions to Practice Problems

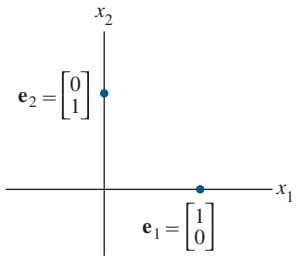
1. A must have five columns for $A\mathbf{x}$ to be defined. A must have two rows for the codomain of T to be $\mathbb{R}^2$.

2. Plot some random points (vectors) on graph paper to see what happens. A point such as $(4, 1)$ maps into $(4, -1)$. The transformation $\mathbf{x} \mapsto A\mathbf{x}$ reflects points through the x-axis (or x_1-axis).

3. Let $\mathbf{x} = t\mathbf{u}$ for some t such that $0 \le t \le 1$. Since T is linear, $T(t\mathbf{u}) = t\,T(\mathbf{u})$, which is a point on the line segment between $\mathbf{0}$ and $T(\mathbf{u})$.

1.9 The Matrix of a Linear Transformation

Whenever a linear transformation T arises geometrically or is described in words, we usually want a "formula" for $T(\mathbf{x})$. The discussion that follows shows that every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is actually a matrix transformation $\mathbf{x} \mapsto A\mathbf{x}$ and that important properties of T are intimately related to familiar properties of A. The key to finding A is to observe that T is completely determined by what it does to the columns of the $n \times n$ identity matrix I_n.



EXAMPLE 1 The columns of $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ are $\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Suppose T is a linear transformation from $\mathbb{R}^2$ into $\mathbb{R}^3$ such that

$$T(\mathbf{e}_1) = \begin{bmatrix} 5 \\ -7 \\ 2 \end{bmatrix} \quad \text{and} \quad T(\mathbf{e}_2) = \begin{bmatrix} -3 \\ 8 \\ 0 \end{bmatrix}$$

With no additional information, find a formula for the image of an arbitrary $\mathbf{x}$ in $\mathbb{R}^2$.

SOLUTION Write

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 \tag{1}$$

Since T is a *linear* transformation,

$$T(\mathbf{x}) = x_1 T(\mathbf{e}_1) + x_2 T(\mathbf{e}_2) \tag{2}$$

$$= x_1 \begin{bmatrix} 5 \\ -7 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} -3 \\ 8 \\ 0 \end{bmatrix} = \begin{bmatrix} 5x_1 - 3x_2 \\ -7x_1 + 8x_2 \\ 2x_1 + 0 \end{bmatrix}$$

■

The step from equation (1) to equation (2) explains why knowledge of $T(\mathbf{e}_1)$ and $T(\mathbf{e}_2)$ is sufficient to determine $T(\mathbf{x})$ for any $\mathbf{x}$. Moreover, since (2) expresses $T(\mathbf{x})$ as a linear combination of vectors, we can put these vectors into the columns of a matrix A and write (2) as

$$T(\mathbf{x}) = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A\mathbf{x}$$

THEOREM 10

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then there exists a unique matrix A such that

$$T(\mathbf{x}) = A\mathbf{x} \quad \text{for all } \mathbf{x} \text{ in } \mathbb{R}^n$$

In fact, A is the $m \times n$ matrix whose jth column is the vector $T(\mathbf{e}_j)$, where $\mathbf{e}_j$ is the jth column of the identity matrix in $\mathbb{R}^n$:

$$A = \begin{bmatrix} T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \end{bmatrix} \tag{3}$$

PROOF Write $\mathbf{x} = I_n\mathbf{x} = [\,\mathbf{e}_1 \;\; \cdots \;\; \mathbf{e}_n\,]\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$, and use the linearity of T to compute

$$T(\mathbf{x}) = T(x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n) = x_1T(\mathbf{e}_1) + \cdots + x_nT(\mathbf{e}_n)$$

$$= \begin{bmatrix} T(\mathbf{e}_1) & \cdots & T(\mathbf{e}_n) \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = A\mathbf{x}$$

The uniqueness of A is treated in Exercise 41. ■

The matrix A in (3) is called the **standard matrix for the linear transformation** T.

We know now that every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ can be viewed as a matrix transformation, and vice versa. The term *linear transformation* focuses on a property of a mapping, while *matrix transformation* describes how such a mapping is implemented, as Examples 2 and 3 illustrate.

EXAMPLE 2 Find the standard matrix A for the dilation transformation $T(\mathbf{x}) = 3\mathbf{x}$, for $\mathbf{x}$ in $\mathbb{R}^2$.

SOLUTION Write

$$T(\mathbf{e}_1) = 3\mathbf{e}_1 = \begin{bmatrix} 3 \\ 0 \end{bmatrix} \quad \text{and} \quad T(\mathbf{e}_2) = 3\mathbf{e}_2 = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$$

■

EXAMPLE 3 Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation that rotates each point in $\mathbb{R}^2$ about the origin through an angle φ, with counterclockwise rotation for a positive angle. We could show geometrically that such a transformation is linear. (See Figure 6 in Section 1.8.) Find the standard matrix A of this transformation.

SOLUTION $\begin{bmatrix}1\\0\end{bmatrix}$ rotates into $\begin{bmatrix}\cos\varphi\\ \sin\varphi\end{bmatrix}$, and $\begin{bmatrix}0\\1\end{bmatrix}$ rotates into $\begin{bmatrix}-\sin\varphi\\ \cos\varphi\end{bmatrix}$. See Figure 1. By Theorem 10,

$$A = \begin{bmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{bmatrix}$$

Example 5 in Section 1.8 is a special case of this transformation, with $\varphi = \pi/2$. ■

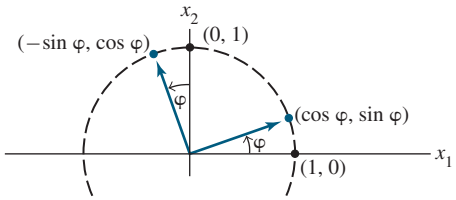


FIGURE 1 A rotation transformation.

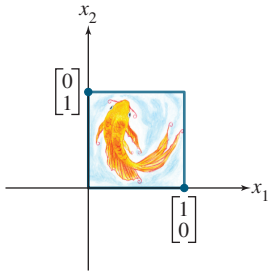


FIGURE 2
The unit square.

Geometric Linear Transformations of $\mathbb{R}^2$

Examples 2 and 3 illustrate linear transformations that are described geometrically. Tables 1–4 illustrate other common geometric linear transformations of the plane. Because the transformations are linear, they are determined completely by what they do to the columns of I_2. Instead of showing only the images of $\mathbf{e}_1$ and $\mathbf{e}_2$, the tables show what a transformation does to the unit square (Figure 2).

Other transformations can be constructed from those listed in Tables 1–4 by applying one transformation after another. For instance, a horizontal shear could be followed by a reflection in the x_2-axis. Section 2.1 will show that such a *composition* of linear transformations is linear. (Also, see Exercise 44.)

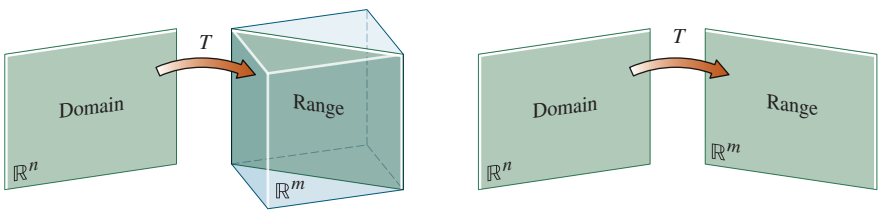
Existence and Uniqueness Questions

The concept of a linear transformation provides a new way to understand the existence and uniqueness questions asked earlier. The next two definitions give the appropriate terminology for transformations.

DEFINITION

A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **onto** $\mathbb{R}^m$ if each $\mathbf{b}$ in $\mathbb{R}^m$ is the image of *at least one* $\mathbf{x}$ in $\mathbb{R}^n$.

Equivalently, T is onto $\mathbb{R}^m$ when the range of T is all of the codomain $\mathbb{R}^m$. That is, T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ if, for each $\mathbf{b}$ in the codomain $\mathbb{R}^m$, there exists at least one solution of $T(\mathbf{x}) = \mathbf{b}$. "Does T map $\mathbb{R}^n$ onto $\mathbb{R}^m$?" is an existence question. The mapping T is *not* onto when there is some $\mathbf{b}$ in $\mathbb{R}^m$ for which the equation $T(\mathbf{x}) = \mathbf{b}$ has no solution. See Figure 3.



T is *not* onto $\mathbb{R}^m$ T is onto $\mathbb{R}^m$

FIGURE 3 Is the range of T all of $\mathbb{R}^m$?

TABLE 1 Reflections

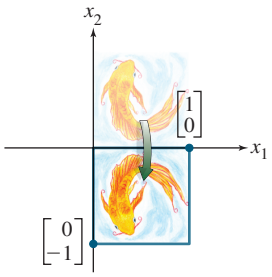
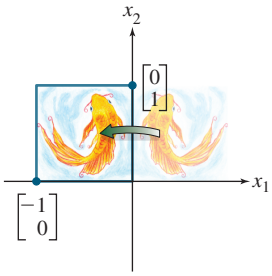
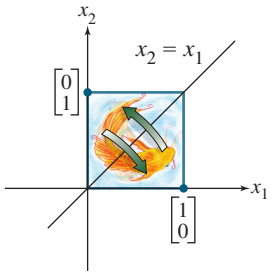
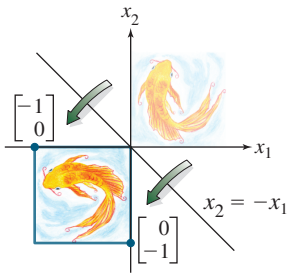
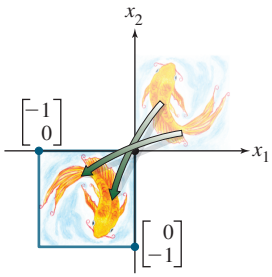
Transformation	Image of the Unit Square	Standard Matrix
Reflection through the x_1-axis		$\begin{bmatrix}1 & 0\\0 & -1\end{bmatrix}$
Reflection through the x_2-axis		$\begin{bmatrix}-1 & 0\\0 & 1\end{bmatrix}$
Reflection through the line $x_2 = x_1$		$\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}$
Reflection through the line $x_2 = -x_1$		$\begin{bmatrix}0 & -1\\-1 & 0\end{bmatrix}$
Reflection through the origin		$\begin{bmatrix}-1 & 0\\0 & -1\end{bmatrix}$

TABLE 2 Contractions and Expansions

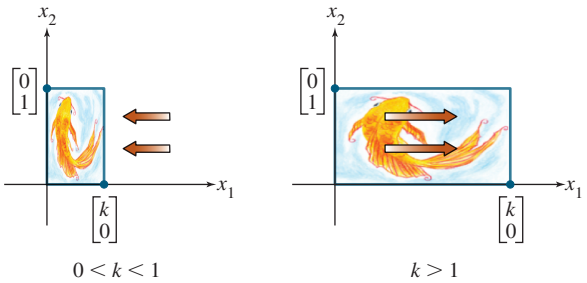
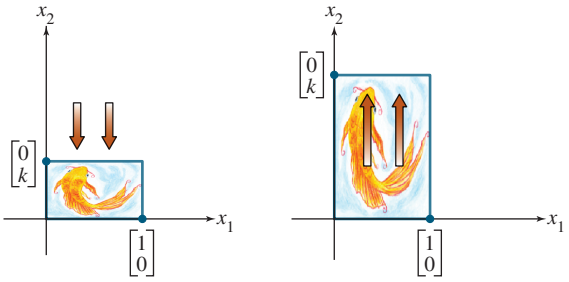
Transformation	Image of the Unit Square	Standard Matrix
Horizontal contraction and expansion		$\begin{bmatrix}k & 0\\0 & 1\end{bmatrix}$
Vertical contraction and expansion		$\begin{bmatrix}1 & 0\\0 & k\end{bmatrix}$

TABLE 3 Shears

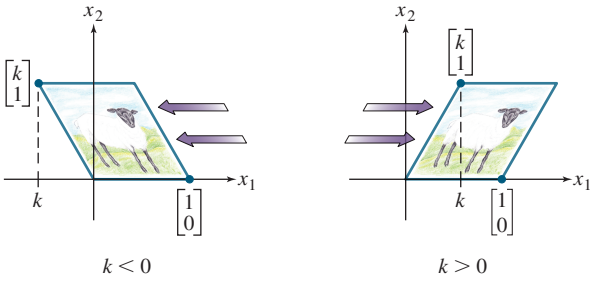
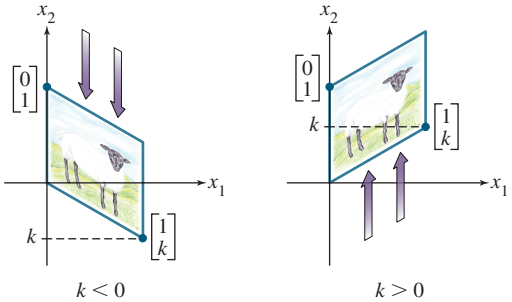
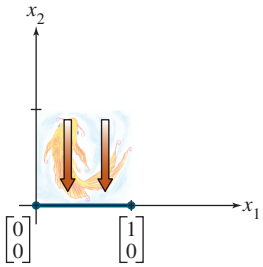
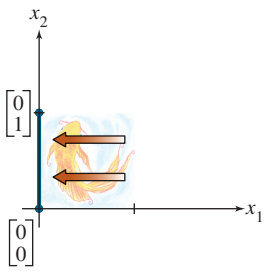
Transformation	Image of the Unit Square	Standard Matrix
Horizontal shear		$\begin{bmatrix}1 & k\\0 & 1\end{bmatrix}$
Vertical shear		$\begin{bmatrix}1 & 0\\k & 1\end{bmatrix}$

TABLE 4 Projections

Transformation	Image of the Unit Square	Standard Matrix
Projection onto the x_1-axis		$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$
Projection onto the x_2-axis		$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$

DEFINITION

A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **one-to-one** if each $\mathbf{b}$ in $\mathbb{R}^m$ is the image of *at most one* $\mathbf{x}$ in $\mathbb{R}^n$.

Equivalently, T is one-to-one if, for each $\mathbf{b}$ in $\mathbb{R}^m$, the equation $T(\mathbf{x}) = \mathbf{b}$ has either a unique solution or none at all. "Is T one-to-one?" is a uniqueness question. The mapping T is *not* one-to-one when some $\mathbf{b}$ in $\mathbb{R}^m$ is the image of more than one vector in $\mathbb{R}^n$. If there is no such $\mathbf{b}$, then T is one-to-one. See Figure 4.

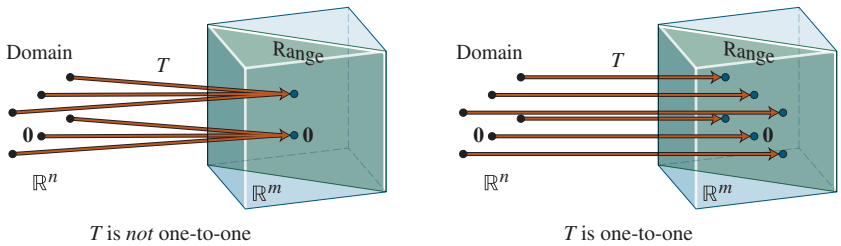


FIGURE 4 Is every $\mathbf{b}$ the image of at most one vector?

The projection transformations shown in Table 4 are *not* one-to-one and do *not* map $\mathbb{R}^2$ onto $\mathbb{R}^2$. The transformations in Tables 1, 2, and 3 are one-to-one *and* do map $\mathbb{R}^2$ onto $\mathbb{R}^2$. Other possibilities are shown in the two examples below.

Example 4 and the theorems that follow show how the function properties of being one-to-one and mapping onto are related to important concepts studied earlier in this chapter.

EXAMPLE 4 Let T be the linear transformation whose standard matrix is

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

Does T map $\mathbb{R}^4$ onto $\mathbb{R}^3$? Is T a one-to-one mapping?

SOLUTION Since A happens to be in echelon form, we can see at once that A has a pivot position in each row. By Theorem 4 in Section 1.4, for each $\mathbf{b}$ in $\mathbb{R}^3$, the equation $A\mathbf{x} = \mathbf{b}$ is consistent. In other words, the linear transformation T maps $\mathbb{R}^4$ (its domain) onto $\mathbb{R}^3$. However, since the equation $A\mathbf{x} = \mathbf{b}$ has a free variable (because there are four variables and only three basic variables), each $\mathbf{b}$ is the image of more than one $\mathbf{x}$. That is, T is *not* one-to-one. ■

THEOREM 11

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then T is one-to-one if and only if the equation $T(\mathbf{x}) = \mathbf{0}$ has only the trivial solution.

Remark: To prove a theorem that says "statement P is true if and only if statement Q is true," one must establish two things: (1) If P is true, then Q is true and (2) If Q is true, then P is true. The second requirement can also be established by showing (2a): If P is false, then Q is false. (This is called contrapositive reasoning.) This proof uses (1) and (2a) to show that P and Q are either both true or both false.

PROOF Since T is linear, $T(\mathbf{0}) = \mathbf{0}$. If T is one-to-one, then the equation $T(\mathbf{x}) = \mathbf{0}$ has at most one solution and hence only the trivial solution. If T is not one-to-one, then there is a $\mathbf{b}$ that is the image of at least two different vectors in $\mathbb{R}^n$—say, $\mathbf{u}$ and $\mathbf{v}$. That is, $T(\mathbf{u}) = \mathbf{b}$ and $T(\mathbf{v}) = \mathbf{b}$. But then, since T is linear,

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

The vector $\mathbf{u} - \mathbf{v}$ is not zero, since $\mathbf{u} \neq \mathbf{v}$. Hence the equation $T(\mathbf{x}) = \mathbf{0}$ has more than one solution. So, either the two conditions in the theorem are both true or they are both false. ■

THEOREM 12

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and let A be the standard matrix for T. Then:

a. T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ if and only if the columns of A span $\mathbb{R}^m$;

b. T is one-to-one if and only if the columns of A are linearly independent.

Remark: "If and only if" statements can be linked together. For example if "P if and only if Q" is known and "Q if and only if R" is known, then one can conclude "P if and only if R." This strategy is used repeatedly in this proof.

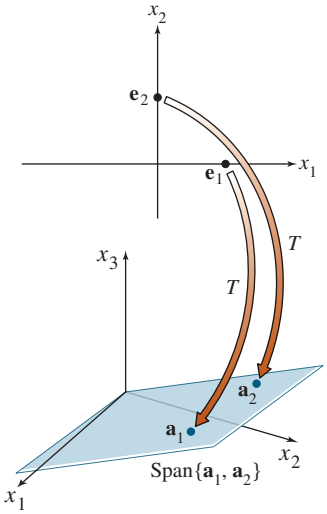
PROOF

a. By Theorem 4 in Section 1.4, the columns of A span $\mathbb{R}^m$ if and only if for each $\mathbf{b}$ in $\mathbb{R}^m$ the equation $A\mathbf{x} = \mathbf{b}$ is consistent—in other words, if and only if for every $\mathbf{b}$, the equation $T(\mathbf{x}) = \mathbf{b}$ has at least one solution. This is true if and only if T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$.

b. The equations $T(\mathbf{x}) = \mathbf{0}$ and $A\mathbf{x} = \mathbf{0}$ are the same except for notation. So, by Theorem 11, T is one-to-one if and only if $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. This happens if and only if the columns of A are linearly independent, as was already noted in the boxed statement (3) in Section 1.7. ■

Statement (a) in Theorem 12 is equivalent to the statement "T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ if and only if every vector in $\mathbb{R}^m$ is a linear combination of the columns of A." See Theorem 4 in Section 1.4.

In the next example and in some exercises that follow, column vectors are written in rows, such as $\mathbf{x} = (x_1, x_2)$, and $T(\mathbf{x})$ is written as $T(x_1, x_2)$ instead of the more formal $T((x_1, x_2))$.



The transformation T is not onto $\mathbb{R}^3$.

EXAMPLE 5 Let $T(x_1, x_2) = (3x_1 + x_2, 5x_1 + 7x_2, x_1 + 3x_2)$. Show that T is a one-to-one linear transformation. Does T map $\mathbb{R}^2$ onto $\mathbb{R}^3$?

SOLUTION When $\mathbf{x}$ and $T(\mathbf{x})$ are written as column vectors, you can determine the standard matrix of T by inspection, visualizing the row–vector computation of each entry in $A\mathbf{x}$.

$$T(\mathbf{x}) = \begin{bmatrix} 3x_1 + x_2 \\ 5x_1 + 7x_2 \\ x_1 + 3x_2 \end{bmatrix} = \underbrace{\begin{bmatrix} ? & ? \\ ? & ? \\ ? & ? \end{bmatrix}}_{A} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 5 & 7 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{4}$$

So T is indeed a linear transformation, with its standard matrix A shown in (4). The columns of A are linearly independent because they are not multiples. By Theorem 12(b), T is one-to-one. To decide if T is onto $\mathbb{R}^3$, examine the span of the columns of A. Since A is 3×2, the columns of A span $\mathbb{R}^3$ if and only if A has 3 pivot positions, by Theorem 4. This is impossible, since A has only 2 columns. So the columns of A do not span $\mathbb{R}^3$, and the associated linear transformation is not onto $\mathbb{R}^3$. ■

Practice Problems

1. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the transformation that first performs a horizontal shear that maps $\mathbf{e}_2$ into $\mathbf{e}_2 - .5\mathbf{e}_1$ (but leaves $\mathbf{e}_1$ unchanged) and then reflects the result through the x_2-axis. Assuming that T is linear, find its standard matrix. [*Hint:* Determine the final location of the images of $\mathbf{e}_1$ and $\mathbf{e}_2$.]
2. Suppose A is a 7×5 matrix with 5 pivots. Let $T(\mathbf{x}) = A\mathbf{x}$ be a linear transformation from $\mathbb{R}^5$ into $\mathbb{R}^7$. Is T a one-to-one linear transformation? Is T onto $\mathbb{R}^7$?

1.9 Exercises

In Exercises 1–10, assume that T is a linear transformation. Find the standard matrix of T.

1. $T : \mathbb{R}^2 \to \mathbb{R}^4$, $T(\mathbf{e}_1) = (2, 1, 2, 1)$ and $T(\mathbf{e}_2) = (-5, 2, 0, 0)$, where $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$.
2. $T : \mathbb{R}^3 \to \mathbb{R}^2$, $T(\mathbf{e}_1) = (1, 3)$, $T(\mathbf{e}_2) = (4, 2)$, and $T(\mathbf{e}_3) = (-5, 4)$, where $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ are the columns of the 3×3 identity matrix.
3. $T : \mathbb{R}^2 \to \mathbb{R}^2$ rotates points (about the origin) through $3\pi/2$ radians (in the counterclockwise direction).

4. $T : \mathbb{R}^2 \to \mathbb{R}^2$ rotates points (about the origin) through $-\pi/4$ radians (since the number is negative, the actual rotation is clockwise). [*Hint:* $T(\mathbf{e}_1) = (1/\sqrt{2}, -1/\sqrt{2})$.]

5. $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a vertical shear transformation that maps $\mathbf{e}_1$ into $\mathbf{e}_1 - 2\mathbf{e}_2$ but leaves the vector $\mathbf{e}_2$ unchanged.

6. $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a horizontal shear transformation that leaves $\mathbf{e}_1$ unchanged and maps $\mathbf{e}_2$ into $\mathbf{e}_2 + 5\mathbf{e}_1$.

7. $T : \mathbb{R}^2 \to \mathbb{R}^2$ first rotates points through $-3\pi/4$ radians (since the number is negative, the actual rotation is clockwise) and then reflects points through the horizontal x_1-axis. [*Hint:* $T(\mathbf{e}_1) = (-1/\sqrt{2}, 1/\sqrt{2})$.]

8. $T : \mathbb{R}^2 \to \mathbb{R}^2$ first reflects points through the vertical x_2-axis and then reflects points through the line $x_2 = x_1$.

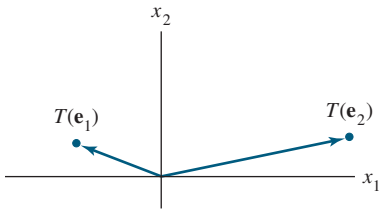
9. $T : \mathbb{R}^2 \to \mathbb{R}^2$ first performs a horizontal shear that transforms $\mathbf{e}_2$ into $\mathbf{e}_2 - 3\mathbf{e}_1$ (leaving $\mathbf{e}_1$ unchanged) and then reflects points through the line $x_2 = -x_1$.

10. $T : \mathbb{R}^2 \to \mathbb{R}^2$ first reflects points through the vertical x_2-axis and then rotates points $3\pi/2$ radians.

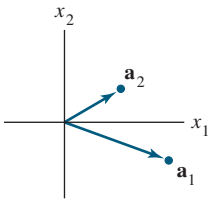
11. A linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ first reflects points through the x_1-axis and then reflects points through the x_2-axis. Show that T can also be described as a linear transformation that rotates points about the origin. What is the angle of that rotation?

12. Show that the transformation in Exercise 8 is merely a rotation about the origin. What is the angle of the rotation?

13. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation such that $T(\mathbf{e}_1)$ and $T(\mathbf{e}_2)$ are the vectors shown in the figure. Using the figure, sketch the vector $T(2, 1)$.



14. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation with standard matrix $A = [\mathbf{a}_1 \quad \mathbf{a}_2]$, where $\mathbf{a}_1$ and $\mathbf{a}_2$ are shown in the figure. Using the figure, draw the image of $\begin{bmatrix} -1 \\ 3 \end{bmatrix}$ under the transformation T.



In Exercises 15 and 16, fill in the missing entries of the matrix, assuming that the equation holds for all values of the variables.

15. $\begin{bmatrix} ? & ? & ? \\ ? & ? & ? \\ ? & ? & ? \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2x_1 - 3x_3 \\ 4x_1 \\ x_1 - x_2 + x_3 \end{bmatrix}$

16. $\begin{bmatrix} ? & ? \\ ? & ? \\ ? & ? \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - 3x_2 \\ -2x_1 + x_2 \\ x_1 \end{bmatrix}$

In Exercises 17–20, show that T is a linear transformation by finding a matrix that implements the mapping. Note that $x_1, x_2, \ldots$ are not vectors but are entries in vectors.

17. $T(x_1, x_2, x_3, x_4) = (0, x_1 + x_2, x_2 + x_3, x_3 + x_4)$

18. $T(x_1, x_2) = (2x_2 - 3x_1, x_1 - 4x_2, 0, x_2)$

19. $T(x_1, x_2, x_3) = (x_1 - 5x_2 + 4x_3, x_2 - 6x_3)$

20. $T(x_1, x_2, x_3, x_4) = 2x_1 + 3x_3 - 4x_4 \qquad (T : \mathbb{R}^4 \to \mathbb{R})$

21. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation such that $T(x_1, x_2) = (x_1 + x_2, 4x_1 + 5x_2)$. Find $\mathbf{x}$ such that $T(\mathbf{x}) = (3, 8)$.

22. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be a linear transformation such that $T(x_1, x_2) = (x_1 - 2x_2, -x_1 + 3x_2, 3x_1 - 2x_2)$. Find $\mathbf{x}$ such that $T(\mathbf{x}) = (-1, 4, 9)$.

In Exercises 23–32, mark each statement True or False **(T/F)**. Justify each answer.

23. **(T/F)** A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is completely determined by its effect on the columns of the $n \times n$ identity matrix.

24. **(T/F)** A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is one-to-one if each vector in $\mathbb{R}^n$ maps onto a unique vector in $\mathbb{R}^m$.

25. **(T/F)** If $T : \mathbb{R}^2 \to \mathbb{R}^2$ rotates vectors about the origin through an angle ϕ, then T is a linear transformation.

26. **(T/F)** The columns of the standard matrix for a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ are the images of the columns of the $n \times n$ identity matrix.

27. **(T/F)** When two linear transformations are performed one after another, the combined effect may not always be a linear transformation.

28. **(T/F)** Not every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is a matrix transformation.

29. **(T/F)** A mapping $T : \mathbb{R}^n \to \mathbb{R}^m$ is onto $\mathbb{R}^m$ if every vector $\mathbf{x}$ in $\mathbb{R}^n$ maps onto some vector in $\mathbb{R}^m$.

30. **(T/F)** The standard matrix of a linear transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$ that reflects points through the horizontal axis, the vertical axis, or the origin has the form $\begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}$, where a and d are ± 1.

31. **(T/F)** A is a 3×2 matrix, then the transformation $\mathbf{x} \mapsto A\mathbf{x}$ cannot be one-to-one.

32. **(T/F)** A is a 3×2 matrix, then the transformation $\mathbf{x} \mapsto A\mathbf{x}$ cannot map $\mathbb{R}^2$ onto $\mathbb{R}^3$.

In Exercises 33–36, determine if the specified linear transformation is (a) one-to-one and (b) onto. Justify each answer.

33. The transformation in Exercise 17

34. The transformation in Exercise 2

35. The transformation in Exercise 19

36. The transformation in Exercise 14

In Exercises 37 and 38, describe the possible echelon forms of the standard matrix for a linear transformation T. Use the notation of Example 1 in Section 1.2.

37. $T : \mathbb{R}^3 \to \mathbb{R}^4$ is one-to-one.

38. $T : \mathbb{R}^4 \to \mathbb{R}^3$ is onto.

39. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, with A its standard matrix. Complete the following statement to make it true: "T is one-to-one if and only if A has ____ pivot columns." Explain why the statement is true. [*Hint:* Look in the exercises for Section 1.7.]

40. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, with A its standard matrix. Complete the following statement to make it true: "T maps $\mathbb{R}^n$ onto $\mathbb{R}^m$ if and only if A has ____ pivot columns." Find some theorems that explain why the statement is true.

41. Verify the uniqueness of A in Theorem 10. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation such that $T(\mathbf{x}) = B\mathbf{x}$ for some $m \times n$ matrix B. Show that if A is the standard matrix for T, then $A = B$. [*Hint:* Show that A and B have the same columns.]

42. Why is the question "Is the linear transformation T onto?" an existence question?

43. If a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ maps $\mathbb{R}^n$ *onto* $\mathbb{R}^m$, can you give a relation between m and n? If T is one-to-one, what can you say about m and n?

44. Let $S : \mathbb{R}^p \to \mathbb{R}^n$ and $T : \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations. Show that the mapping $\mathbf{x} \mapsto T(S(\mathbf{x}))$ is a linear transformation (from $\mathbb{R}^p$ to $\mathbb{R}^m$). [*Hint:* Compute $T(S(c\mathbf{u} + d\mathbf{v}))$ for $\mathbf{u}, \mathbf{v}$ in $\mathbb{R}^p$ and scalars c and d. Justify each step of the computation, and explain why this computation gives the desired conclusion.]

T In Exercises 45–48, let T be the linear transformation whose standard matrix is given. In Exercises 45 and 46, decide if T is a one-to-one mapping. In Exercises 47 and 48, decide if T maps $\mathbb{R}^5$ onto $\mathbb{R}^5$. Justify your answers.

45. $\begin{bmatrix} -5 & 10 & -5 & 4 \\ 8 & 3 & -4 & 7 \\ 4 & -9 & 5 & -3 \\ -3 & -2 & 5 & 4 \end{bmatrix}$

46. $\begin{bmatrix} 7 & 5 & 4 & -9 \\ 10 & 6 & 16 & -4 \\ 12 & 8 & 12 & 7 \\ -8 & -6 & -2 & 5 \end{bmatrix}$

47. $\begin{bmatrix} 4 & -7 & 3 & 7 & 5 \\ 6 & -8 & 5 & 12 & -8 \\ -7 & 10 & -8 & -9 & 14 \\ 3 & -5 & 4 & 2 & -6 \\ -5 & 6 & -6 & -7 & 3 \end{bmatrix}$

48. $\begin{bmatrix} 9 & 13 & 5 & 6 & -1 \\ 14 & 15 & -7 & -6 & 4 \\ -8 & -9 & 12 & -5 & -9 \\ -5 & -6 & -8 & 9 & 8 \\ 13 & 14 & 15 & 2 & 11 \end{bmatrix}$

STUDY GUIDE offers additional resources for mastering existence and uniqueness.

Solution to Practice Problems

1. Follow what happens to $\mathbf{e}_1$ and $\mathbf{e}_2$. See Figure 5. First, $\mathbf{e}_1$ is unaffected by the shear and then is reflected into $-\mathbf{e}_1$. So $T(\mathbf{e}_1) = -\mathbf{e}_1$. Second, $\mathbf{e}_2$ goes to $\mathbf{e}_2 - .5\mathbf{e}_1$ by the shear transformation. Since reflection through the x_2-axis changes $\mathbf{e}_1$ into $-\mathbf{e}_1$ and leaves $\mathbf{e}_2$ unchanged, the vector $\mathbf{e}_2 - .5\mathbf{e}_1$ goes to $\mathbf{e}_2 + .5\mathbf{e}_1$. So $T(\mathbf{e}_2) = \mathbf{e}_2 + .5\mathbf{e}_1$.

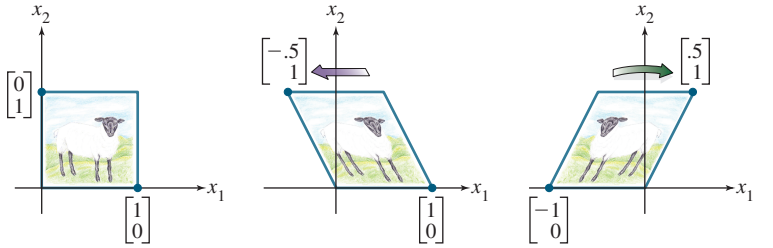
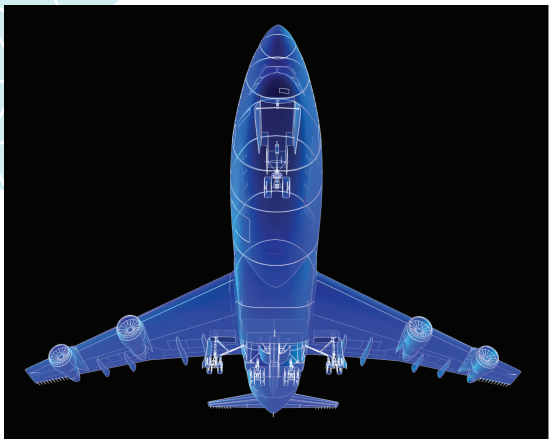


FIGURE 5 The composition of two transformations.

2 Matrix Algebra



Introductory Example

COMPUTER MODELS IN AIRCRAFT DESIGN

To design the next generation of commercial and military aircraft, engineers at Boeing's Phantom Works use 3D modeling and computational fluid dynamics (CFD). They study the airflow around a virtual airplane to answer important design questions before physical models are created. This has drastically reduced design cycle times and cost—and linear algebra plays a crucial role in the process.

The virtual airplane begins as a mathematical "wire-frame" model that exists only in computer memory and on graphics display terminals. (Model of a Boeing 747 is shown.) This mathematical model organizes and influences each step of the design and manufacture of the airplane—both the exterior and interior. The CFD analysis concerns the exterior surface.

Although the finished skin of a plane may seem smooth, the geometry of the surface is complicated. In addition to wings and a fuselage, an aircraft has nacelles, stabilizers, slats, flaps, and ailerons. The way air flows around these structures determines how the plane moves through the sky. Equations that describe the airflow are complicated, and they must account for engine intake, engine exhaust, and the wakes left by the wings of the plane. To study the airflow, engineers need a highly refined description of the plane's surface.

A computer creates a model of the surface by first superimposing a three-dimensional grid of "boxes" on the original wire-frame model. Boxes in this grid lie either completely inside or completely outside the plane, or they intersect the surface of the plane. The computer selects the boxes that intersect the surface and subdivides them, retaining only the smaller boxes that still intersect the surface. The subdividing process is repeated until the grid is extremely fine. A typical grid can include more than 400,000 boxes.

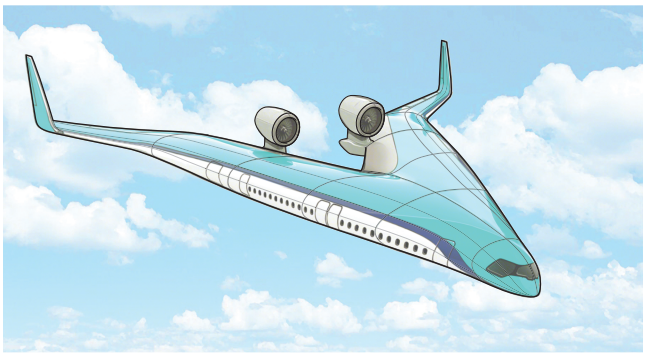
The process for finding the airflow around the plane involves repeatedly solving a system of linear equations $A\mathbf{x} = \mathbf{b}$ that may involve up to 2 million equations and variables. The vector $\mathbf{b}$ changes each time, based on data from the grid and solutions of previous equations. Using the fastest computers available commercially, a Phantom Works team can spend from a few hours to several days setting up and solving a single airflow problem. After the team analyzes the solution, they may make small changes to the airplane surface and begin the whole process again. Thousands of CFD runs may be required.

This chapter presents two important concepts that assist in the solution of such massive systems of equations:

- *Partitioned matrices:* A typical CFD system of equations has a "sparse" coefficient matrix with mostly zero entries. Grouping the variables correctly leads to a partitioned matrix with many zero blocks. Section 2.4 introduces such matrices and describes some of their applications.

- *Matrix factorizations:* Even when written with partitioned matrices, the system of equations is complicated. To further simplify the computations, the CFD software at Boeing uses what is called an LU factorization of the coefficient matrix. Section 2.5 discusses LU and other useful matrix factorizations. Further details about factorizations appear at several points later in the text.

To analyze a solution of an airflow system, engineers want to visualize the airflow over the surface of the plane. They use computer graphics, and linear algebra provides the engine for the graphics. The wire-frame model of the plane's surface is stored as data in many matrices. Once the image has been rendered on a computer screen, engineers can change its scale, zoom in or out of small regions, and rotate the image to see parts that may be hidden from view.



TU-Delft and Air France-KLM are investigating a flying V aircraft design because of its potential for significantly better fuel economy.

Each of these operations is accomplished by appropriate matrix multiplications. Section 2.7 explains the basic ideas.

Our ability to analyze and solve equations will be greatly enhanced when we can perform algebraic operations with matrices. Furthermore, the definitions and theorems in this chapter provide some basic tools for handling the many applications of linear algebra that involve two or more matrices. For $n \times n$ matrices, the Invertible Matrix Theorem in Section 2.3 ties together most of the concepts treated earlier in the text. Sections 2.4 and 2.5 examine partitioned matrices and matrix factorizations, which appear in most modern uses of linear algebra. Sections 2.6 and 2.7 describe two interesting applications of matrix algebra: to economics and to computer graphics. Sections 2.8 and 2.9 provide readers enough information about subspaces to move directly into Chapters 5, 6, and 7, without covering Chapter 4. You may want to omit these two sections if you plan to cover Chapter 4 before moving to Chapter 5.

2.1 Matrix Operations

If A is an $m \times n$ matrix—that is, a matrix with m rows and n columns—then the scalar entry in the ith row and jth column of A is denoted by a_{ij} and is called the (i, j)-entry of A. See Figure 1. For instance, the $(3, 2)$-entry is the number a_{32} in the third row, second column. Each column of A is a list of m real numbers, which identifies a vector in $\mathbb{R}^m$. Often, these columns are denoted by $\mathbf{a}_1, \ldots, \mathbf{a}_n$, and the matrix A is written as

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$$

Observe that the number a_{ij} is the ith entry (from the top) of the jth column vector $\mathbf{a}_j$.

The **diagonal entries** in an $m \times n$ matrix $A = [\, a_{ij} \,]$ are $a_{11}, a_{22}, a_{33}, \ldots$, and they form the **main diagonal** of A. A **diagonal matrix** is a square $n \times n$ matrix whose nondiagonal entries are zero. An example is the $n \times n$ identity matrix, I_n. An $m \times n$ matrix whose entries are all zero is a **zero matrix** and is written as 0. The size of a zero matrix is usually clear from the context.

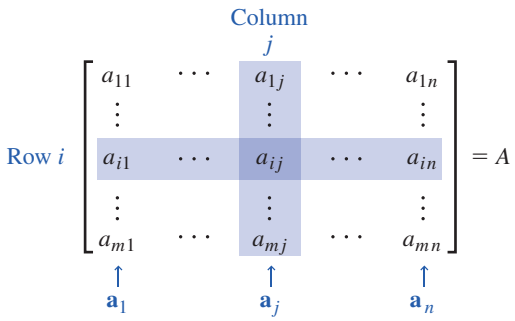


FIGURE 1 Matrix notation.

Sums and Scalar Multiples

The arithmetic for vectors described earlier has a natural extension to matrices. We say that two matrices are **equal** if they have the same size (i.e., the same number of rows and the same number of columns) and if their corresponding columns are equal, which amounts to saying that their corresponding entries are equal. If A and B are $m \times n$ matrices, then the **sum** $A + B$ is the $m \times n$ matrix whose columns are the sums of the corresponding columns in A and B. Since vector addition of the columns is done entrywise, each entry in $A + B$ is the sum of the corresponding entries in A and B. The sum $A + B$ is defined only when A and B are the same size.

EXAMPLE 1 Let

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

Then

$$A + B = \begin{bmatrix} 5 & 1 & 6 \\ 2 & 8 & 9 \end{bmatrix}$$

but $A + C$ is not defined because A and C have different sizes. ■

If r is a scalar and A is a matrix, then the **scalar multiple** rA is the matrix whose columns are r times the corresponding columns in A. As with vectors, $-A$ stands for $(-1)A$, and $A - B$ is the same as $A + (-1)B$.

EXAMPLE 2 If A and B are the matrices in Example 1, then

$$2B = 2\begin{bmatrix} 1 & 1 & 1 \\ 3 & 5 & 7 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 2 \\ 6 & 10 & 14 \end{bmatrix}$$

$$A - 2B = \begin{bmatrix} 4 & 0 & 5 \\ -1 & 3 & 2 \end{bmatrix} - \begin{bmatrix} 2 & 2 & 2 \\ 6 & 10 & 14 \end{bmatrix} = \begin{bmatrix} 2 & -2 & 3 \\ -7 & -7 & -12 \end{bmatrix}$$ ■

It was unnecessary in Example 2 to compute $A - 2B$ as $A + (-1)2B$ because the usual rules of algebra apply to sums and scalar multiples of matrices, as the following theorem shows.

THEOREM 1

Let A, B, and C be matrices of the same size, and let r and s be scalars.

a. $A + B = B + A$
b. $(A + B) + C = A + (B + C)$
c. $A + 0 = A$
d. $r(A + B) = rA + rB$
e. $(r + s)A = rA + sA$
f. $r(sA) = (rs)A$

Each equality in Theorem 1 is verified by showing that the matrix on the left side has the same size as the matrix on the right and that corresponding columns are equal. Size is no problem because A, B, and C are equal in size. The equality of columns follows immediately from analogous properties of vectors. For instance, if the jth columns of A, B, and C are $\mathbf{a}_j$, $\mathbf{b}_j$, and $\mathbf{c}_j$, respectively, then the jth columns of $(A + B) + C$ and $A + (B + C)$ are

$$(\mathbf{a}_j + \mathbf{b}_j) + \mathbf{c}_j \quad \text{and} \quad \mathbf{a}_j + (\mathbf{b}_j + \mathbf{c}_j)$$

respectively. Since these two vector sums are equal for each j, property (b) is verified.

Because of the associative property of addition, we can simply write $A + B + C$ for the sum, which can be computed either as $(A + B) + C$ or as $A + (B + C)$. The same applies to sums of four or more matrices.

Matrix Multiplication

When a matrix B multiplies a vector $\mathbf{x}$, it transforms $\mathbf{x}$ into the vector $B\mathbf{x}$. If this vector is then multiplied in turn by a matrix A, the resulting vector is $A(B\mathbf{x})$. See Figure 2.

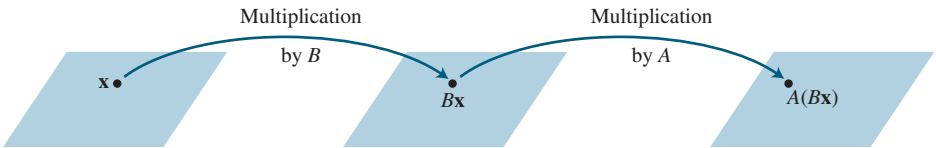


FIGURE 2 Multiplication by B and then A.

Thus $A(B\mathbf{x})$ is produced from $\mathbf{x}$ by a *composition* of mappings—the linear transformations studied in Section 1.8. Our goal is to represent this composite mapping as multiplication by a single matrix, denoted by AB, so that

$$A(B\mathbf{x}) = (AB)\mathbf{x} \tag{1}$$

See Figure 3.

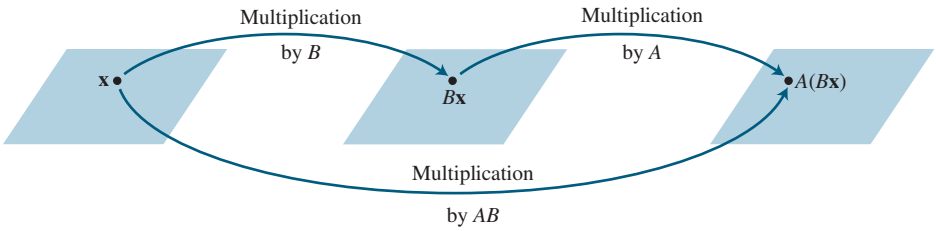


FIGURE 3 Multiplication by AB.

If A is $m \times n$, B is $n \times p$, and $\mathbf{x}$ is in $\mathbb{R}^p$, denote the columns of B by $\mathbf{b}_1, \ldots, \mathbf{b}_p$ and the entries in $\mathbf{x}$ by $x_1, \ldots, x_p$. Then

$$B\mathbf{x} = x_1\mathbf{b}_1 + \cdots + x_p\mathbf{b}_p$$

By the linearity of multiplication by A,

$$\begin{aligned} A(B\mathbf{x}) &= A(x_1\mathbf{b}_1) + \cdots + A(x_p\mathbf{b}_p) \\ &= x_1A\mathbf{b}_1 + \cdots + x_pA\mathbf{b}_p \end{aligned}$$

The vector $A(B\mathbf{x})$ is a linear combination of the vectors $A\mathbf{b}_1, \ldots, A\mathbf{b}_p$, using the entries in $\mathbf{x}$ as weights. In matrix notation, this linear combination is written as

$$A(B\mathbf{x}) = [\, A\mathbf{b}_1 \quad A\mathbf{b}_2 \quad \cdots \quad A\mathbf{b}_p \,]\mathbf{x}$$

Thus multiplication by $[\, A\mathbf{b}_1 \quad A\mathbf{b}_2 \quad \cdots \quad A\mathbf{b}_p \,]$ transforms $\mathbf{x}$ into $A(B\mathbf{x})$. We have found the matrix we sought!

DEFINITION

If A is an $m \times n$ matrix, and if B is an $n \times p$ matrix with columns $\mathbf{b}_1, \ldots, \mathbf{b}_p$, then the product AB is the $m \times p$ matrix whose columns are $A\mathbf{b}_1, \ldots, A\mathbf{b}_p$. That is,

$$AB = A\big[\mathbf{b}_1 \quad \mathbf{b}_2 \quad \cdots \quad \mathbf{b}_p\big] = \big[A\mathbf{b}_1 \quad A\mathbf{b}_2 \quad \cdots \quad A\mathbf{b}_p\big]$$

This definition makes equation (1) true for all $\mathbf{x}$ in $\mathbb{R}^p$. Equation (1) proves that the composite mapping in Figure 3 is a linear transformation and that its standard matrix is AB. *Multiplication of matrices corresponds to composition of linear transformations.*

EXAMPLE 3 Compute AB, where $A = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 3 & 6 \\ 1 & -2 & 3 \end{bmatrix}$.

SOLUTION Write $B = [\, \mathbf{b}_1 \quad \mathbf{b}_2 \quad \mathbf{b}_3 \,]$, and compute:

$$A\mathbf{b}_1 = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 11 \\ -1 \end{bmatrix}, \quad A\mathbf{b}_2 = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 13 \end{bmatrix}, \quad A\mathbf{b}_3 = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 6 \\ 3 \end{bmatrix} = \begin{bmatrix} 21 \\ -9 \end{bmatrix}$$

Then

$$AB = A[\, \mathbf{b}_1 \quad \mathbf{b}_2 \quad \mathbf{b}_3 \,] = \begin{bmatrix} 11 & 0 & 21 \\ -1 & 13 & -9 \end{bmatrix}$$

$A\mathbf{b}_1 \quad A\mathbf{b}_2 \quad A\mathbf{b}_3$

■

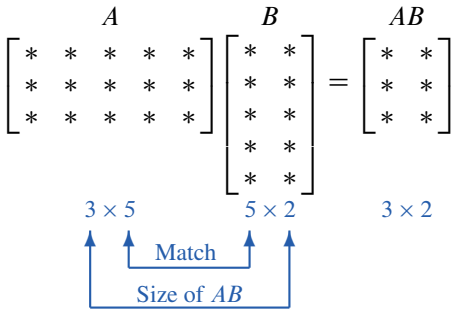
Notice that since the first column of AB is $A\mathbf{b}_1$, this column is a linear combination of the columns of A using the entries in $\mathbf{b}_1$ as weights. A similar statement is true for each column of AB.

Each column of AB is a linear combination of the columns of A using weights from the corresponding column of B.

Obviously, the number of columns of A must match the number of rows in B in order for a linear combination such as $A\mathbf{b}_1$ to be defined. Also, the definition of AB shows that *AB has the same number of rows as A and the same number of columns as B.*

EXAMPLE 4 If A is a 3×5 matrix and B is a 5×2 matrix, what are the sizes of AB and BA, if they are defined?

SOLUTION Since A has 5 columns and B has 5 rows, the product AB is defined and is a 3×2 matrix:



The product BA is *not* defined because the 2 columns of B do not match the 3 rows of A. ■

The definition of AB is important for theoretical work and applications, but the following rule provides a more efficient method for calculating the individual entries in AB when working small problems by hand.

ROW–COLUMN RULE FOR COMPUTING *AB*

If the product AB is defined, then the entry in row i and column j of AB is the sum of the products of corresponding entries from row i of A and column j of B. If $(AB)_{ij}$ denotes the (i, j)-entry in AB, and if A is an $m \times n$ matrix, then

$$(AB)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}$$

To verify this rule, let $B = [\,\mathbf{b}_1 \;\; \cdots \;\; \mathbf{b}_p\,]$. Column j of AB is $A\mathbf{b}_j$, and we can compute $A\mathbf{b}_j$ by the row–vector rule for computing $A\mathbf{x}$ from Section 1.4. The ith entry in $A\mathbf{b}_j$ is the sum of the products of corresponding entries from row i of A and the vector $\mathbf{b}_j$, which is precisely the computation described in the rule for computing the (i, j)-entry of AB.

EXAMPLE 5 Use the row–column rule to compute two of the entries in AB for the matrices in Example 3. An inspection of the numbers involved will make it clear how the two methods for calculating AB produce the same matrix.

SOLUTION To find the entry in row 1 and column 3 of AB, consider row 1 of A and column 3 of B. Multiply corresponding entries and add the results, as shown below:

$$AB = \begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 4 & 3 & 6 \\ 1 & -2 & 3 \end{bmatrix} = \begin{bmatrix} \square & \square & 2(6)+3(3) \\ \square & \square & \square \end{bmatrix} = \begin{bmatrix} \square & \square & 21 \\ \square & \square & \square \end{bmatrix}$$

For the entry in row 2 and column 2 of AB, use row 2 of A and column 2 of B:

$$\begin{bmatrix} 2 & 3 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 4 & 3 & 6 \\ 1 & -2 & 3 \end{bmatrix} = \begin{bmatrix} \square & \square & 21 \\ \square & 1(3)+-5(-2) & \square \end{bmatrix} = \begin{bmatrix} \square & \square & 21 \\ \square & 13 & \square \end{bmatrix}$$

■

EXAMPLE 6 Find the entries in the second row of AB, where

$$A = \begin{bmatrix} 2 & -5 & 0 \\ -1 & 3 & -4 \\ 6 & -8 & -7 \\ -3 & 0 & 9 \end{bmatrix}, \qquad B = \begin{bmatrix} 4 & -6 \\ 7 & 1 \\ 3 & 2 \end{bmatrix}$$

SOLUTION By the row–column rule, the entries of the second row of AB come from row 2 of A (and the columns of B):

$$\begin{bmatrix} 2 & -5 & 0 \\ -1 & 3 & -4 \\ 6 & -8 & -7 \\ -3 & 0 & 9 \end{bmatrix} \begin{bmatrix} 4 & -6 \\ 7 & 1 \\ 3 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} \square & \square \\ -4 + 21 - 12 & 6 + 3 - 8 \\ \square & \square \\ \square & \square \end{bmatrix} = \begin{bmatrix} \square & \square \\ 5 & 1 \\ \square & \square \\ \square & \square \end{bmatrix}$$

■

Notice that since Example 6 requested only the second row of AB, we could have written just the second row of A to the left of B and computed

$$\begin{bmatrix} -1 & 3 & -4 \end{bmatrix} \begin{bmatrix} 4 & -6 \\ 7 & 1 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 1 \end{bmatrix}$$

This observation about rows of AB is true in general and follows from the row–column rule. Let $\text{row}_i(A)$ denote the ith row of a matrix A. Then

$$\text{row}_i(AB) = \text{row}_i(A) \cdot B \tag{2}$$

Properties of Matrix Multiplication

The following theorem lists the standard properties of matrix multiplication. Recall that I_m represents the $m \times m$ identity matrix and $I_m\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^m$.

THEOREM 2

Let A be an $m \times n$ matrix, and let B and C have sizes for which the indicated sums and products are defined.

a. $A(BC) = (AB)C$ (associative law of multiplication)

b. $A(B + C) = AB + AC$ (left distributive law)

c. $(B + C)A = BA + CA$ (right distributive law)

d. $r(AB) = (rA)B = A(rB)$ for any scalar r

e. $I_m A = A = AI_n$ (identity for matrix multiplication)

PROOF Properties (b)–(e) are considered in the exercises. Property (a) follows from the fact that matrix multiplication corresponds to composition of linear transformations (which are functions), and it is known (or easy to check) that the composition of functions

is associative. Here is another proof of (a) that rests on the "column definition" of the product of two matrices. Let

$$C = [\,\mathbf{c}_1 \quad \cdots \quad \mathbf{c}_p\,]$$

By the definition of matrix multiplication,

$$BC = [\,B\mathbf{c}_1 \quad \cdots \quad B\mathbf{c}_p\,]$$
$$A(BC) = [\,A(B\mathbf{c}_1) \quad \cdots \quad A(B\mathbf{c}_p)\,]$$

Recall from equation (1) that the definition of AB makes $A(B\mathbf{x}) = (AB)\mathbf{x}$ for all $\mathbf{x}$, so

$$A(BC) = [\,(AB)\mathbf{c}_1 \quad \cdots \quad (AB)\mathbf{c}_p\,] = (AB)C$$ ■

The associative and distributive laws in Theorems 1 and 2 say essentially that pairs of parentheses in matrix expressions can be inserted and deleted in the same way as in the algebra of real numbers. In particular, we can write ABC for the product, which can be computed either as $A(BC)$ or as $(AB)C$.[1] Similarly, a product $ABCD$ of four matrices can be computed as $A(BCD)$ or $(ABC)D$ or $A(BC)D$, and so on. It does not matter how we group the matrices when computing the product, so long as the left-to-right order of the matrices is preserved.

The left-to-right order in products is critical because AB and BA are usually not the same. This is not surprising, because the columns of AB are linear combinations of the columns of A, whereas the columns of BA are constructed from the columns of B. The position of the factors in the product AB is emphasized by saying that A is *right-multiplied* by B or that B is *left-multiplied* by A. If $AB = BA$, we say that A and B **commute** with one another.

EXAMPLE 7 Let $A = \begin{bmatrix} 5 & 1 \\ 3 & -2 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 0 \\ 4 & 3 \end{bmatrix}$. Show that these matrices do not commute. That is, verify that $AB \neq BA$.

SOLUTION

$$AB = \begin{bmatrix} 5 & 1 \\ 3 & -2 \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 14 & 3 \\ -2 & -6 \end{bmatrix}$$
$$BA = \begin{bmatrix} 2 & 0 \\ 4 & 3 \end{bmatrix}\begin{bmatrix} 5 & 1 \\ 3 & -2 \end{bmatrix} = \begin{bmatrix} 10 & 2 \\ 29 & -2 \end{bmatrix}$$ ■

Example 7 illustrates the first of the following list of important differences between matrix algebra and the ordinary algebra of real numbers. See Exercises 9–12 for examples of these situations.

***Warnings*:**

1. In general, $AB \neq BA$.
2. The cancellation laws do *not* hold for matrix multiplication. That is, if $AB = AC$, then it is *not* true in general that $B = C$. (See Exercise 10.)
3. If a product AB is the zero matrix, you *cannot* conclude in general that either $A = 0$ or $B = 0$. (See Exercise 12.)

[1] When B is square and C has fewer columns than A has rows, it is more efficient to compute $A(BC)$ than $(AB)C$.

Powers of a Matrix

If A is an $n \times n$ matrix and if k is a positive integer, then A^k denotes the product of k copies of A:

$$A^k = \underbrace{A \cdots A}_{k}$$

If A is nonzero and if $\mathbf{x}$ is in $\mathbb{R}^n$, then $A^k\mathbf{x}$ is the result of left-multiplying $\mathbf{x}$ by A repeatedly k times. If $k = 0$, then $A^0\mathbf{x}$ should be $\mathbf{x}$ itself. Thus A^0 is interpreted as the identity matrix. Matrix powers are useful in both theory and applications (Sections 2.6, 5.9, and later in the text).

The Transpose of a Matrix

Given an $m \times n$ matrix A, the **transpose** of A is the $n \times m$ matrix, denoted by A^T, whose columns are formed from the corresponding rows of A.

EXAMPLE 8 Let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad B = \begin{bmatrix} -5 & 2 \\ 1 & -3 \\ 0 & 4 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 & 1 & 1 \\ -3 & 5 & -2 & 7 \end{bmatrix}$$

Then

$$A^T = \begin{bmatrix} a & c \\ b & d \end{bmatrix}, \quad B^T = \begin{bmatrix} -5 & 1 & 0 \\ 2 & -3 & 4 \end{bmatrix}, \quad C^T = \begin{bmatrix} 1 & -3 \\ 1 & 5 \\ 1 & -2 \\ 1 & 7 \end{bmatrix}$$

■

THEOREM 3

Let A and B denote matrices whose sizes are appropriate for the following sums and products.

a. $(A^T)^T = A$

b. $(A + B)^T = A^T + B^T$

c. For any scalar r, $(rA)^T = rA^T$

d. $(AB)^T = B^TA^T$

Proofs of (a)–(c) are straightforward and are omitted. For (d), see Exercise 41. Usually, $(AB)^T$ is not equal to A^TB^T, even when A and B have sizes such that the product A^TB^T is defined.

The generalization of Theorem 3(d) to products of more than two factors can be stated in words as follows:

The transpose of a product of matrices equals the product of their transposes in the *reverse* order.

The exercises contain numerical examples that illustrate properties of transposes.

Artificial intelligence (AI) involves having a computer learn to recognize important information about anything that can be presented in a digitized format. One important area of AI is identifying whether the object in a picture matches a chosen object such as a number, fingerprint, or face.

In the next example, matrix transposition and matrix multiplication are used to tell whether or not a 2×2 block of colored squares matches the chosen checkerboard pattern in Figure 4.

EXAMPLE 9 In order to feed a 2×2 colored block into the computer, it first gets converted into a 4×1 vector by assigning a 1 to each block that is blue and a 0 to each block that is white. Then, the computer converts the block of numbers into a vector by placing the numbers in each column below the numbers in the column to its left.

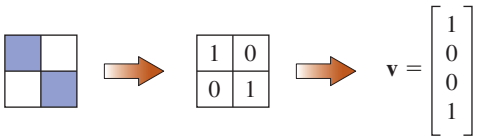
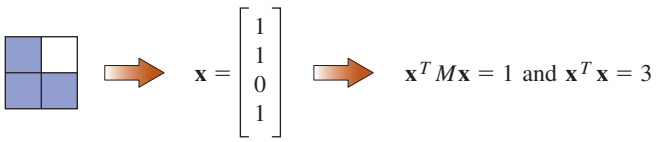


FIGURE 4

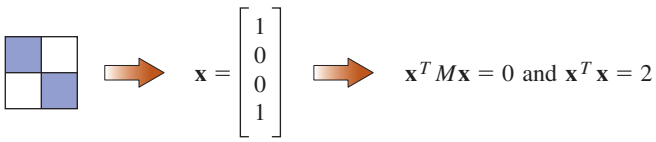
Let $M = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}$.

Notice that $\mathbf{v}^T M \mathbf{v} = \begin{bmatrix} 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} = 0,$

and $\mathbf{w}^T M \mathbf{w} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = 0$, where $\mathbf{w}$ is the vector generated by a 2×2 block of all white squares. It can be verified that for any other vector $\mathbf{x}$ generated from a 2×2 block of white and blue squares, if $\mathbf{x}$ is not $\mathbf{v}$ or $\mathbf{w}$, then the product $\mathbf{x}^T M \mathbf{x}$ is nonzero. Thus, if a computer checks the value of $\mathbf{x}^T M \mathbf{x}$ and finds it is nonzero, the computer knows that the pattern corresponding to $\mathbf{x}$ is not the checkerboard with a blue square in the top left corner.



This pattern is not the checkerboard pattern since $\mathbf{x}^T M \mathbf{x} \neq 0$.



This pattern is the checkerboard pattern since $\mathbf{x}^T M \mathbf{x} = 0$, but $\mathbf{x}^T \mathbf{x} \neq 0$.

FIGURE 5

However, if the computer finds that $\mathbf{x}^T M\mathbf{x} = 0$, then $\mathbf{x}$ could be either $\mathbf{v}$ or $\mathbf{w}$. To distinguish between the two, the computer can calculate the product $\mathbf{x}^T\mathbf{x}$, for $\mathbf{x}^T\mathbf{x}$ is zero if and only if $\mathbf{x}$ is $\mathbf{w}$.[2] Thus, to conclude that $\mathbf{x}$ is equal to $\mathbf{v}$, the computer must have $\mathbf{x}^T M\mathbf{x} = 0$ and $\mathbf{x}^T\mathbf{x} \neq 0$. ■

Example 5 of Section 6.3 illustrates one way to choose a matrix M so that matrix multiplication and transposition can be used to identify a particular pattern of colored squares.

Another important aspect of AI starts even before the data is fed to the machine. In Section 1.9, it is illustrated how matrix multiplication can be used to move vectors around in space. In the next example, matrix multiplication is used to *scrub* data and prepare it for processing.

EXAMPLE 10 The dates of ground crew accidents for January and February of 2020 are listed in the columns of matrix T for Toronto Pearson Airport and matrix C for Chicago O'Hare Airport:

$$\text{Toronto: } T = \begin{bmatrix} 1 & 12 & 14 & 15 & 21 & 22 & 23 & 1 & 2 & 3 & 12 & 15 & 17 & 19 & 26 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \end{bmatrix}$$

$$\text{Chicago: } C = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 \\ 1 & 11 & 22 & 23 & 24 & 1 & 2 & 5 & 20 & 21 \end{bmatrix}$$

Clearly the data is listed differently in the two matrices. Canada and the United states have different traditions for whether the month or day comes first when writing a date. For matrix T, the day is listed in the first row and the month is listed in the second row. For matrix C, the month is listed in the first row and the day is listed in the second row. In order to use this data, the first and second rows need to be swapped in one of the matrices. Reviewing the effects of matrix multiplication in Table 1 of Section 1.9, notice that the matrix $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ switches the x_1 and x_2 coordinates of any vector $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ it is applied to and indeed

$$AT = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\ 1 & 12 & 14 & 15 & 21 & 22 & 23 & 1 & 2 & 3 & 12 & 15 & 17 & 19 & 26 \end{bmatrix}$$

has the data listed in the same order as it is listed in matrix C. The matrices AT and C can now be fed into the same machine. ■

In Exercises 51 and 52 you will be asked to *scrub* further data for this project.[3]

[2] To see why $\mathbf{x}^T\mathbf{x}$ is zero if and only if $\mathbf{x}$ is $\mathbf{w}$, let $\mathbf{x}^T = [x_1\ x_2\ x_3\ x_4]$. Then $\mathbf{x}^T\mathbf{x} = x_1^2 + x_2^2 + x_3^2 + x_4^2$ and this sum is zero if and only if the coordinates of $\mathbf{x}$ are all zero. That is, if and only if $\mathbf{x} = \mathbf{w}$.

[3] Although the data in this example and the corresponding exercises are fictitious, Data Analytics students at Washington State University identified scrubbing the data they received as an important first step in their actual analysis of ground crew accidents at three major airports in the United States.

Numerical Notes

1. The fastest way to obtain AB on a computer depends on the way in which the computer stores matrices in its memory. The standard high-performance algorithms, such as in LAPACK, calculate AB by columns, as in our definition of the product. (A version of LAPACK written in C++ calculates AB by rows.)
2. The definition of AB lends itself well to parallel processing on a computer. The columns of B are assigned individually or in groups to different processors, which independently and hence simultaneously compute the corresponding columns of AB.

Practice Problems

1. Since vectors in $\mathbb{R}^n$ may be regarded as $n \times 1$ matrices, the properties of transposes in Theorem 3 apply to vectors, too. Let

$$A = \begin{bmatrix} 1 & -3 \\ -2 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

Compute $(A\mathbf{x})^T$, $\mathbf{x}^T A^T$, $\mathbf{x}\mathbf{x}^T$, and $\mathbf{x}^T\mathbf{x}$. Is $A^T\mathbf{x}^T$ defined?

2. Let A be a 4×4 matrix and let $\mathbf{x}$ be a vector in $\mathbb{R}^4$. What is the fastest way to compute $A^2\mathbf{x}$? Count the multiplications.
3. Suppose A is an $m \times n$ matrix, all of whose rows are identical. Suppose B is an $n \times p$ matrix, all of whose columns are identical. What can be said about the entries in AB?

2.1 Exercises

In Exercises 1 and 2, compute each matrix sum or product if it is defined. If an expression is undefined, explain why. Let

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

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

1. $-2A$, $B - 2A$, AC, CD
2. $A + 2B$, $3C - E$, CB, EB

In the rest of this exercise set and in those to follow, you should assume that each matrix expression is defined. That is, the sizes of the matrices (and vectors) involved "match" appropriately.

3. Let $A = \begin{bmatrix} 4 & -1 \\ 5 & -2 \end{bmatrix}$. Compute $3I_2 - A$ and $(3I_2)A$.
4. Compute $A - 5I_3$ and $(5I_3)A$, when

$$A = \begin{bmatrix} 9 & -1 & 3 \\ -8 & 7 & -3 \\ -4 & 1 & 8 \end{bmatrix}.$$

In Exercises 5 and 6, compute the product AB in two ways: (a) by the definition, where $A\mathbf{b}_1$ and $A\mathbf{b}_2$ are computed separately, and (b) by the row–column rule for computing AB.

5. $A = \begin{bmatrix} -1 & 2 \\ 5 & 4 \\ 2 & -3 \end{bmatrix}$, $B = \begin{bmatrix} 3 & -4 \\ -2 & 1 \end{bmatrix}$
6. $A = \begin{bmatrix} 4 & -2 \\ -3 & 0 \\ 3 & 5 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 3 \\ 4 & -1 \end{bmatrix}$
7. If a matrix A is 5×3 and the product AB is 5×7, what is the size of B?
8. How many rows does B have if BC is a 3×4 matrix?
9. Let $A = \begin{bmatrix} 3 & 4 \\ -2 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 5 & -6 \\ 3 & k \end{bmatrix}$. What value(s) of k, if any, will make $AB = BA$?
10. Let $A = \begin{bmatrix} 3 & -6 \\ -4 & 8 \end{bmatrix}$, $B = \begin{bmatrix} 8 & 6 \\ 5 & 7 \end{bmatrix}$, $C = \begin{bmatrix} 6 & -2 \\ 4 & 3 \end{bmatrix}$. Verify that $AB = AC$ and yet $B \neq C$.
11. Let $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 5 \end{bmatrix}$ and $D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$. Compute AD and DA. Explain how the columns or rows of A change when A is multiplied by D on the right or on the left. Find a 3×3 matrix B, not the identity matrix or the zero matrix, such that $AB = BA$.

12. Let $A = \begin{bmatrix} 2 & -8 \\ -1 & 4 \end{bmatrix}$. Construct a 2×2 matrix B such that AB is the zero matrix. Use two different nonzero columns for B.

13. Let $\mathbf{r}_1, \ldots, \mathbf{r}_p$ be vectors in $\mathbb{R}^n$, and let Q be an $m \times n$ matrix. Write the matrix $[\, Q\mathbf{r}_1 \cdots Q\mathbf{r}_p \,]$ as a *product* of two matrices (neither of which is an identity matrix).

14. Let U be the 3×2 cost matrix described in Example 6 of Section 1.8. The first column of U lists the costs per dollar of output for manufacturing product B, and the second column lists the costs per dollar of output for product C. (The costs are categorized as materials, labor, and overhead.) Let $\mathbf{q}_1$ be a vector in $\mathbb{R}^2$ that lists the output (measured in dollars) of products B and C manufactured during the first quarter of the year, and let $\mathbf{q}_2, \mathbf{q}_3$, and $\mathbf{q}_4$ be the analogous vectors that list the amounts of products B and C manufactured in the second, third, and fourth quarters, respectively. Give an economic description of the data in the matrix UQ, where $Q = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \mathbf{q}_3 \quad \mathbf{q}_4]$.

Exercises 15–24 concern arbitrary matrices A, B, and C for which the indicated sums and products are defined. Mark each statement True or False **(T/F)**. Justify each answer.

15. **(T/F)** If A and B are 2×2 with columns $\mathbf{a}_1, \mathbf{a}_2$, and $\mathbf{b}_1, \mathbf{b}_2$, respectively, then $AB = [\mathbf{a}_1\mathbf{b}_1 \quad \mathbf{a}_2\mathbf{b}_2]$.

16. **(T/F)** If A and B are 3×3 and $B = [\mathbf{b}_1 \quad \mathbf{b}_2 \quad \mathbf{b}_3]$, then $AB = [A\mathbf{b}_1 + A\mathbf{b}_2 + A\mathbf{b}_3]$.

17. **(T/F)** Each column of AB is a linear combination of the columns of B using weights from the corresponding column of A.

18. **(T/F)** The second row of AB is the second row of A multiplied on the right by B.

19. **(T/F)** $AB + AC = A(B + C)$

20. **(T/F)** $A^T + B^T = (A + B)^T$

21. **(T/F)** $(AB)C = (AC)B$

22. **(T/F)** $(AB)^T = A^T B^T$

23. **(T/F)** The transpose of a product of matrices equals the product of their transposes in the same order.

24. **(T/F)** The transpose of a sum of matrices equals the sum of their transposes.

25. If $A = \begin{bmatrix} 1 & -3 \\ -3 & 8 \end{bmatrix}$ and $AB = \begin{bmatrix} -1 & 3 & -2 \\ 1 & -7 & 3 \end{bmatrix}$, determine the first and second columns of B.

26. Suppose the first two columns, $\mathbf{b}_1$ and $\mathbf{b}_2$, of B are equal. What can you say about the columns of AB (if AB is defined)? Why?

27. Suppose the third column of B is the sum of the first two columns. What can you say about the third column of AB? Why?

28. Suppose the second column of B is all zeros. What can you say about the second column of AB?

29. Suppose the last column of AB is all zeros, but B itself has no column of zeros. What can you say about the columns of A?

30. Show that if the columns of B are linearly dependent, then so are the columns of AB.

31. Suppose $CA = I_n$ (the $n \times n$ identity matrix). Show that the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Explain why A cannot have more columns than rows.

32. Suppose $AD = I_m$ (the $m \times m$ identity matrix). Show that for any $\mathbf{b}$ in $\mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a solution. [*Hint:* Think about the equation $AD\mathbf{b} = \mathbf{b}$.] Explain why A cannot have more rows than columns.

33. Suppose A is an $m \times n$ matrix and there exist $n \times m$ matrices C and D such that $CA = I_n$ and $AD = I_m$. Prove that $m = n$ and $C = D$. [*Hint:* Think about the product CAD.]

34. Suppose A is a $3 \times n$ matrix whose columns span $\mathbb{R}^3$. Explain how to construct an $n \times 3$ matrix D such that $AD = I_3$.

In Exercises 35 and 36, view vectors in $\mathbb{R}^n$ as $n \times 1$ matrices. For $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, the matrix product $\mathbf{u}^T\mathbf{v}$ is a 1×1 matrix, called the **scalar product**, or **inner product**, of $\mathbf{u}$ and $\mathbf{v}$. It is usually written as a single real number without brackets. The matrix product $\mathbf{u}\mathbf{v}^T$ is an $n \times n$ matrix, called the **outer product** of $\mathbf{u}$ and $\mathbf{v}$. The products $\mathbf{u}^T\mathbf{v}$ and $\mathbf{u}\mathbf{v}^T$ will appear later in the text.

35. Let $\mathbf{u} = \begin{bmatrix} -2 \\ 3 \\ -4 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$. Compute $\mathbf{u}^T\mathbf{v}, \mathbf{v}^T\mathbf{u}, \mathbf{u}\mathbf{v}^T$, and $\mathbf{v}\mathbf{u}^T$.

36. If $\mathbf{u}$ and $\mathbf{v}$ are in $\mathbb{R}^n$, how are $\mathbf{u}^T\mathbf{v}$ and $\mathbf{v}^T\mathbf{u}$ related? How are $\mathbf{u}\mathbf{v}^T$ and $\mathbf{v}\mathbf{u}^T$ related?

37. Prove Theorem 2(b) and 2(c). Use the row–column rule. The (i, j)-entry in $A(B + C)$ can be written as

$$a_{i1}(b_{1j} + c_{1j}) + \cdots + a_{in}(b_{nj} + c_{nj}) \text{ or } \sum_{k=1}^{n} a_{ik}(b_{kj} + c_{kj})$$

38. Prove Theorem 2(d). [*Hint:* The (i, j)-entry in $(rA)B$ is $(ra_{i1})b_{1j} + \cdots + (ra_{in})b_{nj}$.]

39. Show that $I_m A = A$ when A is an $m \times n$ matrix. You can assume $I_m\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^m$.

40. Show that $AI_n = A$ when A is an $m \times n$ matrix. [*Hint:* Use the (column) definition of AI_n.]

41. Prove Theorem 3(d). [*Hint:* Consider the jth row of $(AB)^T$.]

42. Give a formula for $(AB\mathbf{x})^T$, where $\mathbf{x}$ is a vector and A and B are matrices of appropriate sizes.

T 43. Use a web search engine such as Google to find documentation for your matrix program, and write the commands that

will produce the following matrices (without keying in each entry of the matrix).

a. A 5×6 matrix of zeros

b. A 3×5 matrix of ones

c. The 6×6 identity matrix

d. A 5×5 diagonal matrix, with diagonal entries 3, 5, 7, 2, 4

A useful way to test new ideas in matrix algebra, or to make conjectures, is to make calculations with matrices selected at random. Checking a property for a few matrices does not prove that the property holds in general, but it makes the property more believable. Also, if the property is actually false, you may discover this when you make a few calculations.

T 44. Write the command(s) that will create a 6×4 matrix with random entries. In what range of numbers do the entries lie? Tell how to create a 3×3 matrix with random integer entries between -9 and 9. [*Hint:* If x is a random number such that $0 < x < 1$, then $-9.5 < 19(x - .5) < 9.5$.]

T 45. Construct a random 4×4 matrix A and test whether $(A + I)(A - I) = A^2 - I$. The best way to do this is to compute $(A + I)(A - I) - (A^2 - I)$ and verify that this difference is the zero matrix. Do this for three random matrices. Then test $(A + B)(A - B) = A^2 - B^2$ the same way for three pairs of random 4×4 matrices. Report your conclusions.

T 46. Use at least three pairs of random 4×4 matrices A and B to test the equalities $(A + B)^T = A^T + B^T$ and $(AB)^T = A^T B^T$. (See Exercise 45.) Report your conclusions. [*Note:* Most matrix programs use A' for A^T.]

T 47. Let

$$S = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Compute S^k for $k = 2, \ldots, 6$.

T 48. Describe in words what happens when you compute A^5, A^{10}, A^{20}, and A^{30} for

$$A = \begin{bmatrix} 1/6 & 1/2 & 1/3 \\ 1/2 & 1/4 & 1/4 \\ 1/3 & 1/4 & 5/12 \end{bmatrix}$$

T 49. The matrix M can detect a particular 2×2 colored pattern like in Example 9. Create a nonzero 4×1 vector $\mathbf{x}$ by choosing each entry to be a zero or one. Test to see if $\mathbf{x}$ corresponds to the right pattern by calculating $\mathbf{x}^T M\mathbf{x}$. If $\mathbf{x}^T M\mathbf{x} = 0$, then $\mathbf{x}$ is the pattern identified by M. If $\mathbf{x}^T M\mathbf{x} \neq 0$, try a different nonzero vector of zeros and ones. You may want to be systematic in the way that you choose each $\mathbf{x}$ in order to avoid testing the same vector twice. You are using "guess and check" to determine which pattern of 2×2 colored squares the matrix M detects.

$$M = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

T 50. Repeat Exercise 49 with the matrix

$$M = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ -1 & -1 & 0 & 2 \end{bmatrix}$$

T 51. Use the matrix $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ to switch the first and second rows of the matrix M containing dates of accidents at the Montreal Trudeau Airport.

Montreal:

$$M = \begin{bmatrix} 2 & 3 & 16 & 24 & 25 & 26 & 6 & 7 & 19 & 26 \\ 1 & 1 & 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 \end{bmatrix}$$

This data in matrix M has been scrubbed in matrix AM and can be fed into the same machine as the other data from Example 10.

T 52. Use the matrix $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ to remove the last row from the matrix N containing dates of accidents at the New York JFK Airport.

New York:

$$N = \begin{bmatrix} 1 & 1 & 1 & 1 & 2 & 2 & 2 \\ 1 & 12 & 21 & 22 & 3 & 20 & 21 \\ 2020 & 2020 & 2020 & 2020 & 2020 & 2020 & 2020 \end{bmatrix}$$

The data in matrix N has been scrubbed in matrix BN and can be fed into the same machine as the other data from Example 10.

Solutions to Practice Problems

1. $A\mathbf{x} = \begin{bmatrix} 1 & -3 \\ -2 & 4 \end{bmatrix}\begin{bmatrix} 5 \\ 3 \end{bmatrix} = \begin{bmatrix} -4 \\ 2 \end{bmatrix}$. So $(A\mathbf{x})^T = \begin{bmatrix} -4 & 2 \end{bmatrix}$. Also,

$$\mathbf{x}^T A^T = \begin{bmatrix} 5 & 3 \end{bmatrix}\begin{bmatrix} 1 & -2 \\ -3 & 4 \end{bmatrix} = \begin{bmatrix} -4 & 2 \end{bmatrix}.$$

The quantities $(A\mathbf{x})^T$ and $\mathbf{x}^T A^T$ are equal, by Theorem 3(d). Next,

$$\mathbf{x}\mathbf{x}^T = \begin{bmatrix} 5 \\ 3 \end{bmatrix}\begin{bmatrix} 5 & 3 \end{bmatrix} = \begin{bmatrix} 25 & 15 \\ 15 & 9 \end{bmatrix}$$

$$\mathbf{x}^T\mathbf{x} = \begin{bmatrix} 5 & 3 \end{bmatrix}\begin{bmatrix} 5 \\ 3 \end{bmatrix} = [\,25 + 9\,] = 34$$

A 1×1 matrix such as $\mathbf{x}^T\mathbf{x}$ is usually written without the brackets. Finally, $A^T\mathbf{x}^T$ is not defined, because $\mathbf{x}^T$ does not have two rows to match the two columns of A^T.

2. The fastest way to compute $A^2\mathbf{x}$ is to compute $A(A\mathbf{x})$. The product $A\mathbf{x}$ requires 16 multiplications, 4 for each entry, and $A(A\mathbf{x})$ requires 16 more. In contrast, the product A^2 requires 64 multiplications, 4 for each of the 16 entries in A^2. After that, $A^2\mathbf{x}$ takes 16 more multiplications, for a total of 80.

3. First observe that by the definition of matrix multiplication,

$$AB = [A\mathbf{b}_1 \quad A\mathbf{b}_2 \quad \cdots \quad A\mathbf{b}_n] = [A\mathbf{b}_1 \quad A\mathbf{b}_1 \quad \cdots \quad A\mathbf{b}_1],$$

so the columns of AB are identical. Next, recall that $\text{row}_i(AB) = \text{row}_i(A) \cdot B$. Since all the rows of A are identical, all the rows of AB are identical. Putting this information about the rows and columns together, it follows that all the entries in AB are the same.

2.2 The Inverse of a Matrix

Matrix algebra provides tools for manipulating matrix equations and creating various useful formulas in ways similar to doing ordinary algebra with real numbers. This section investigates the matrix analogue of the reciprocal, or multiplicative inverse, of a nonzero number.

Recall that the multiplicative inverse of a number such as 5 is 1/5 or 5^{-1}. This inverse satisfies the equations

$$5^{-1}(5) = 1 \quad \text{and} \quad 5(5^{-1}) = 1$$

The matrix generalization requires *both* equations and avoids the slanted-line notation (for division) because matrix multiplication is not commutative. Furthermore, a full generalization is possible only if the matrices involved are square.[1]

An $n \times n$ matrix A is said to be **invertible** if there is an $n \times n$ matrix C such that

$$CA = I \quad \text{and} \quad AC = I$$

where $I = I_n$, the $n \times n$ identity matrix. In this case, C is an **inverse** of A. In fact, C is uniquely determined by A, because if B were another inverse of A, then $B = BI = B(AC) = (BA)C = IC = C$. This unique inverse is denoted by A^{-1}, so that

$$A^{-1}A = I \quad \text{and} \quad AA^{-1} = I$$

A matrix that is *not* invertible is sometimes called a **singular matrix**, and an invertible matrix is called a **nonsingular matrix**.

[1] One could say that an $m \times n$ matrix A is invertible if there exist $n \times m$ matrices C and D such that $CA = I_n$ and $AD = I_m$. However, these equations imply that A is square and $C = D$. Thus, A is invertible as defined above. See Exercises 31–33 in Section 2.1.

EXAMPLE 1 If $A = \begin{bmatrix} 2 & 5 \\ -3 & -7 \end{bmatrix}$ and $C = \begin{bmatrix} -7 & -5 \\ 3 & 2 \end{bmatrix}$, then

$$AC = \begin{bmatrix} 2 & 5 \\ -3 & -7 \end{bmatrix}\begin{bmatrix} -7 & -5 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and}$$

$$CA = \begin{bmatrix} -7 & -5 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} 2 & 5 \\ -3 & -7 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Thus $C = A^{-1}$. ■

Here is a simple formula for the inverse of a 2×2 matrix, along with a test to tell if the inverse exists.

THEOREM 4

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. If $ad - bc \neq 0$, then A is invertible and

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

If $ad - bc = 0$, then A is not invertible.

The simple proof of Theorem 4 is outlined in Exercises 35 and 36. The quantity $ad - bc$ is called the **determinant** of A, and we write

$$\det A = ad - bc$$

Theorem 4 says that a 2×2 matrix A is invertible if and only if $\det A \neq 0$.

EXAMPLE 2 Find the inverse of $A = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}$.

SOLUTION Since $\det A = 3(6) - 4(5) = -2 \neq 0$, A is invertible, and

$$A^{-1} = \frac{1}{-2}\begin{bmatrix} 6 & -4 \\ -5 & 3 \end{bmatrix} = \begin{bmatrix} 6/(-2) & -4/(-2) \\ -5/(-2) & 3/(-2) \end{bmatrix} = \begin{bmatrix} -3 & 2 \\ 5/2 & -3/2 \end{bmatrix}$$ ■

Invertible matrices are indispensable in linear algebra—mainly for algebraic calculations and formula derivations, as in the next theorem. There are also occasions when an inverse matrix provides insight into a mathematical model of a real-life situation, as in Example 3.

THEOREM 5

If A is an invertible $n \times n$ matrix, then for each $\mathbf{b}$ in $\mathbb{R}^n$, the equation $A\mathbf{x} = \mathbf{b}$ has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.

PROOF Take any $\mathbf{b}$ in $\mathbb{R}^n$. A solution exists because if $A^{-1}\mathbf{b}$ is substituted for $\mathbf{x}$, then $A\mathbf{x} = A(A^{-1}\mathbf{b}) = (AA^{-1})\mathbf{b} = I\mathbf{b} = \mathbf{b}$. So $A^{-1}\mathbf{b}$ is a solution. To prove that the solution is unique, show that if $\mathbf{u}$ is any solution, then $\mathbf{u}$, in fact, must be $A^{-1}\mathbf{b}$. Indeed, if $A\mathbf{u} = \mathbf{b}$, we can multiply both sides by A^{-1} and obtain

$$A^{-1}A\mathbf{u} = A^{-1}\mathbf{b}, \quad I\mathbf{u} = A^{-1}\mathbf{b}, \quad \text{and} \quad \mathbf{u} = A^{-1}\mathbf{b}$$ ■

EXAMPLE 3 A horizontal elastic beam is supported at each end and is subjected to forces at points 1, 2, and 3, as shown in Figure 1. Let $\mathbf{f}$ in $\mathbb{R}^3$ list the forces at these points, and let $\mathbf{y}$ in $\mathbb{R}^3$ list the amounts of deflection (that is, movement) of the beam at the three points. Using Hooke's law from physics, it can be shown that

$$\mathbf{y} = D\mathbf{f}$$

where D is a *flexibility matrix*. Its inverse is called the *stiffness matrix*. Describe the physical significance of the columns of D and D^{-1}.

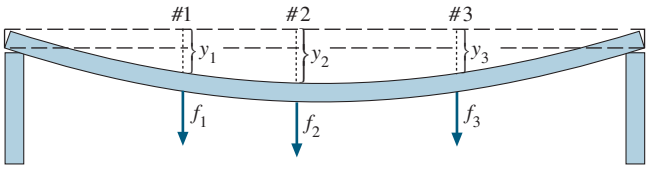


FIGURE 1 Deflection of an elastic beam.

SOLUTION Write $I_3 = [\,\mathbf{e}_1 \quad \mathbf{e}_2 \quad \mathbf{e}_3\,]$ and observe that

$$D = DI_3 = [\,D\mathbf{e}_1 \quad D\mathbf{e}_2 \quad D\mathbf{e}_3\,]$$

Interpret the vector $\mathbf{e}_1 = (1, 0, 0)$ as a unit force applied downward at point 1 on the beam (with zero force at the other two points). Then $D\mathbf{e}_1$, the first column of D, lists the beam deflections due to a unit force at point 1. Similar descriptions apply to the second and third columns of D.

To study the stiffness matrix D^{-1}, observe that the equation $\mathbf{f} = D^{-1}\mathbf{y}$ computes a force vector $\mathbf{f}$ when a deflection vector $\mathbf{y}$ is given. Write

$$D^{-1} = D^{-1}I_3 = [\,D^{-1}\mathbf{e}_1 \quad D^{-1}\mathbf{e}_2 \quad D^{-1}\mathbf{e}_3\,]$$

Now interpret $\mathbf{e}_1$ as a deflection vector. Then $D^{-1}\mathbf{e}_1$ lists the forces that create the deflection. That is, the first column of D^{-1} lists the forces that must be applied at the three points to produce a unit deflection at point 1 and zero deflections at the other points. Similarly, columns 2 and 3 of D^{-1} list the forces required to produce unit deflections at points 2 and 3, respectively. In each column, one or two of the forces must be negative (point upward) to produce a unit deflection at the desired point and zero deflections at the other two points. If the flexibility is measured, for example, in inches of deflection per pound of load, then the stiffness matrix entries are given in pounds of load per inch of deflection. ■

The formula in Theorem 5 is seldom used to solve an equation $A\mathbf{x} = \mathbf{b}$ numerically because row reduction of $[\,A \quad \mathbf{b}\,]$ is nearly always faster. (Row reduction is usually more accurate, too, when computations involve rounding off numbers.) One possible exception is the 2×2 case. In this case, mental computations to solve $A\mathbf{x} = \mathbf{b}$ are sometimes easier using the formula for A^{-1}, as in the next example.

EXAMPLE 4 Use the inverse of the matrix A in Example 2 to solve the system

$$\begin{aligned} 3x_1 + 4x_2 &= 3 \\ 5x_1 + 6x_2 &= 7 \end{aligned}$$

SOLUTION This system is equivalent to $A\mathbf{x} = \mathbf{b}$, so

$$\mathbf{x} = A^{-1}\mathbf{b} = \begin{bmatrix} -3 & 2 \\ 5/2 & -3/2 \end{bmatrix}\begin{bmatrix} 3 \\ 7 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$$

■

The next theorem provides three useful facts about invertible matrices.

THEOREM 6

a. If A is an invertible matrix, then A^{-1} is invertible and

$$(A^{-1})^{-1} = A$$

b. If A and B are $n \times n$ invertible matrices, then so is AB, and the inverse of AB is the product of the inverses of A and B in the reverse order. That is,

$$(AB)^{-1} = B^{-1}A^{-1}$$

c. If A is an invertible matrix, then so is A^T, and the inverse of A^T is the transpose of A^{-1}. That is,

$$(A^T)^{-1} = (A^{-1})^T$$

PROOF To verify statement (a), find a matrix C such that

$$A^{-1}C = I \quad \text{and} \quad CA^{-1} = I$$

In fact, these equations are satisfied with A in place of C. Hence A^{-1} is invertible, and A is its inverse. Next, to prove statement (b), compute:

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$$

A similar calculation shows that $(B^{-1}A^{-1})(AB) = I$. For statement (c), use Theorem 3(d), read from right to left, $(A^{-1})^T A^T = (AA^{-1})^T = I^T = I$. Similarly, $A^T(A^{-1})^T = I^T = I$. Hence A^T is invertible, and its inverse is $(A^{-1})^T$. ■

Remark: Part (b) illustrates the important role that definitions play in proofs. The theorem claims that $B^{-1}A^{-1}$ is the inverse of AB. The proof establishes this by showing that $B^{-1}A^{-1}$ satisfies the definition of what it means to be the inverse of AB. Now, the inverse of AB is a matrix that when multiplied on the left (or right) by AB, the product is the identity matrix I. So the proof consists of showing that $B^{-1}A^{-1}$ has this property.

The following generalization of Theorem 6(b) is needed later.

The product of $n \times n$ invertible matrices is invertible, and the inverse is the product of their inverses in the reverse order.

There is an important connection between invertible matrices and row operations that leads to a method for computing inverses. As we shall see, an invertible matrix A is row equivalent to an identity matrix, and we can find A^{-1} by *watching the row reduction of A to I.*

Elementary Matrices

An **elementary matrix** is one that is obtained by performing a single elementary row operation on an identity matrix. The next example illustrates the three kinds of elementary matrices.

EXAMPLE 5 Let

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix},$$

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

Compute E_1A, E_2A, and E_3A, and describe how these products can be obtained by elementary row operations on A.

SOLUTION Verify that

$$E_1A = \begin{bmatrix} a & b & c \\ d & e & f \\ g - 4a & h - 4b & i - 4c \end{bmatrix}, \quad E_2A = \begin{bmatrix} d & e & f \\ a & b & c \\ g & h & i \end{bmatrix},$$

$$E_3A = \begin{bmatrix} a & b & c \\ d & e & f \\ 5g & 5h & 5i \end{bmatrix}.$$

Addition of -4 times row 1 of A to row 3 produces E_1A. (This is a row replacement operation.) An interchange of rows 1 and 2 of A produces E_2A, and multiplication of row 3 of A by 5 produces E_3A. ■

Left-multiplication (that is, multiplication on the left) by E_1 in Example 5 has the same effect on any $3 \times n$ matrix. It adds -4 times row 1 to row 3. In particular, since $E_1 \cdot I = E_1$, we see that E_1 *itself* is produced by this same row operation on the identity. Thus Example 5 illustrates the following general fact about elementary matrices. See Exercises 37 and 38.

If an elementary row operation is performed on an $m \times n$ matrix A, the resulting matrix can be written as EA, where the $m \times m$ matrix E is created by performing the same row operation on I_m.

Since row operations are reversible, as shown in Section 1.1, elementary matrices are invertible, for if E is produced by a row operation on I, then there is another row operation of the same type that changes E back into I. Hence there is an elementary matrix F such that $FE = I$. Since E and F correspond to reverse operations, $EF = I$, too.

Each elementary matrix E is invertible. The inverse of E is the elementary matrix of the same type that transforms E back into I.

EXAMPLE 6 Find the inverse of $E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix}$.

SOLUTION To transform E_1 into I, add $+4$ times row 1 to row 3. The elementary matrix that does this is

$$E_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ +4 & 0 & 1 \end{bmatrix}$$

■

The following theorem provides the best way to "visualize" an invertible matrix, and the theorem leads immediately to a method for finding the inverse of a matrix.

THEOREM 7

An $n \times n$ matrix A is invertible if and only if A is row equivalent to I_n, and in this case, any sequence of elementary row operations that reduces A to I_n also transforms I_n into A^{-1}.

Remark: The comment on the proof of Theorem 11 in Chapter 1 noted that "P if and only if Q" is equivalent to two statements: (1) "If P then Q" and (2) "If Q then P." The second statement is called the *converse* of the first and explains the use of the word *conversely* in the second paragraph of this proof.

PROOF Suppose that A is invertible. Then, since the equation $A\mathbf{x} = \mathbf{b}$ has a solution for each $\mathbf{b}$ (Theorem 5), A has a pivot position in every row (Theorem 4 in Section 1.4). Because A is square, the n pivot positions must be on the diagonal, which implies that the reduced echelon form of A is I_n. That is, $A \sim I_n$.

Now suppose, conversely, that $A \sim I_n$. Then, since each step of the row reduction of A corresponds to left-multiplication by an elementary matrix, there exist elementary matrices $E_1, \ldots, E_p$ such that

$$A \sim E_1 A \sim E_2(E_1 A) \sim \cdots \sim E_p(E_{p-1} \cdots E_1 A) = I_n$$

That is,

$$E_p \cdots E_1 A = I_n \tag{1}$$

Since the product $E_p \cdots E_1$ of invertible matrices is invertible, (1) leads to

$$\begin{aligned}(E_p \cdots E_1)^{-1}(E_p \cdots E_1)A &= (E_p \cdots E_1)^{-1} I_n \\ A &= (E_p \cdots E_1)^{-1}\end{aligned}$$

Thus A is invertible, as it is the inverse of an invertible matrix (Theorem 6). Also,

$$A^{-1} = [\,(E_p \cdots E_1)^{-1}\,]^{-1} = E_p \cdots E_1$$

Then $A^{-1} = E_p \cdots E_1 I_n$, which says that A^{-1} results from applying $E_1, \ldots, E_p$ successively to I_n. This is the same sequence in (1) that reduced A to I_n. ■

An Algorithm for Finding A^{-1}

If we place A and I side by side to form an augmented matrix $[\,A \quad I\,]$, then row operations on this matrix produce identical operations on A and on I. By Theorem 7, either there are row operations that transform A to I_n and I_n to A^{-1} or else A is not invertible.

ALGORITHM FOR FINDING A^{-1}

Row reduce the augmented matrix $[\,A \quad I\,]$. If A is row equivalent to I, then $[\,A \quad I\,]$ is row equivalent to $[\,I \quad A^{-1}\,]$. Otherwise, A does not have an inverse.

EXAMPLE 7 Find the inverse of the matrix $A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix}$, if it exists.

SOLUTION

$$[\,A \quad I\,] = \begin{bmatrix} 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 0 & 3 & 0 & 1 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 4 & -3 & 8 & 0 & 0 & 1 \end{bmatrix}$$

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

$$\sim \begin{bmatrix} 1 & 0 & 3 & 0 & 1 & 0 \\ 0 & 1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 3/2 & -2 & 1/2 \end{bmatrix}$$

$$\sim \begin{bmatrix} 1 & 0 & 0 & -9/2 & 7 & -3/2 \\ 0 & 1 & 0 & -2 & 4 & -1 \\ 0 & 0 & 1 & 3/2 & -2 & 1/2 \end{bmatrix}$$

Theorem 7 shows, since $A \sim I$, that A is invertible, and

$$A^{-1} = \begin{bmatrix} -9/2 & 7 & -3/2 \\ -2 & 4 & -1 \\ 3/2 & -2 & 1/2 \end{bmatrix}$$

■

Reasonable Answers

Once you have found a candidate for the inverse of a matrix, you can check that your answer is correct by finding the product of A with A^{-1}. For the inverse found for matrix A in Example 7, notice

$$AA^{-1} = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 0 & 3 \\ 4 & -3 & 8 \end{bmatrix} \begin{bmatrix} -9/2 & 7 & -3/2 \\ -2 & 4 & -1 \\ 3/2 & -2 & 1/2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

confirming that answer is correct. It is not necessary to check that $A^{-1}A = I$ since A is invertible.

Another View of Matrix Inversion

Denote the columns of I_n by $\mathbf{e}_1, \ldots, \mathbf{e}_n$. Then row reduction of $[\,A \quad I\,]$ to $[\,I \quad A^{-1}\,]$ can be viewed as the simultaneous solution of the n systems

$$A\mathbf{x} = \mathbf{e}_1, \quad A\mathbf{x} = \mathbf{e}_2, \quad \ldots, \quad A\mathbf{x} = \mathbf{e}_n \tag{2}$$

where the "augmented columns" of these systems have all been placed next to A to form $[\,A \quad \mathbf{e}_1 \quad \mathbf{e}_2 \quad \cdots \quad \mathbf{e}_n\,] = [\,A \quad I\,]$. The equation $AA^{-1} = I$ and the definition of matrix multiplication show that the columns of A^{-1} are precisely the solutions of the systems in (2). This observation is useful because some applied problems may require finding only one or two columns of A^{-1}. In this case, only the corresponding systems in (2) need to be solved.

Numerical Note

In practical work, A^{-1} is seldom computed, unless the entries of A^{-1} are needed. Computing both A^{-1} and $A^{-1}\mathbf{b}$ takes about three times as many arithmetic operations as solving $A\mathbf{x} = \mathbf{b}$ by row reduction, and row reduction may be more accurate.

Practice Problems

1. Use determinants to determine which of the following matrices are invertible.

a. $\begin{bmatrix} 3 & -9 \\ 2 & 6 \end{bmatrix}$ b. $\begin{bmatrix} 4 & -9 \\ 0 & 5 \end{bmatrix}$ c. $\begin{bmatrix} 6 & -9 \\ -4 & 6 \end{bmatrix}$

2. Find the inverse of the matrix $A = \begin{bmatrix} 1 & -2 & -1 \\ -1 & 5 & 6 \\ 5 & -4 & 5 \end{bmatrix}$, if it exists.

3. If A is an invertible matrix, prove that $5A$ is an invertible matrix.

2.2 Exercises

Find the inverses of the matrices in Exercises 1–4.

1. $\begin{bmatrix} 8 & 3 \\ 5 & 2 \end{bmatrix}$ **2.** $\begin{bmatrix} 5 & 4 \\ 9 & 7 \end{bmatrix}$

3. $\begin{bmatrix} 8 & 3 \\ -7 & -3 \end{bmatrix}$ **4.** $\begin{bmatrix} 3 & -2 \\ 7 & -4 \end{bmatrix}$

5. Verify that the inverse you found in Exercise 1 is correct.

6. Verify that the inverse you found in Exercise 2 is correct.

7. Use the inverse found in Exercise 1 to solve the system

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

8. Use the inverse found in Exercise 2 to solve the system

$$\begin{aligned} 5x_1 + 4x_2 &= -3 \\ 9x_1 + 7x_2 &= -5 \end{aligned}$$

9. Let $A = \begin{bmatrix} 1 & 2 \\ 5 & 12 \end{bmatrix}$, $\mathbf{b}_1 = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$, $\mathbf{b}_2 = \begin{bmatrix} 1 \\ -5 \end{bmatrix}$, $\mathbf{b}_3 = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$, and $\mathbf{b}_4 = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$.

a. Find A^{-1}, and use it to solve the four equations $A\mathbf{x} = \mathbf{b}_1$, $A\mathbf{x} = \mathbf{b}_2$, $A\mathbf{x} = \mathbf{b}_3$, $A\mathbf{x} = \mathbf{b}_4$

b. The four equations in part (a) can be solved by the same set of row operations, since the coefficient matrix is the same in each case. Solve the four equations in part (a) by row reducing the augmented matrix $\begin{bmatrix} A & \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 & \mathbf{b}_4 \end{bmatrix}$

10. Use matrix algebra to show that if A is invertible and D satisfies $AD = I$, then $D = A^{-1}$.

In Exercises 11–20, mark each statement True or False **(T/F)**. Justify each answer.

11. **(T/F)** In order for a matrix B to be the inverse of A, both equations $AB = I$ and $BA = I$ must be true.

12. **(T/F)** A product of invertible $n \times n$ matrices is invertible, and the inverse of the product is the product of their inverses in the same order.

13. **(T/F)** If A and B are $n \times n$ and invertible, then $A^{-1}B^{-1}$ is the inverse of AB.

14. **(T/F)** If A is invertible, then the inverse of A^{-1} is A itself.

15. **(T/F)** If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $ab - cd \neq 0$, then A is invertible.

16. **(T/F)** If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $ad = bc$, then A is not invertible.

17. **(T/F)** If A is an invertible $n \times n$ matrix, then the equation $A\mathbf{x} = \mathbf{b}$ is consistent for *each* $\mathbf{b}$ in $\mathbb{R}^n$.

18. **(T/F)** If A can be row reduced to the identity matrix, then A must be invertible.

19. **(T/F)** Each elementary matrix is invertible.

20. **(T/F)** If A is invertible, then the elementary row operations that reduce A to the identity I_n also reduce A^{-1} to I_n.

21. Let A be an invertible $n \times n$ matrix, and let B be an $n \times p$ matrix. Show that the equation $AX = B$ has a unique solution $A^{-1}B$.

22. Let A be an invertible $n \times n$ matrix, and let B be an $n \times p$ matrix. Explain why $A^{-1}B$ can be computed by row reduction:

$$\text{If } [A \;\; B] \sim \cdots \sim [I \;\; X], \text{ then } X = A^{-1}B.$$

If A is larger than 2×2, then row reduction of $[A \;\; B]$ is much faster than computing both A^{-1} and $A^{-1}B$.

23. Suppose $AB = AC$, where B and C are $n \times p$ matrices and A is invertible. Show that $B = C$. Is this true, in general, when A is not invertible?

24. Suppose $(B - C)D = 0$, where B and C are $m \times n$ matrices and D is invertible. Show that $B = C$.

25. Suppose A, B, and C are invertible $n \times n$ matrices. Show that ABC is also invertible by producing a matrix D such that $(ABC)\,D = I$ and $D\,(ABC) = I$.

26. Suppose A and B are $n \times n$, B is invertible, and AB is invertible. Show that A is invertible. [*Hint:* Let $C = AB$, and solve this equation for A.]

27. Solve the equation $AB = BC$ for A, assuming that A, B, and C are square and B is invertible.

28. Suppose P is invertible and $A = PBP^{-1}$. Solve for B in terms of A.

29. If A, B, and C are $n \times n$ invertible matrices, does the equation $C^{-1}(A + X)B^{-1} = I_n$ have a solution, X? If so, find it.

30. Suppose A, B, and X are $n \times n$ matrices with A, X, and $A - AX$ invertible, and suppose

$$(A - AX)^{-1} = X^{-1}B \tag{3}$$

a. Explain why B is invertible.

b. Solve (3) for X. If you need to invert a matrix, explain why that matrix is invertible.

31. Explain why the columns of an $n \times n$ matrix A are linearly independent when A is invertible.

32. Explain why the columns of an $n \times n$ matrix A span $\mathbb{R}^n$ when A is invertible. [*Hint:* Review Theorem 4 in Section 1.4.]

33. Suppose A is $n \times n$ and the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Explain why A has n pivot columns and A is row equivalent to I_n. By Theorem 7, this shows that A must be invertible. (This exercise and Exercise 34 will be cited in Section 2.3.)

34. Suppose A is $n \times n$ and the equation $A\mathbf{x} = \mathbf{b}$ has a solution for each $\mathbf{b}$ in $\mathbb{R}^n$. Explain why A must be invertible. [*Hint:* Is A row equivalent to I_n?]

Exercises 35 and 36 prove Theorem 4 for $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

35. Show that if $ad - bc = 0$, then the equation $A\mathbf{x} = \mathbf{0}$ has more than one solution. Why does this imply that A is not invertible? [*Hint:* First, consider $a = b = 0$. Then, if a and b are not both zero, consider the vector $\mathbf{x} = \begin{bmatrix} -b \\ a \end{bmatrix}$.]

36. Show that if $ad - bc \neq 0$, the formula for A^{-1} works.

Exercises 37 and 38 prove special cases of the facts about elementary matrices stated in the box following Example 5. Here A is a 3×3 matrix and $I = I_3$. (A general proof would require slightly more notation.)

37. a. Use equation (1) from Section 2.1 to show that $\text{row}_i(A) = \text{row}_i(I) \cdot A$, for $i = 1, 2, 3$.

b. Show that if rows 1 and 2 of A are interchanged, then the result may be written as EA, where E is an elementary matrix formed by interchanging rows 1 and 2 of I.

c. Show that if row 3 of A is multiplied by 5, then the result may be written as EA, where E is formed by multiplying row 3 of I by 5.

38. Show that if row 3 of A is replaced by $\text{row}_3(A) - 4\text{row}_1(A)$, the result is EA, where E is formed from I by replacing $\text{row}_3(I)$ by $\text{row}_3(I) - 4\text{row}_1(I)$.

Find the inverses of the matrices in Exercises 39–42, if they exist. Use the algorithm introduced in this section.

39. $\begin{bmatrix} 1 & 2 \\ 4 & 7 \end{bmatrix}$

40. $\begin{bmatrix} 9 & 7 \\ 8 & 6 \end{bmatrix}$

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

42. $\begin{bmatrix} 1 & -2 & 1 \\ 4 & -7 & 3 \\ -2 & 6 & -4 \end{bmatrix}$

43. Use the algorithm from this section to find the inverses of

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

Let A be the corresponding $n \times n$ matrix, and let B be its inverse. Guess the form of B, and then prove that $AB = I$ and $BA = I$.

44. Repeat the strategy of Exercise 43 to guess the inverse of $A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 2 & 0 & & 0 \\ 1 & 2 & 3 & & 0 \\ \vdots & & & \ddots & \vdots \\ 1 & 2 & 3 & \cdots & n \end{bmatrix}$. Prove that your guess is correct.

45. Let $A = \begin{bmatrix} -2 & -7 & -9 \\ 2 & 5 & 6 \\ 1 & 3 & 4 \end{bmatrix}$. Find the third column of A^{-1} without computing the other columns.

T 46. Let $A = \begin{bmatrix} -25 & -9 & -27 \\ 546 & 180 & 537 \\ 154 & 50 & 149 \end{bmatrix}$. Find the second and third columns of A^{-1} without computing the first column.

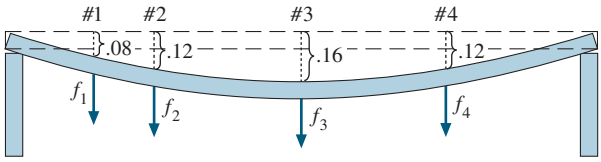
47. Let $A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 5 \end{bmatrix}$. Construct a 2×3 matrix C (by trial and error) using only 1, -1, and 0 as entries, such that $CA = I_2$. Compute AC and note that $AC \neq I_3$.

48. Let $A = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix}$. Construct a 4×2 matrix D using only 1 and 0 as entries, such that $AD = I_2$. Is it possible that $CA = I_4$ for some 4×2 matrix C? Why or why not?

49. Let $D = \begin{bmatrix} .005 & .002 & .001 \\ .002 & .004 & .002 \\ .001 & .002 & .005 \end{bmatrix}$ be a flexibility matrix, with flexibility measured in inches per pound. Suppose that forces of 30, 50, and 20 lb are applied at points 1, 2, and 3, respectively, in Figure 1 of Example 3. Find the corresponding deflections.

T 50. Compute the stiffness matrix D^{-1} for D in Exercise 49. List the forces needed to produce a deflection of .04 in. at point 3, with zero deflections at the other points.

T 51. Let $D = \begin{bmatrix} .0040 & .0030 & .0010 & .0005 \\ .0030 & .0050 & .0030 & .0010 \\ .0010 & .0030 & .0050 & .0030 \\ .0005 & .0010 & .0030 & .0040 \end{bmatrix}$ be a flexibility matrix for an elastic beam with four points at which force is applied. Units are centimeters per newton of force. Measurements at the four points show deflections of .08, .12, .16, and .12 cm. Determine the forces at the four points.



Deflection of elastic beam in Exercises 51 and 52.

T 52. With D as in Exercise 51, determine the forces that produce a deflection of .24 cm at the second point on the beam, with zero deflections at the other three points. How is the answer related to the entries in D^{-1}? [*Hint:* First answer the question when the deflection is 1 cm at the second point.]

Solutions to Practice Problems

1. a. $\det \begin{bmatrix} 3 & -9 \\ 2 & 6 \end{bmatrix} = 3 \cdot 6 - (-9) \cdot 2 = 18 + 18 = 36$. The determinant is nonzero, so the matrix is invertible.

b. $\det \begin{bmatrix} 4 & -9 \\ 0 & 5 \end{bmatrix} = 4 \cdot 5 - (-9) \cdot 0 = 20 \neq 0$. The matrix is invertible.

c. $\det \begin{bmatrix} 6 & -9 \\ -4 & 6 \end{bmatrix} = 6 \cdot 6 - (-9)(-4) = 36 - 36 = 0$. The matrix is not invertible.

2.
$$[\,A \quad I\,] \sim \begin{bmatrix} 1 & -2 & -1 & 1 & 0 & 0 \\ -1 & 5 & 6 & 0 & 1 & 0 \\ 5 & -4 & 5 & 0 & 0 & 1 \end{bmatrix}$$
$$\sim \begin{bmatrix} 1 & -2 & -1 & 1 & 0 & 0 \\ 0 & 3 & 5 & 1 & 1 & 0 \\ 0 & 6 & 10 & -5 & 0 & 1 \end{bmatrix}$$
$$\sim \begin{bmatrix} 1 & -2 & -1 & 1 & 0 & 0 \\ 0 & 3 & 5 & 1 & 1 & 0 \\ 0 & 0 & 0 & -7 & -2 & 1 \end{bmatrix}$$

So $[\,A \quad I\,]$ is row equivalent to a matrix of the form $[\,B \quad D\,]$, where B is square and has a row of zeros. Further row operations will not transform B into I, so we stop. A does not have an inverse.

3. Since A is an invertible matrix, there exists a matrix C such that $AC = I = CA$. The goal is to find a matrix D so that $(5A)D = I = D(5A)$. Set $D = 1/5\,C$. Applying Theorem 2 from Section 2.1 establishes that $(5A)(1/5\,C) = (5)(1/5)(AC) = 1\,I = I$, and $(1/5\,C)(5A) = (1/5)(5)(CA) = 1\,I = I$. Thus $1/5\,C$ is indeed the inverse of A, proving that A is invertible.

3 Determinants



Introductory Example

WEIGHING DIAMONDS

How is the value of a diamond determined? Jewelers use the four cs: cut, clarity, color, and carats; a carat is a unit of mass equal to 0.2 grams. When a jeweler receives a supply of diamonds, it is vital that they be weighed accurately as part of determining their value. The difference of half a carat can have a large impact on a diamond's value.

When weighing small objects, such as diamonds or other gemstones, one strategy is to weigh the objects individually, but there are more accurate strategies that involve weighing the objects in groups and then deducing the individual weights from the results.

Suppose there are n small objects to be weighed, labeled $s_1, s_2, \cdots, s_n$. One method of determining the weight of each small object uses a two-pan balance. A *weighing* consists of placing some of the small objects in the left pan and the rest in the right pan. The balance records the difference between the weights in the pans.



The jeweler (or other individual weighing small light objects) plans her strategy in advance by creating a design matrix D with entries determined by the following scheme: If gemstone s_j is placed in the left pan during the ith weighing, then $d_{ij} = -1$ and if gemstone s_j is placed in the right pan during the ith weighing the $d_{ij} = 1$. Each row of the matrix D corresponds to a particular weighing. The jth column of D tells you where to put s_j at each weighing. Thus D is an $m \times n$ matrix, where m corresponds to the number of weighings and n corresponds to the number of objects. It has been shown that the accuracy of a weighing design is highest when a design matrix that maximizes the value of the *determinant* of $D^T D$ is chosen.

For example, consider the design matrix $D = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix}$ for weighing the gems s_1, s_2, s_3, s_4. For this design, the first weighing has all four gems in the right tray (the first row of D consists of all ones). For the second weighing, gem s_2 is in the left tray and the rest of the gems are in the right tray (the second row of D has a -1 in the second column). For the third weighing, gem s_3 is in the left tray and the rest of the gems are in the right tray (the third row of D has a -1 in the third column). In the last weighing, gem s_4 is in the left tray and the remaining gems are in the right tray (the fourth row of D has a -1 in the fourth column). The determinant of $D^T D$ is 64.

However, this is not the best design for using four weighings to determine the weight of four objects.

If $D = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & 1 & -1 \end{bmatrix}$, then the determinant of $D^TD = 256$, and hence this is a better design. Notice that the first weighing of this design is the same as the previous one, but then the remaining weighings each have two objects in each pan.

Calculating determinants of matrices and understanding their properties is the theme of this chapter. As you learn more about determinants, you may also come up with strategies for good and bad choices for a weighing design.

Another important use of the determinant is to calculate the area of a parallelogram or the volume of a parallelepiped. In Section 1.9, we saw that matrix multiplication can be used to change the shape of a box or other object. The determinant of the matrix used determines how much the area changes when it is multiplied by a matrix, just as a fish story can transform the size of the fish caught.

Indeed, the determinant has so many uses that a summary of the applications known in the early 1900s filled a four-volume treatise by Thomas Muir. With changes in emphasis and the greatly increased sizes of the matrices used in modem applications, many uses that were important then are no longer critical today. Nevertheless, the determinant still plays many important theoretical and practical roles.

Beyond introducing the determinant in Section 3.1, this chapter presents two important ideas. Section 3.2 derives an invertibility criterion for a square matrix that plays a pivotal role in Chapter 5. Section 3.3 shows how the determinant measures the amount by which a linear transformation changes the area of a figure. When applied locally, this technique answers the question of a map's expansion rate near the poles. This idea plays a critical role in multivariable calculus in the form of the Jacobian.

3.1 Introduction to Determinants

Recall from Section 2.2 that a 2×2 matrix is invertible if and only if its determinant is nonzero. To extend this useful fact to larger matrices, we need a definition for the determinant of an $n \times n$ matrix. We can discover the definition for the 3×3 case by watching what happens when an invertible 3×3 matrix A is row reduced.

Consider $A = [a_{ij}]$ with $a_{11} \neq 0$. If we multiply the second and third rows of A by a_{11} and then subtract appropriate multiples of the first row from the other two rows, we find that A is row equivalent to the following two matrices:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{11}a_{21} & a_{11}a_{22} & a_{11}a_{23} \\ a_{11}a_{31} & a_{11}a_{32} & a_{11}a_{33} \end{bmatrix} \sim \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{11}a_{22} - a_{12}a_{21} & a_{11}a_{23} - a_{13}a_{21} \\ 0 & a_{11}a_{32} - a_{12}a_{31} & a_{11}a_{33} - a_{13}a_{31} \end{bmatrix} \quad (1)$$

Since A is invertible, either the $(2,2)$-entry or the $(3,2)$-entry on the right in (1) is nonzero. Let us suppose that the $(2,2)$-entry is nonzero. (Otherwise, we can make a row interchange before proceeding.) Multiply row 3 by $a_{11}a_{22} - a_{12}a_{21}$, and then to the new row 3 add $-(a_{11}a_{32} - a_{12}a_{31})$ times row 2. This will show that

$$A \sim \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{11}a_{22} - a_{12}a_{21} & a_{11}a_{23} - a_{13}a_{21} \\ 0 & 0 & a_{11}\Delta \end{bmatrix}$$

where

$$\Delta = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} \quad (2)$$

Since A is invertible, Δ must be nonzero. The converse is true, too, as we will see in Section 3.2. We call Δ in (2) the **determinant** of the 3×3 matrix A.

Recall that the determinant of a 2×2 matrix, $A = [a_{ij}]$, is the number

$$\det A = a_{11}a_{22} - a_{12}a_{21}$$

For a 1×1 matrix—say, $A = [a_{11}]$—we define $\det A = a_{11}$. To generalize the definition of the determinant to larger matrices, we'll use 2×2 determinants to rewrite the 3×3 determinant Δ described above. Since the terms in Δ can be grouped as $(a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32}) - (a_{12}a_{21}a_{33} - a_{12}a_{23}a_{31}) + (a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31})$,

$$\Delta = a_{11} \det \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} - a_{12} \det \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix} + a_{13} \det \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$$

For brevity, write

$$\Delta = a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13} \quad (3)$$

where A_{11}, A_{12}, and A_{13} are obtained from A by deleting the first row and one of the three columns. For any square matrix A, let A_{ij} denote the submatrix formed by deleting the ith row and jth column of A. For instance, if

$$A = \begin{bmatrix} 1 & -2 & 5 & 0 \\ 2 & 0 & 4 & -1 \\ 3 & 1 & 0 & 7 \\ 0 & 4 & -2 & 0 \end{bmatrix}$$

then A_{32} is obtained by crossing out row 3 and column 2,

$$\begin{bmatrix} 1 & -2 & 5 & 0 \\ 2 & 0 & 4 & -1 \\ 3 & 1 & 0 & 7 \\ 0 & 4 & -2 & 0 \end{bmatrix}$$

so that

$$A_{32} = \begin{bmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{bmatrix}$$

We can now give a *recursive* definition of a determinant. When $n = 3$, $\det A$ is defined using determinants of the 2×2 submatrices A_{1j}, as in (3) above. When $n = 4$, $\det A$ uses determinants of the 3×3 submatrices A_{1j}. In general, an $n \times n$ determinant is defined by determinants of $(n-1) \times (n-1)$ submatrices.

DEFINITION

For $n \geq 2$, the **determinant** of an $n \times n$ matrix $A = [a_{ij}]$ is the sum of n terms of the form $\pm a_{1j} \det A_{1j}$, with plus and minus signs alternating, where the entries $a_{11}, a_{12}, \ldots, a_{1n}$ are from the first row of A. In symbols,

$$\begin{aligned} \det A &= a_{11} \det A_{11} - a_{12} \det A_{12} + \cdots + (-1)^{1+n} a_{1n} \det A_{1n} \\ &= \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det A_{1j} \end{aligned}$$

EXAMPLE 1 Compute the determinant of

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

SOLUTION Compute $\det A = a_{11} \det A_{11} - a_{12} \det A_{12} + a_{13} \det A_{13}$:

$$\begin{aligned} \det A &= 1 \det \begin{bmatrix} 4 & -1 \\ -2 & 0 \end{bmatrix} - 5 \det \begin{bmatrix} 2 & -1 \\ 0 & 0 \end{bmatrix} + 0 \det \begin{bmatrix} 2 & 4 \\ 0 & -2 \end{bmatrix} \\ &= 1(0-2) - 5(0-0) + 0(-4-0) = -2 \end{aligned}$$

■

Another common notation for the determinant of a matrix uses a pair of vertical lines in place of brackets. Thus the calculation in Example 1 can be written as

$$\det A = 1 \begin{vmatrix} 4 & -1 \\ -2 & 0 \end{vmatrix} - 5 \begin{vmatrix} 2 & -1 \\ 0 & 0 \end{vmatrix} + 0 \begin{vmatrix} 2 & 4 \\ 0 & -2 \end{vmatrix} = \cdots = -2$$

To state the next theorem, it is convenient to write the definition of $\det A$ in a slightly different form. Given $A = [a_{ij}]$, the **(i, j)-cofactor** of A is the number C_{ij} given by

$$C_{ij} = (-1)^{i+j} \det A_{ij} \tag{4}$$

Then

$$\det A = a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n}$$

This formula is called a **cofactor expansion across the first row** of A. We omit the proof of the following fundamental theorem to avoid a lengthy digression.

THEOREM 1

The determinant of an $n \times n$ matrix A can be computed by a cofactor expansion across any row or down any column. The expansion across the ith row using the cofactors in (4) is

$$\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}$$

The cofactor expansion down the jth column is

$$\det A = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}$$

The plus or minus sign in the (i, j)-cofactor depends on the position of a_{ij} in the matrix, regardless of the sign of a_{ij} itself. The factor $(-1)^{i+j}$ determines the following checkerboard pattern of signs:

$$\begin{bmatrix} + & - & + & \cdots \\ - & + & - & \\ + & - & + & \\ \vdots & & & \ddots \end{bmatrix}$$

EXAMPLE 2 Use a cofactor expansion across the third row to compute $\det A$, where

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

SOLUTION Compute

$$\begin{aligned}\det A &= a_{31}C_{31} + a_{32}C_{32} + a_{33}C_{33} \\ &= (-1)^{3+1}a_{31}\det A_{31} + (-1)^{3+2}a_{32}\det A_{32} + (-1)^{3+3}a_{33}\det A_{33} \\ &= 0\begin{vmatrix} 5 & 0 \\ 4 & -1 \end{vmatrix} - (-2)\begin{vmatrix} 1 & 0 \\ 2 & -1 \end{vmatrix} + 0\begin{vmatrix} 1 & 5 \\ 2 & 4 \end{vmatrix} \\ &= 0 + 2(-1) + 0 = -2\end{aligned}$$

■

Theorem 1 is helpful for computing the determinant of a matrix that contains many zeros. For example, if a row is mostly zeros, then the cofactor expansion across that row has many terms that are zero, and the cofactors in those terms need not be calculated. The same approach works with a column that contains many zeros.

EXAMPLE 3 Compute $\det A$, where

$$A = \begin{bmatrix} 3 & -7 & 8 & 9 & -6 \\ 0 & 2 & -5 & 7 & 3 \\ 0 & 0 & 1 & 5 & 0 \\ 0 & 0 & 2 & 4 & -1 \\ 0 & 0 & 0 & -2 & 0 \end{bmatrix}$$

SOLUTION The cofactor expansion down the first column of A has all terms equal to zero except the first. Thus

$$\det A = 3\begin{vmatrix} 2 & -5 & 7 & 3 \\ 0 & 1 & 5 & 0 \\ 0 & 2 & 4 & -1 \\ 0 & 0 & -2 & 0 \end{vmatrix} + 0\,C_{21} + 0\,C_{31} + 0\,C_{41} + 0\,C_{51}$$

Henceforth we will omit the zero terms in the cofactor expansion. Next, expand this 4×4 determinant down the first column to take advantage of the zeros there. We have

$$\det A = 3(2)\begin{vmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{vmatrix}$$

This 3×3 determinant was computed in Example 1 and found to equal -2. Hence $\det A = 3(2)(-2) = -12$. ■

The matrix in Example 3 was nearly triangular. The method in that example is easily adapted to prove the following theorem.

THEOREM 2

If A is a triangular matrix, then $\det A$ is the product of the entries on the main diagonal of A.

The strategy in Example 3 of looking for zeros works extremely well when an entire row or column consists of zeros. In such a case, the cofactor expansion along such a row or column is a sum of zeros! So the determinant is zero. Unfortunately, most cofactor expansions are not so quickly evaluated.

Reasonable Answers

How big can a determinant be? Let A be an $n \times n$ matrix. Notice that taking the determinant of A consists of adding and subtracting terms with n products each. If p is the product of the n largest elements in absolute value (the same number may be repeated if it occurs more than once as a matrix entry), then the determinant must be between $-np$ and np. For example, consider $A = \begin{bmatrix} 6 & 5 \\ -7 & 9 \end{bmatrix}$ and $B = \begin{bmatrix} 7 & 6 \\ 7 & -9 \end{bmatrix}$. The largest number in absolute value of each matrix is 9, and the second largest number is 7. In these two cases, $p = 7(9) = 63$ and $np = 126$. The determinant of each of these matrices should be a number between -126 and 126. Notice that $\det A = 6(9) - 5(-7) = 54 + 35 = 89$, $\det B = 7(-9) - 6(7) = -63 - 42 = -105$, illustrating that because the products are added and subtracted, any number in the range between -126 and 126 could turn out to be the determinant.

Next, consider $C = \begin{bmatrix} 7 & 9 \\ 7 & 9 \end{bmatrix}$ and $D = \begin{bmatrix} -9 & 9 \\ 9 & 9 \end{bmatrix}$. In matrices C and D, the number 9 appears twice and so should be selected twice. In this case, $p = 9(9) = 81$ and $np = 162$, so the determinants of C and D should be numbers between -162 and 162. Indeed, $\det C = (7)(9) - (7)(9) = 0$ and $\det D = (-9)(9) - (9)(9) = -162$. Notice that it is important to choose 9 twice as the two largest numbers in matrix D in order to get the correct bounds for the determinant of D.

Numerical Note

By today's standards, a 25×25 matrix is small. Yet it would be impossible to calculate a 25×25 determinant by cofactor expansion. In general, a cofactor expansion requires more than $n!$ multiplications, and 25! is approximately 1.55×10^{25}.

If a computer performs one trillion multiplications per second, it would have to run for almost 500,000 years to compute a 25×25 determinant by this method. Fortunately, there are faster methods, as we'll soon discover.

Exercises 19–38 explore important properties of determinants, mostly for the 2×2 case. The results from Exercises 33–36 will be used in the next section to derive the analogous properties for $n \times n$ matrices.

Practice Problem

Compute $\begin{vmatrix} 5 & -7 & 2 & 2 \\ 0 & 3 & 0 & -4 \\ -5 & -8 & 0 & 3 \\ 0 & 5 & 0 & -6 \end{vmatrix}$.

3.1 Exercises

Compute the determinants in Exercises 1–8 using a cofactor expansion across the first row. In Exercises 1–4, also compute the determinant by a cofactor expansion down the second column.

1. $\begin{vmatrix} 3 & 0 & 4 \\ 2 & 3 & 2 \\ 0 & 5 & -1 \end{vmatrix}$ **2.** $\begin{vmatrix} 0 & 4 & 1 \\ 5 & -3 & 0 \\ 2 & 4 & 1 \end{vmatrix}$

3. $\begin{vmatrix} 2 & -2 & 3 \\ 3 & 1 & 2 \\ 1 & 3 & -1 \end{vmatrix}$ **4.** $\begin{vmatrix} 1 & 2 & 4 \\ 3 & 1 & 1 \\ 2 & 4 & 2 \end{vmatrix}$

5. $\begin{vmatrix} 4 & 5 & -8 \\ 1 & 0 & 2 \\ 7 & 3 & 6 \end{vmatrix}$ **6.** $\begin{vmatrix} 6 & -3 & 2 \\ 0 & 5 & -5 \\ 3 & -7 & 8 \end{vmatrix}$

7. $\begin{vmatrix} 4 & 3 & 0 \\ 6 & 5 & 2 \\ 9 & 7 & 3 \end{vmatrix}$ **8.** $\begin{vmatrix} 4 & 1 & 2 \\ 4 & 0 & 3 \\ 3 & -2 & 5 \end{vmatrix}$

Compute the determinants in Exercises 9–14 by cofactor expansions. At each step, choose a row or column that involves the least amount of computation.

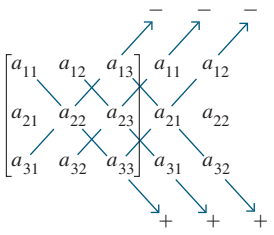
9. $\begin{vmatrix} 7 & 6 & 8 & 4 \\ 0 & 0 & 0 & 6 \\ 8 & 7 & 9 & 3 \\ 0 & 4 & 0 & 5 \end{vmatrix}$ **10.** $\begin{vmatrix} 1 & -2 & 4 & 2 \\ 0 & 0 & 3 & 0 \\ 2 & -4 & -3 & 5 \\ 2 & 0 & 3 & 5 \end{vmatrix}$

11. $\begin{vmatrix} 2 & -3 & 4 & 5 \\ 0 & 5 & 3 & -1 \\ 0 & 0 & -2 & 7 \\ 0 & 0 & 0 & 4 \end{vmatrix}$ **12.** $\begin{vmatrix} 3 & 0 & 0 & 0 \\ 7 & -2 & 0 & 0 \\ 2 & 6 & 3 & 0 \\ 3 & -8 & 4 & -3 \end{vmatrix}$

13. $\begin{vmatrix} 4 & 0 & -7 & 3 & -5 \\ 0 & 0 & 2 & 0 & 0 \\ 7 & 3 & -6 & 4 & -8 \\ 5 & 0 & 5 & 2 & -3 \\ 0 & 0 & 9 & -1 & 2 \end{vmatrix}$

14. $\begin{vmatrix} 6 & 0 & 2 & 4 & 0 \\ 9 & 0 & -4 & 1 & 0 \\ 8 & -5 & 6 & 7 & 1 \\ 2 & 0 & 0 & 0 & 0 \\ 4 & 2 & 3 & 2 & 0 \end{vmatrix}$

The expansion of a 3×3 determinant can be remembered by the following device. Write a second copy of the first two columns to the right of the matrix, and compute the determinant by multiplying entries on six diagonals:



Add the downward diagonal products and subtract the upward products. Use this method to compute the determinants in Exercises 15–18. ***Warning:*** *This trick does not generalize in any reasonable way to* 4×4 *or larger matrices.*

15. $\begin{vmatrix} 1 & 0 & 4 \\ 2 & 3 & 2 \\ 0 & 5 & -2 \end{vmatrix}$ **16.** $\begin{vmatrix} 6 & 5 & 0 \\ 4 & 3 & -2 \\ 2 & 0 & 1 \end{vmatrix}$

17. $\begin{vmatrix} 2 & -3 & 3 \\ 3 & 2 & 2 \\ 1 & 3 & -1 \end{vmatrix}$ **18.** $\begin{vmatrix} 1 & 4 & 5 \\ 3 & 4 & 3 \\ 3 & 3 & 4 \end{vmatrix}$

In Exercises 19–24, explore the effect of an elementary row operation on the determinant of a matrix. In each case, state the row operation and describe how it affects the determinant.

19. $\begin{bmatrix} a & b \\ c & d \end{bmatrix}, \begin{bmatrix} c & d \\ a & b \end{bmatrix}$

20. $\begin{bmatrix} a & b \\ c & d \end{bmatrix}, \begin{bmatrix} a & b \\ kc & kd \end{bmatrix}$

21. $\begin{bmatrix} 6 & 5 \\ 3 & 4 \end{bmatrix}, \begin{bmatrix} 6 & 5 \\ 3+6k & 4+5k \end{bmatrix}$

22. $\begin{bmatrix} a & b \\ c & d \end{bmatrix}, \begin{bmatrix} a+kc & b+kd \\ c & d \end{bmatrix}$

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

24. $\begin{bmatrix} a & b & c \\ 1 & 4 & 5 \\ 2 & 3 & 6 \end{bmatrix}, \begin{bmatrix} 2 & 3 & 6 \\ 1 & 4 & 5 \\ a & b & c \end{bmatrix}$

Compute the determinants of the elementary matrices given in Exercises 25–30. (See Section 2.2, Examples 5 and 6.)

25. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & k & 1 \end{bmatrix}$ **26.** $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

27. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ k & 0 & 1 \end{bmatrix}$ **28.** $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$

29. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & 1 \end{bmatrix}$ **30.** $\begin{bmatrix} k & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Use Exercises 25–30 to answer the questions in Exercises 31 and 32. Give reasons for your answers.

31. What is the determinant of an elementary row replacement matrix?

32. What is the determinant of an elementary scaling matrix with k on the diagonal?

In Exercises 33–36, verify that $\det EA = (\det E)(\det A)$, where E is the elementary matrix shown and $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

33. $\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$ **34.** $\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$

35. $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ **36.** $\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$

37. Let $A = \begin{bmatrix} 6 & 5 \\ 3 & 4 \end{bmatrix}$. Write $2A$. Is $\det 2A = 2 \det A$?

38. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and let k be a scalar. Find a formula that relates $\det kA$ to k and $\det A$.

In Exercises 39 through 42, A is an $n \times n$ matrix. Mark each statement True or False **(T/F)**. Justify each answer.

39. (T/F) An $n \times n$ determinant is defined by determinants of $(n-1) \times (n-1)$ submatrices.

40. (T/F) The (i, j)-cofactor of a matrix A is the matrix A_{ij} obtained by deleting from A its ith row and jth column.

41. (T/F) The cofactor expansion of $\det A$ down a column is equal to the cofactor expansion along a row.

42. (T/F) The determinant of a triangular matrix is the sum of the entries on the main diagonal.

43. Let $\mathbf{u} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Compute the area of the parallelogram determined by $\mathbf{u}$, $\mathbf{v}$, $\mathbf{u} + \mathbf{v}$, and $\mathbf{0}$, and compute the determinant of $[\,\mathbf{u} \quad \mathbf{v}\,]$. How do they compare? Replace the first entry of $\mathbf{v}$ by an arbitrary number x, and repeat the problem. Draw a picture and explain what you find.

44. Let $\mathbf{u} = \begin{bmatrix} a \\ b \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} c \\ 0 \end{bmatrix}$, where a, b, and c are positive (for simplicity). Compute the area of the parallelogram determined by $\mathbf{u}$, $\mathbf{v}$, $\mathbf{u} + \mathbf{v}$, and $\mathbf{0}$, and compute the determinants of the matrices $[\,\mathbf{u} \quad \mathbf{v}\,]$ and $[\,\mathbf{v} \quad \mathbf{u}\,]$. Draw a picture and explain what you find.

45. Let A be a 2×2 matrix all of whose entries are numbers that are greater than or equal to -10 and less than or equal to 10. Decide if each of the following is a reasonable answer for $\det A$.

a. 0

b. 202

c. -110

d. 555

46. Let A be a 3×3 matrix all of whose entries are numbers that are greater than or equal to -5 and less than or equal to 5. Decide if each of the following is a reasonable answer for $\det A$.

a. 300

b. -220

c. 1000

d. 10

T 47. Construct a random 4×4 matrix A with integer entries between -9 and 9. How is $\det A^{-1}$ related to $\det A$? Experiment with random $n \times n$ integer matrices for $n = 4$, 5, and 6, and make a conjecture. *Note:* In the unlikely event that you encounter a matrix with a zero determinant, reduce it to echelon form and discuss what you find.

T 48. Is it true that $\det AB = (\det A)(\det B)$? To find out, generate random 5×5 matrices A and B, and compute $\det AB - (\det A \det B)$. Repeat the calculations for three other pairs of $n \times n$ matrices, for various values of n. Report your results.

T 49. Is it true that $\det(A + B) = \det A + \det B$? Experiment with four pairs of random matrices as in Exercise 48, and make a conjecture.

T 50. Construct a random 4×4 matrix A with integer entries between -9 and 9, and compare $\det A$ with $\det A^T$, $\det(-A)$, $\det(2A)$, and $\det(10A)$. Repeat with two other random 4×4 integer matrices, and make conjectures about how these determinants are related. (Refer to Exercise 44 in Section 2.1.) Then check your conjectures with several random 5×5 and 6×6 integer matrices. Modify your conjectures, if necessary, and report your results.

T 51. Recall from the introductory section that the larger the determinant of $D^T D$, where D is the design matrix, the better will be the accuracy of the calculated weights for small light objects. Which of the following matrices corresponds to the best design for four weighings of four objects? Describe which of the objects s_1, s_2, s_3, and s_4 you would put in the left and right pans for each weighing corresponding to the best design matrix.

a. $D = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 \end{bmatrix}$

b. $D = \begin{bmatrix} -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix}$

c. $D = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 \end{bmatrix}$

T 52. Repeat Exercise 51 for the case of five weighings of four objects and the following design matrices.

a. $D = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix}$

b. $D = \begin{bmatrix} -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & 1 & 1 & -1 \end{bmatrix}$

c. $D = \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & 1 & -1 \\ -1 & -1 & -1 & -1 \end{bmatrix}$

Solution to Practice Problem

Take advantage of the zeros. Begin with a cofactor expansion down the third column to obtain a 3×3 matrix, which may be evaluated by an expansion down its first column.

$$\begin{vmatrix} 5 & -7 & 2 & 2 \\ 0 & 3 & 0 & -4 \\ -5 & -8 & 0 & 3 \\ 0 & 5 & 0 & -6 \end{vmatrix} = (-1)^{1+3}(2)\begin{vmatrix} 0 & 3 & -4 \\ -5 & -8 & 3 \\ 0 & 5 & -6 \end{vmatrix}$$

$$= 2\,(-1)^{2+1}(-5)\begin{vmatrix} 3 & -4 \\ 5 & -6 \end{vmatrix} = 20$$

The $(-1)^{2+1}$ in the next-to-last calculation came from the $(2, 1)$-position of the -5 in the 3×3 determinant.

3.2 Properties of Determinants

The secret of determinants lies in how they change when row operations are performed. The following theorem generalizes the results of Exercises 19–24 in Section 3.1. The proof is at the end of this section.

THEOREM 3

Row Operations

Let A be a square matrix.

a. If a multiple of one row of A is added to another row to produce a matrix B, then $\det B = \det A$.

b. If two rows of A are interchanged to produce B, then $\det B = -\det A$.

c. If one row of A is multiplied by k to produce B, then $\det B = k \det A$.

The following examples show how to use Theorem 3 to find determinants efficiently.

EXAMPLE 1 Compute $\det A$, where $A = \begin{bmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ -1 & 7 & 0 \end{bmatrix}$.

SOLUTION The strategy is to reduce A to echelon form and then to use the fact that the determinant of a triangular matrix is the product of the diagonal entries. The first two row replacements in column 1 do not change the determinant:

$$\det A = \begin{vmatrix} 1 & -4 & 2 \\ -2 & 8 & -9 \\ -1 & 7 & 0 \end{vmatrix} = \begin{vmatrix} 1 & -4 & 2 \\ 0 & 0 & -5 \\ -1 & 7 & 0 \end{vmatrix} = \begin{vmatrix} 1 & -4 & 2 \\ 0 & 0 & -5 \\ 0 & 3 & 2 \end{vmatrix}$$

An interchange of rows 2 and 3 reverses the sign of the determinant, so

$$\det A = -\begin{vmatrix} 1 & -4 & 2 \\ 0 & 3 & 2 \\ 0 & 0 & -5 \end{vmatrix} = -(1)(3)(-5) = 15$$ ■

A common use of Theorem 3(c) in hand calculations is to *factor out a common multiple of one row* of a matrix. For instance,

$$\begin{vmatrix} * & * & * \\ 5k & -2k & 3k \\ * & * & * \end{vmatrix} = k \begin{vmatrix} * & * & * \\ 5 & -2 & 3 \\ * & * & * \end{vmatrix}$$

where the starred entries are unchanged. We use this step in the next example.

EXAMPLE 2 Compute $\det A$, where $A = \begin{bmatrix} 2 & -8 & 6 & 8 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 1 & -4 & 0 & 6 \end{bmatrix}$.

SOLUTION To simplify the arithmetic, we want a 1 in the upper-left corner. We could interchange rows 1 and 4. Instead, we factor out 2 from the top row, and then proceed with row replacements in the first column:

$$\det A = 2\begin{vmatrix} 1 & -4 & 3 & 4 \\ 3 & -9 & 5 & 10 \\ -3 & 0 & 1 & -2 \\ 1 & -4 & 0 & 6 \end{vmatrix} = 2\begin{vmatrix} 1 & -4 & 3 & 4 \\ 0 & 3 & -4 & -2 \\ 0 & -12 & 10 & 10 \\ 0 & 0 & -3 & 2 \end{vmatrix}$$

Next, we could factor out another 2 from row 3 or use the 3 in the second column as a pivot. We choose the latter operation, adding 4 times row 2 to row 3:

$$\det A = 2\begin{vmatrix} 1 & -4 & 3 & 4 \\ 0 & 3 & -4 & -2 \\ 0 & 0 & -6 & 2 \\ 0 & 0 & -3 & 2 \end{vmatrix}$$

Finally, adding $-1/2$ times row 3 to row 4, and computing the "triangular" determinant, we find that

$$\det A = 2\begin{vmatrix} 1 & -4 & 3 & 4 \\ 0 & 3 & -4 & -2 \\ 0 & 0 & -6 & 2 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 2\,(1)(3)(-6)(1) = -36$$ ■

$$U = \begin{bmatrix} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & \blacksquare & * \\ 0 & 0 & 0 & \blacksquare \end{bmatrix}$$

$\det U \neq 0$

$$U = \begin{bmatrix} \blacksquare & * & * & * \\ 0 & \blacksquare & * & * \\ 0 & 0 & 0 & \blacksquare \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$\det U = 0$

FIGURE 1
Typical echelon forms of square matrices.

Suppose a square matrix A has been reduced to an echelon form U by row replacements and row interchanges. (This is always possible. See the row reduction algorithm in Section 1.2.) If there are r interchanges, then Theorem 3 shows that

$$\det A = (-1)^r \det U$$

Since U is in echelon form, it is triangular, and so $\det U$ is the product of the diagonal entries $u_{11}, \ldots, u_{nn}$. If A is invertible, the entries u_{ii} are all pivots (because $A \sim I_n$ and the u_{ii} have not been scaled to 1's). Otherwise, at least u_{nn} is zero, and the product $u_{11} \cdots u_{nn}$ is zero. See Figure 1. Thus

$$\det A = \begin{cases} (-1)^r \begin{pmatrix} \text{product of} \\ \text{pivots in } U \end{pmatrix} & \text{when } A \text{ is invertible} \\ 0 & \text{when } A \text{ is not invertible} \end{cases} \tag{1}$$

It is interesting to note that although the echelon form U described above is not unique (because it is not completely row reduced), and the pivots are not unique, the *product* of the pivots *is* unique, except for a possible minus sign.

Formula (1) not only gives a concrete interpretation of $\det A$ but also proves the main theorem of this section:

THEOREM 4

A square matrix A is invertible if and only if $\det A \neq 0$.

Theorem 4 adds the statement "$\det A \neq 0$" to the Invertible Matrix Theorem. A useful corollary is that $\det A = 0$ when the columns of A are linearly dependent. Also, $\det A = 0$ when the *rows* of A are linearly dependent. (Rows of A are columns of A^T, and linearly dependent columns of A^T make A^T singular. When A^T is singular, so is A, by the Invertible Matrix Theorem.) In practice, linear dependence is obvious when two columns or two rows are the same or a column or a row is zero.

EXAMPLE 3 Compute $\det A$, where $A = \begin{bmatrix} 3 & -1 & 2 & -5 \\ 0 & 5 & -3 & -6 \\ -6 & 7 & -7 & 4 \\ -5 & -8 & 0 & 9 \end{bmatrix}$.

SOLUTION Add 2 times row 1 to row 3 to obtain

$$\det A = \det \begin{bmatrix} 3 & -1 & 2 & -5 \\ 0 & 5 & -3 & -6 \\ 0 & 5 & -3 & -6 \\ -5 & -8 & 0 & 9 \end{bmatrix} = 0$$

because the second and third rows of the second matrix are equal. ■

Numerical Notes

1. Most computer programs that compute $\det A$ for a general matrix A use the method of formula (1) above.
2. It can be shown that evaluation of an $n \times n$ determinant using row operations requires about $2n^3/3$ arithmetic operations. Any modern microcomputer can calculate a 25×25 determinant in a fraction of a second, since only about 10,000 operations are required.

Computers can also handle large "sparse" matrices, with special routines that take advantage of the presence of many zeros. Of course, zero entries can speed hand computations, too. The calculations in the next example combine the power of row operations with the strategy from Section 3.1 of using zero entries in cofactor expansions.

EXAMPLE 4 Compute $\det A$, where $A = \begin{bmatrix} 0 & 1 & 2 & -1 \\ 2 & 5 & -7 & 3 \\ 0 & 3 & 6 & 2 \\ -2 & -5 & 4 & -2 \end{bmatrix}$.

SOLUTION A good way to begin is to use the 2 in column 1 as a pivot, eliminating the -2 below it. Then use a cofactor expansion to reduce the size of the determinant, followed by another row replacement operation. Thus

$$\det A = \begin{vmatrix} 0 & 1 & 2 & -1 \\ 2 & 5 & -7 & 3 \\ 0 & 3 & 6 & 2 \\ 0 & 0 & -3 & 1 \end{vmatrix} = -2 \begin{vmatrix} 1 & 2 & -1 \\ 3 & 6 & 2 \\ 0 & -3 & 1 \end{vmatrix} = -2 \begin{vmatrix} 1 & 2 & -1 \\ 0 & 0 & 5 \\ 0 & -3 & 1 \end{vmatrix}$$

An interchange of rows 2 and 3 would produce a "triangular determinant." Another approach is to make a cofactor expansion down the first column:

$$\det A = (-2)(1) \begin{vmatrix} 0 & 5 \\ -3 & 1 \end{vmatrix} = -2\,(15) = -30$$ ■

Column Operations

We can perform operations on the columns of a matrix in a way that is analogous to the row operations we have considered. The next theorem shows that column operations have the same effects on determinants as row operations.

Remark: The Principle of Mathematical Induction says the following: Let $P(n)$ be a statement that is either true or false for each natural number n. Then $P(n)$ is true for all $n \geq 1$ provided that $P(1)$ is true, and for each natural number k, if $P(k)$ is true, then $P(k+1)$ is true. The Principle of Mathematical Induction is used to prove the next theorem.

THEOREM 5 If A is an $n \times n$ matrix, then $\det A^T = \det A$.

PROOF The theorem is obvious for $n = 1$. Suppose the theorem is true for $k \times k$ determinants and let $n = k + 1$. Then the cofactor of a_{1j} in A equals the cofactor of a_{j1} in A^T, because the cofactors involve $k \times k$ determinants. Hence the cofactor expansion of $\det A$ along the first *row* equals the cofactor expansion of $\det A^T$ down the first *column*. That is, A and A^T have equal determinants. The theorem is true for $n = 1$, and the truth of the theorem for one value of n implies its truth for the next value of n. By the Principle of Mathematical Induction, the theorem is true for all $n \geq 1$. ■

Because of Theorem 5, each statement in Theorem 3 is true when the word *row* is replaced everywhere by *column*. To verify this property, one merely applies the original Theorem 3 to A^T. A row operation on A^T amounts to a column operation on A.

Column operations are useful for both theoretical purposes and hand computations. However, for simplicity we'll perform only row operations in numerical calculations.

Determinants and Matrix Products

The proof of the following useful theorem is at the end of the section. Applications are in the exercises.

THEOREM 6

Multiplicative Property

If A and B are $n \times n$ matrices, then $\det AB = (\det A)(\det B)$.

EXAMPLE 5 Verify Theorem 6 for $A = \begin{bmatrix} 6 & 1 \\ 3 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix}$.

SOLUTION

$$AB = \begin{bmatrix} 6 & 1 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} 4 & 3 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 25 & 20 \\ 14 & 13 \end{bmatrix}$$

and

$$\det AB = 25(13) - 20(14) = 325 - 280 = 45$$

Since $\det A = 9$ and $\det B = 5$,

$$(\det A)(\det B) = 9(5) = 45 = \det AB$$ ■

Warning: A common misconception is that Theorem 6 has an analogue for *sums* of matrices. However, $\det(A + B)$ is *not* equal to $\det A + \det B$, in general.

A Linearity Property of the Determinant Function

For an $n \times n$ matrix A, we can consider $\det A$ as a function of the n column vectors in A. We will show that if all columns except one are held fixed, then $\det A$ is a *linear function* of that one (vector) variable.

Suppose that the jth column of A is allowed to vary, and write

$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_{j-1} & \mathbf{x} & \mathbf{a}_{j+1} & \cdots & \mathbf{a}_n \end{bmatrix}$$

Define a transformation T from $\mathbb{R}^n$ to $\mathbb{R}$ by

$$T(\mathbf{x}) = \det \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_{j-1} & \mathbf{x} & \mathbf{a}_{j+1} & \cdots & \mathbf{a}_n \end{bmatrix}$$

Then,

$$T(c\mathbf{x}) = cT(\mathbf{x}) \quad \text{for all scalars } c \text{ and all } \mathbf{x} \text{ in } \mathbb{R}^n \tag{2}$$

$$T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}) \quad \text{for all } \mathbf{u}, \mathbf{v} \text{ in } \mathbb{R}^n \tag{3}$$

Property (2) is Theorem 3(c) applied to the columns of A. A proof of property (3) follows from a cofactor expansion of $\det A$ down the jth column. (See Exercise 49.) This (multi-) linearity property of the determinant turns out to have many useful consequences that are studied in more advanced courses.

Proofs of Theorems 3 and 6

It is convenient to prove Theorem 3 when it is stated in terms of the elementary matrices discussed in Section 2.2. We call an elementary matrix E a *row replacement* (*matrix*) if E is obtained from the identity I by adding a multiple of one row to another row; E is an *interchange* if E is obtained by interchanging two rows of I; and E is *a scale by r* if E is obtained by multiplying a row of I by a nonzero scalar r. With this terminology, Theorem 3 can be reformulated as follows:

If A is an $n \times n$ matrix and E is an $n \times n$ elementary matrix, then

$$\det EA = (\det E)(\det A)$$

where

$$\det E = \begin{cases} 1 & \text{if } E \text{ is a row replacement} \\ -1 & \text{if } E \text{ is an interchange} \\ r & \text{if } E \text{ is a scale by } r \end{cases}$$

PROOF OF THEOREM 3 The proof is by induction on the size of A. The case of a 2×2 matrix was verified in Exercises 33–36 of Section 3.1. Suppose the theorem has been verified for determinants of $k \times k$ matrices with $k \geq 2$, let $n = k + 1$, and let A be $n \times n$. The action of E on A involves either two rows or only one row. So we can expand $\det EA$ across a row that is unchanged by the action of E, say, row i. Let A_{ij} (respectively, B_{ij}) be the matrix obtained by deleting row i and column j from A (respectively, EA). Then the rows of B_{ij} are obtained from the rows of A_{ij} by the same type of elementary row operation that E performs on A. Since these submatrices are only $k \times k$, the induction assumption implies that

$$\det B_{ij} = \alpha \det A_{ij}$$

where $\alpha = 1$, -1, or r, depending on the nature of E. The cofactor expansion across row i is

$$\begin{aligned} \det EA &= a_{i1}(-1)^{i+1} \det B_{i1} + \cdots + a_{in}(-1)^{i+n} \det B_{in} \\ &= \alpha a_{i1}(-1)^{i+1} \det A_{i1} + \cdots + \alpha a_{in}(-1)^{i+n} \det A_{in} \\ &= \alpha \det A \end{aligned}$$

In particular, taking $A = I_n$, we see that $\det E = 1$, -1, or r, depending on the nature of E. Thus the theorem is true for $n = 2$, and the truth of the theorem for one value of n implies its truth for the next value of n. By the principle of induction, the theorem must be true for $n \geq 2$. The theorem is trivially true for $n = 1$. ■

PROOF OF THEOREM 6 If A is not invertible, then neither is AB, by Exercise 35 in Section 2.3. In this case, $\det AB = (\det A)(\det B)$, because both sides are zero, by Theorem 4. If A is invertible, then A and the identity matrix I_n are row equivalent by the Invertible Matrix Theorem. So there exist elementary matrices $E_1, \ldots, E_p$ such that

$$A = E_p E_{p-1} \cdots E_1 I_n = E_p E_{p-1} \cdots E_1$$

For brevity, write $|A|$ for $\det A$. Then repeated application of Theorem 3, as rephrased above, shows that

$$\begin{aligned}|AB| &= |E_p \cdots E_1 B| = |E_p||E_{p-1} \cdots E_1 B| = \cdots \\ &= |E_p| \cdots |E_1||B| = \cdots = |E_p \cdots E_1||B| \\ &= |A||B|\end{aligned}$$

■

Practice Problems

1. Compute $\begin{vmatrix} 1 & -3 & 1 & -2 \\ 2 & -5 & -1 & -2 \\ 0 & -4 & 5 & 1 \\ -3 & 10 & -6 & 8 \end{vmatrix}$ in as few steps as possible.

2. Use a determinant to decide if $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are linearly independent, when

$$\mathbf{v}_1 = \begin{bmatrix} 5 \\ -7 \\ 9 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -3 \\ 3 \\ -5 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2 \\ -7 \\ 5 \end{bmatrix}$$

3. Let A be an $n \times n$ matrix such that $A^2 = I$. Show that $\det A = \pm 1$.

3.2 Exercises

Each equation in Exercises 1–4 illustrates a property of determinants. State the property.

1. $\begin{vmatrix} 0 & 5 & -2 \\ 1 & -3 & 6 \\ 4 & -1 & 8 \end{vmatrix} = -\begin{vmatrix} 1 & -3 & 6 \\ 0 & 5 & -2 \\ 4 & -1 & 8 \end{vmatrix}$

2. $\begin{vmatrix} 3 & -6 & 9 \\ 3 & 5 & -5 \\ 1 & 3 & 3 \end{vmatrix} = 3\begin{vmatrix} 1 & -2 & 3 \\ 3 & 5 & -5 \\ 1 & 3 & 3 \end{vmatrix}$

3. $\begin{vmatrix} 1 & 2 & 2 \\ 0 & 3 & -4 \\ 2 & 7 & 4 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 2 \\ 0 & 3 & -4 \\ 0 & 3 & 0 \end{vmatrix}$

4. $\begin{vmatrix} 1 & 3 & -4 \\ 2 & 0 & -3 \\ 3 & -5 & 2 \end{vmatrix} = \begin{vmatrix} 1 & 3 & -4 \\ 0 & -6 & 5 \\ 3 & -5 & 2 \end{vmatrix}$

Find the determinants in Exercises 5–10 by row reduction to echelon form.

5. $\begin{vmatrix} 1 & 5 & -4 \\ -1 & -4 & 5 \\ -2 & -8 & 7 \end{vmatrix}$

6. $\begin{vmatrix} 3 & -6 & 6 \\ 3 & -5 & 9 \\ 3 & -4 & 8 \end{vmatrix}$

7. $\begin{vmatrix} 1 & 3 & 0 & 2 \\ -2 & -5 & 7 & 4 \\ 3 & 5 & 2 & 1 \\ 1 & -1 & 2 & -3 \end{vmatrix}$

8. $\begin{vmatrix} 1 & 2 & -3 & 4 \\ 0 & 1 & 5 & 6 \\ -4 & -9 & 7 & -14 \\ 2 & 5 & 0 & 7 \end{vmatrix}$

9. $\begin{vmatrix} 1 & -1 & -3 & 0 \\ 0 & 1 & 5 & 4 \\ -1 & 0 & 5 & 3 \\ 3 & -3 & -2 & 3 \end{vmatrix}$

10. $\begin{vmatrix} 1 & 3 & -1 & 0 & -2 \\ 0 & 1 & -2 & -1 & -3 \\ -2 & -6 & 2 & 3 & 10 \\ 1 & 5 & -6 & 2 & -3 \\ 0 & 2 & -4 & 5 & 9 \end{vmatrix}$

Combine the methods of row reduction and cofactor expansion to compute the determinants in Exercises 11–14.

11. $\begin{vmatrix} 3 & 4 & -3 & -1 \\ 3 & 0 & 1 & -3 \\ -6 & 0 & -4 & 3 \\ 6 & 8 & -4 & -1 \end{vmatrix}$

12. $\begin{vmatrix} -2 & 6 & 0 & 9 \\ 3 & 4 & 8 & 2 \\ 4 & 3 & 0 & 1 \\ 3 & 1 & 2 & -1 \end{vmatrix}$

13. $\begin{vmatrix} 2 & 5 & 4 & 1 \\ 4 & 7 & 6 & 2 \\ 6 & -2 & -4 & 0 \\ -6 & 7 & 7 & 0 \end{vmatrix}$

14. $\begin{vmatrix} 4 & 3 & 2 & 1 \\ 5 & 4 & -3 & 0 \\ 9 & -8 & -7 & 0 \\ 4 & 6 & 2 & 1 \end{vmatrix}$

Find the determinants in Exercises 15–20, where

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = 7.$$

15. $\begin{vmatrix} a & b & c \\ d & e & f \\ 3g & 3h & 3i \end{vmatrix}$ **16.** $\begin{vmatrix} a & b & c \\ d+3g & e+3h & f+3i \\ g & h & i \end{vmatrix}$

17. $\begin{vmatrix} a+d & b+e & c+f \\ d & e & f \\ g & h & i \end{vmatrix}$

18. $\begin{vmatrix} a & b & c \\ 8d & 8e & 8f \\ g & h & i \end{vmatrix}$

19. $\begin{vmatrix} a & b & c \\ 2d+a & 2e+b & 2f+c \\ g & h & i \end{vmatrix}$

20. $\begin{vmatrix} g & h & i \\ a & b & c \\ d & e & f \end{vmatrix}$

In Exercises 21–23, use determinants to find out if the matrix is invertible.

21. $\begin{bmatrix} 1 & 3 & 6 \\ 2 & 4 & 7 \\ 0 & 5 & 8 \end{bmatrix}$ **22.** $\begin{bmatrix} 4 & 5 & 0 \\ 3 & 2 & 1 \\ 1 & -4 & 3 \end{bmatrix}$

23. $\begin{bmatrix} 3 & 0 & 0 & 2 \\ 6 & 8 & 9 & 0 \\ 4 & 5 & 6 & 0 \\ 0 & -8 & -9 & 4 \end{bmatrix}$

In Exercises 24–26, use determinants to decide if the set of vectors is linearly independent.

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

25. $\begin{bmatrix} 7 \\ -4 \\ -6 \end{bmatrix}, \begin{bmatrix} -8 \\ 5 \\ 7 \end{bmatrix}, \begin{bmatrix} 7 \\ 0 \\ -5 \end{bmatrix}$

26. $\begin{bmatrix} 3 \\ 5 \\ -6 \\ 4 \end{bmatrix}, \begin{bmatrix} 2 \\ -6 \\ 0 \\ 7 \end{bmatrix}, \begin{bmatrix} -2 \\ -1 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ -2 \end{bmatrix}$

In Exercises 27–34, A and B are $n \times n$ matrices. Mark each statement True or False **(T/F)**. Justify each answer.

27. **(T/F)** A row replacement operation does not affect the determinant of a matrix.

28. **(T/F)** If $\det A$ is zero, then two rows or two columns are the same, or a row or a column is zero.

29. **(T/F)** If the columns of A are linearly dependent, then $\det A = 0$.

30. **(T/F)** The determinant of A is the product of the diagonal entries in A.

31. **(T/F)** If three row interchanges are made in succession, then the new determinant equals the old determinant.

32. **(T/F)** The determinant of A is the product of the pivots in any echelon form U of A, multiplied by $(-1)^r$, where r is the number of row interchanges made during row reduction from A to U.

33. **(T/F)** $\det(A + B) = \det A + \det B$.

34. **(T/F)** $\det A^{-1} = (-1)\det A$.

35. Compute $\det B^4$, where $B = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix}$.

36. Use Theorem 3 (but not Theorem 4) to show that if two rows of a square matrix A are equal, then $\det A = 0$. The same is true for two columns. Why?

In Exercises 37–42, mention an appropriate theorem in your explanation.

37. Show that if A is invertible, then $\det A^{-1} = \dfrac{1}{\det A}$.

38. Suppose that A is a square matrix such that $\det A^3 = 0$. Explain why A cannot be invertible.

39. Let A and B be square matrices. Show that even though AB and BA may not be equal, it is always true that $\det AB = \det BA$.

40. Let A and P be square matrices, with P invertible. Show that $\det(PAP^{-1}) = \det A$.

41. Let U be a square matrix such that $U^TU = I$. Show that $\det U = \pm 1$.

42. Find a formula for $\det(rA)$ when A is an $n \times n$ matrix.

Verify that $\det AB = (\det A)(\det B)$ for the matrices in Exercises 43 and 44. (Do not use Theorem 6.)

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

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

45. Let A and B be 3×3 matrices, with $\det A = -2$ and $\det B = 3$. Use properties of determinants (in the text and in the preceding exercises) to compute:

a. $\det AB$ b. $\det 5A$ c. $\det B^T$

d. $\det A^{-1}$ e. $\det A^3$

46. Let A and B be 4×4 matrices, with $\det A = 4$ and $\det B = -5$. Compute:

a. $\det AB$ b. $\det 3A$ c. $\det B^4$

d. $\det BA\,B^T$ e. $\det AB\,A^{-1}$

47. Verify that $\det A = \det B + \det C$, where

$$A = \begin{bmatrix} a+e & b+f \\ c & d \end{bmatrix},\ B = \begin{bmatrix} a & b \\ c & d \end{bmatrix},\ C = \begin{bmatrix} e & f \\ c & d \end{bmatrix}$$

48. Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Show that $\det(A+B) = \det A + \det B$ if and only if $a + d = 0$.

49. Verify that $\det A = \det B + \det C$, where

$$A = \begin{bmatrix} a_{11} & a_{12} & u_1+v_1 \\ a_{21} & a_{22} & u_2+v_2 \\ a_{31} & a_{32} & u_3+v_3 \end{bmatrix},$$

$$B = \begin{bmatrix} a_{11} & a_{12} & u_1 \\ a_{21} & a_{22} & u_2 \\ a_{31} & a_{32} & u_3 \end{bmatrix},\ C = \begin{bmatrix} a_{11} & a_{12} & v_1 \\ a_{21} & a_{22} & v_2 \\ a_{31} & a_{32} & v_3 \end{bmatrix}$$

Note, however, that A is *not* the same as $B + C$.

50. Right-multiplication by an elementary matrix E affects the *columns* of A in the same way that left-multiplication affects the *rows*. Use Theorems 5 and 3 and the obvious fact that E^T is another elementary matrix to show that

$$\det AE = (\det E)(\det A)$$

Do not use Theorem 6.

51. Suppose A is an $n \times n$ matrix and a computer suggests that $\det A = 5$ and $\det\left(A^{-1}\right) = 1$. Should you trust these answers? Why or why not?

52. Suppose A and B are $n \times n$ matrices and a computer suggests that $\det A = 5$, $\det B = 2$ and $\det AB = 7$. Should you trust these answers? Why or why not?

T 53. Compute $\det A^T A$ and $\det AA^T$ for several random 4×5 matrices and several random 5×6 matrices. What can you say about $A^T A$ and AA^T when A has more columns than rows?

T 54. If $\det A$ is close to zero, is the matrix A nearly singular? Experiment with the nearly singular 4×4 matrix

$$A = \begin{bmatrix} 4 & 0 & -7 & -7 \\ -6 & 1 & 11 & 9 \\ 7 & -5 & 10 & 19 \\ -1 & 2 & 3 & -1 \end{bmatrix}$$

Compute the determinants of A, $10A$, and $0.1A$. In contrast, compute the condition numbers of these matrices. Repeat these calculations when A is the 4×4 identity matrix. Discuss your results.

Solutions to Practice Problems

1. Perform row replacements to create zeros in the first column, and then create a row of zeros.

$$\begin{vmatrix} 1 & -3 & 1 & -2 \\ 2 & -5 & -1 & -2 \\ 0 & -4 & 5 & 1 \\ -3 & 10 & -6 & 8 \end{vmatrix} = \begin{vmatrix} 1 & -3 & 1 & -2 \\ 0 & 1 & -3 & 2 \\ 0 & -4 & 5 & 1 \\ 0 & 1 & -3 & 2 \end{vmatrix} = \begin{vmatrix} 1 & -3 & 1 & -2 \\ 0 & 1 & -3 & 2 \\ 0 & -4 & 5 & 1 \\ 0 & 0 & 0 & 0 \end{vmatrix} = 0$$

2. $\det[\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3\,] = \begin{vmatrix} 5 & -3 & 2 \\ -7 & 3 & -7 \\ 9 & -5 & 5 \end{vmatrix} = \begin{vmatrix} 5 & -3 & 2 \\ -2 & 0 & -5 \\ 9 & -5 & 5 \end{vmatrix}$ Row 1 added to row 2

$$= -(-3)\begin{vmatrix} -2 & -5 \\ 9 & 5 \end{vmatrix} - (-5)\begin{vmatrix} 5 & 2 \\ -2 & -5 \end{vmatrix}$$ Cofactors of column 2

$$= 3(35) + 5(-21) = 0$$

By Theorem 4, the matrix $[\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3\,]$ is not invertible. The columns are linearly dependent, by the Invertible Matrix Theorem.

3. Recall that $\det I = 1$. By Theorem 6, $\det(AA) = (\det A)(\det A)$. Putting these two observations together results in

$$1 = \det I = \det A^2 = \det(AA) = (\det A)(\det A) = (\det A)^2$$

Taking the square root of both sides establishes that $\det A = \pm 1$.

3.3 Cramer's Rule, Volume, and Linear Transformations

This section applies the theory of the preceding sections to obtain important theoretical formulas and a geometric interpretation of the determinant.

Cramer's Rule

Cramer's rule is needed in a variety of theoretical calculations. For instance, it can be used to study how the solution of $A\mathbf{x} = \mathbf{b}$ is affected by changes in the entries of $\mathbf{b}$. However, the formula is inefficient for hand calculations, except for 2×2 or perhaps 3×3 matrices.

For any $n \times n$ matrix A and any $\mathbf{b}$ in $\mathbb{R}^n$, let $A_i(\mathbf{b})$ be the matrix obtained from A by replacing column i by the vector $\mathbf{b}$.

$$A_i(\mathbf{b}) = [\mathbf{a}_1 \quad \cdots \quad \underset{\text{col } i}{\mathbf{b}} \quad \cdots \quad \mathbf{a}_n]$$

THEOREM 7

Cramer's Rule

Let A be an invertible $n \times n$ matrix. For any $\mathbf{b}$ in $\mathbb{R}^n$, the unique solution $\mathbf{x}$ of $A\mathbf{x} = \mathbf{b}$ has entries given by

$$x_i = \frac{\det A_i(\mathbf{b})}{\det A}, \qquad i = 1, 2, \ldots, n \tag{1}$$

PROOF Denote the columns of A by $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and the columns of the $n \times n$ identity matrix I by $\mathbf{e}_1, \ldots, \mathbf{e}_n$. If $A\mathbf{x} = \mathbf{b}$, the definition of matrix multiplication shows that

$$\begin{aligned} A(I_i(\mathbf{x})) &= A\begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{x} & \cdots & \mathbf{e}_n \end{bmatrix} = \begin{bmatrix} A\mathbf{e}_1 & \cdots & A\mathbf{x} & \cdots & A\mathbf{e}_n \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{b} & \cdots & \mathbf{a}_n \end{bmatrix} = A_i(\mathbf{b}) \end{aligned}$$

By the multiplicative property of determinants,

$$(\det A)(\det I_i(\mathbf{x})) = \det A_i(\mathbf{b})$$

The second determinant on the left is simply x_i. (Make a cofactor expansion along the ith row.) Hence $(\det A)\, x_i = \det A_i(\mathbf{b})$. This proves (1) because A is invertible and $\det A \neq 0$. ■

EXAMPLE 1 Use Cramer's rule to solve the system

$$\begin{aligned} 3x_1 - 2x_2 &= 6 \\ -5x_1 + 4x_2 &= 8 \end{aligned}$$

SOLUTION View the system as $A\mathbf{x} = \mathbf{b}$. Using the notation introduced above,

$$A = \begin{bmatrix} 3 & -2 \\ -5 & 4 \end{bmatrix}, \qquad A_1(\mathbf{b}) = \begin{bmatrix} 6 & -2 \\ 8 & 4 \end{bmatrix}, \qquad A_2(\mathbf{b}) = \begin{bmatrix} 3 & 6 \\ -5 & 8 \end{bmatrix}$$

Since $\det A = 2$, the system has a unique solution. By Cramer's rule,

$$x_1 = \frac{\det A_1(\mathbf{b})}{\det A} = \frac{24 + 16}{2} = 20$$
$$x_2 = \frac{\det A_2(\mathbf{b})}{\det A} = \frac{24 + 30}{2} = 27$$

■

Application to Engineering

A number of important engineering problems, particularly in electrical engineering and control theory, can be analyzed by *Laplace transforms*. This approach converts an appropriate system of linear differential equations into a system of linear algebraic equations whose coefficients involve a parameter. The next example illustrates the type of algebraic system that may arise.

EXAMPLE 2 Consider the following system in which s is an unspecified parameter. Determine the values of s for which the system has a unique solution, and use Cramer's rule to describe the solution.

$$3sx_1 - 2x_2 = 4$$
$$-6x_1 + sx_2 = 1$$

SOLUTION View the system as $A\mathbf{x} = \mathbf{b}$. Then

$$A = \begin{bmatrix} 3s & -2 \\ -6 & s \end{bmatrix}, \quad A_1(\mathbf{b}) = \begin{bmatrix} 4 & -2 \\ 1 & s \end{bmatrix}, \quad A_2(\mathbf{b}) = \begin{bmatrix} 3s & 4 \\ -6 & 1 \end{bmatrix}$$

Since

$$\det A = 3s^2 - 12 = 3(s + 2)(s - 2)$$

the system has a unique solution precisely when $s \neq \pm 2$. For such an s, the solution is (x_1, x_2), where

$$x_1 = \frac{\det A_1(\mathbf{b})}{\det A} = \frac{4s + 2}{3(s + 2)(s - 2)}$$
$$x_2 = \frac{\det A_2(\mathbf{b})}{\det A} = \frac{3s + 24}{3(s + 2)(s - 2)} = \frac{s + 8}{(s + 2)(s - 2)}$$

■

A Formula for A^{-1}

Cramer's rule leads easily to a general formula for the inverse of an $n \times n$ matrix A. The jth column of A^{-1} is a vector $\mathbf{x}$ that satisfies

$$A\mathbf{x} = \mathbf{e}_j$$

where $\mathbf{e}_j$ is the jth column of the identity matrix, and the ith entry of $\mathbf{x}$ is the (i, j)-entry of A^{-1}. By Cramer's rule,

$$\{(i, j)\text{-entry of } A^{-1}\} = x_i = \frac{\det A_i(\mathbf{e}_j)}{\det A} \tag{2}$$

Recall that A_{ji} denotes the submatrix of A formed by deleting row j and column i. A cofactor expansion down column i of $A_i(\mathbf{e}_j)$ shows that

$$\det A_i(\mathbf{e}_j) = (-1)^{i+j} \det A_{ji} = C_{ji} \tag{3}$$

where C_{ji} is a cofactor of A. By (2), the (i, j)-entry of A^{-1} is the cofactor C_{ji} divided by $\det A$. [Note that the subscripts on C_{ji} are the reverse of (i, j).] Thus

$$A^{-1} = \frac{1}{\det A}\begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix} \tag{4}$$

The matrix of cofactors on the right side of (4) is called the **adjugate** (or **classical adjoint**) of A, denoted by adj A. (The term *adjoint* also has another meaning in advanced texts on linear transformations.) The next theorem simply restates (4).

THEOREM 8

An Inverse Formula

Let A be an invertible $n \times n$ matrix. Then

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A$$

EXAMPLE 3 Find the inverse of the matrix $A = \begin{bmatrix} 2 & 1 & 3 \\ 1 & -1 & 1 \\ 1 & 4 & -2 \end{bmatrix}$.

SOLUTION The nine cofactors are

$$C_{11} = +\begin{vmatrix} -1 & 1 \\ 4 & -2 \end{vmatrix} = -2, \quad C_{12} = -\begin{vmatrix} 1 & 1 \\ 1 & -2 \end{vmatrix} = 3, \quad C_{13} = +\begin{vmatrix} 1 & -1 \\ 1 & 4 \end{vmatrix} = 5$$

$$C_{21} = -\begin{vmatrix} 1 & 3 \\ 4 & -2 \end{vmatrix} = 14, \quad C_{22} = +\begin{vmatrix} 2 & 3 \\ 1 & -2 \end{vmatrix} = -7, \quad C_{23} = -\begin{vmatrix} 2 & 1 \\ 1 & 4 \end{vmatrix} = -7$$

$$C_{31} = +\begin{vmatrix} 1 & 3 \\ -1 & 1 \end{vmatrix} = 4, \quad C_{32} = -\begin{vmatrix} 2 & 3 \\ 1 & 1 \end{vmatrix} = 1, \quad C_{33} = +\begin{vmatrix} 2 & 1 \\ 1 & -1 \end{vmatrix} = -3$$

The adjugate matrix is the *transpose* of the matrix of cofactors. [For instance, C_{12} goes in the $(2, 1)$ position.] Thus

$$\operatorname{adj} A = \begin{bmatrix} -2 & 14 & 4 \\ 3 & -7 & 1 \\ 5 & -7 & -3 \end{bmatrix}$$

We could compute $\det A$ directly, but the following computation provides a check on the calculations for adj A *and* produces $\det A$:

$$(\operatorname{adj} A)\, A = \begin{bmatrix} -2 & 14 & 4 \\ 3 & -7 & 1 \\ 5 & -7 & -3 \end{bmatrix}\begin{bmatrix} 2 & 1 & 3 \\ 1 & -1 & 1 \\ 1 & 4 & -2 \end{bmatrix} = \begin{bmatrix} 14 & 0 & 0 \\ 0 & 14 & 0 \\ 0 & 0 & 14 \end{bmatrix} = 14I$$

Since $(\operatorname{adj} A)A = 14I$, Theorem 8 shows that $\det A = 14$ and

$$A^{-1} = \frac{1}{14}\begin{bmatrix} -2 & 14 & 4 \\ 3 & -7 & 1 \\ 5 & -7 & -3 \end{bmatrix} = \begin{bmatrix} -1/7 & 1 & 2/7 \\ 3/14 & -1/2 & 1/14 \\ 5/14 & -1/2 & -3/14 \end{bmatrix}$$

■

Numerical Notes

Theorem 8 is useful mainly for theoretical calculations. The formula for A^{-1} permits one to deduce properties of the inverse without actually calculating it. Except for special cases, the algorithm in Section 2.2 gives a much better way to compute A^{-1}, if the inverse is really needed.

Cramer's rule is also a theoretical tool. It can be used to study how sensitive the solution of $A\mathbf{x} = \mathbf{b}$ is to changes in an entry in $\mathbf{b}$ or in A (perhaps due to experimental error when acquiring the entries for $\mathbf{b}$ or A). When A is a 3×3 matrix with *complex* entries, Cramer's rule is sometimes selected for hand computation because row reduction of $[\,A \quad \mathbf{b}\,]$ with complex arithmetic can be messy, and the determinants are fairly easy to compute. For a larger $n \times n$ matrix (real or complex), Cramer's rule is hopelessly inefficient. Computing just *one* determinant takes about as much work as solving $A\mathbf{x} = \mathbf{b}$ by row reduction.

Determinants as Area or Volume

In the next application, we verify the geometric interpretation of determinants described in the chapter introduction. Although a general discussion of length and distance in $\mathbb{R}^n$ will not be given until Chapter 6, we assume here that the usual Euclidean concepts of length, area, and volume are already understood for $\mathbb{R}^2$ and $\mathbb{R}^3$.

THEOREM 9

If A is a 2×2 matrix, the area of the parallelogram determined by the columns of A is $|\det A|$. If A is a 3×3 matrix, the volume of the parallelepiped determined by the columns of A is $|\det A|$.

STUDY GUIDE provides a geometric proof of the determinant as area.

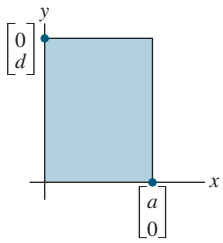


FIGURE 1
Area $= |ad|$.

PROOF The theorem is obviously true for any 2×2 diagonal matrix:

$$\left|\det\begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}\right| = |ad| = \begin{Bmatrix} \text{area of} \\ \text{rectangle} \end{Bmatrix}$$

See Figure 1. It will suffice to show that any 2×2 matrix $A = [\,\mathbf{a}_1 \quad \mathbf{a}_2\,]$ can be transformed into a diagonal matrix in a way that changes neither the area of the associated parallelogram nor $|\det A|$. From Section 3.2, we know that the absolute value of the determinant is unchanged when two columns are interchanged or a multiple of one column is added to another. And it is easy to see that such operations suffice to transform A into a diagonal matrix. Column interchanges do not change the parallelogram at all. So it suffices to prove the following simple geometric observation that applies to vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$:

Let $\mathbf{a}_1$ and $\mathbf{a}_2$ be nonzero vectors. Then for any scalar c, the area of the parallelogram determined by $\mathbf{a}_1$ and $\mathbf{a}_2$ equals the area of the parallelogram determined by $\mathbf{a}_1$ and $\mathbf{a}_2 + c\mathbf{a}_1$.

To prove this statement, we may assume that $\mathbf{a}_2$ is not a multiple of $\mathbf{a}_1$, for otherwise the two parallelograms would be degenerate and have zero area. If L is the line

through $\mathbf{0}$ and $\mathbf{a}_1$, then $\mathbf{a}_2 + L$ is the line through $\mathbf{a}_2$ parallel to L, and $\mathbf{a}_2 + c\mathbf{a}_1$ is on this line. See Figure 2. The points $\mathbf{a}_2$ and $\mathbf{a}_2 + c\mathbf{a}_1$ have the same perpendicular distance to L. Hence the two parallelograms in Figure 2 have the same area, since they share the base from $\mathbf{0}$ to $\mathbf{a}_1$. This completes the proof for $\mathbb{R}^2$.

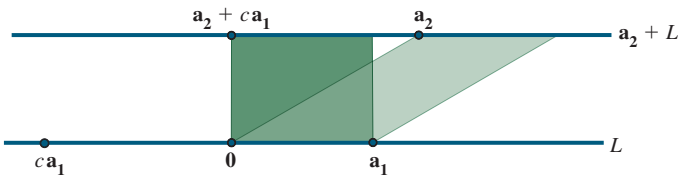


FIGURE 2 Two parallelograms of equal area.

The proof for $\mathbb{R}^3$ is similar. The theorem is obviously true for a 3×3 diagonal matrix. See Figure 3. And any 3×3 matrix A can be transformed into a diagonal matrix using column operations that do not change $|\det A|$. (Think about doing row operations on A^T.) So it suffices to show that these operations do not affect the volume of the parallelepiped determined by the columns of A.

A parallelepiped is shown in Figure 4 as a shaded box with two sloping sides. Its volume is the area of the base in the plane $\text{Span}\{\mathbf{a}_1, \mathbf{a}_3\}$ times the altitude of $\mathbf{a}_2$ above $\text{Span}\{\mathbf{a}_1, \mathbf{a}_3\}$. Any vector $\mathbf{a}_2 + c\mathbf{a}_1$ has the same altitude because $\mathbf{a}_2 + c\mathbf{a}_1$ lies in the plane $\mathbf{a}_2 + \text{Span}\{\mathbf{a}_1, \mathbf{a}_3\}$, which is parallel to $\text{Span}\{\mathbf{a}_1, \mathbf{a}_3\}$. Hence the volume of the parallelepiped is unchanged when $[\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3\,]$ is changed to $[\,\mathbf{a}_1 \quad \mathbf{a}_2 + c\mathbf{a}_1 \quad \mathbf{a}_3\,]$. Thus a column replacement operation does not affect the volume of the parallelepiped. Since column interchanges have no effect on the volume, the proof is complete. ■

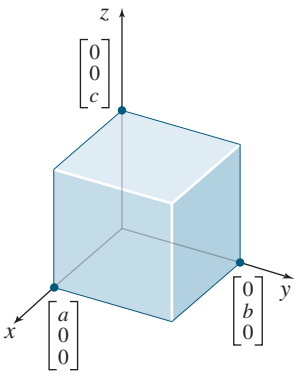


FIGURE 3
Volume $= |abc|$.

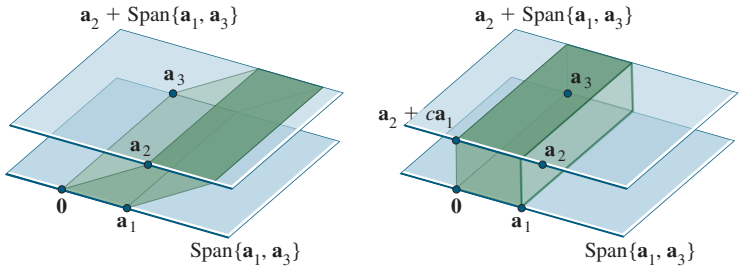


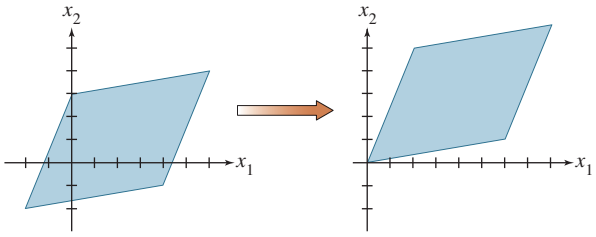
FIGURE 4 Two parallelepipeds of equal volume.

EXAMPLE 4 Calculate the area of the parallelogram determined by the points $(-2, -2)$, $(0, 3)$, $(4, -1)$, and $(6, 4)$. See Figure 5(a).

SOLUTION First translate the parallelogram to one having the origin as a vertex. For example, subtract the vertex $(-2, -2)$ from each of the four vertices. The new parallelogram has the same area, and its vertices are $(0, 0)$, $(2, 5)$, $(6, 1)$, and $(8, 6)$. See Figure 5(b). This parallelogram is determined by the columns of

$$A = \begin{bmatrix} 2 & 6 \\ 5 & 1 \end{bmatrix}$$

Since $|\det A| = |-28|$, the area of the parallelogram is 28. ■



(a) (b)

FIGURE 5 Translating a parallelogram does not change its area.

Linear Transformations

Determinants can be used to describe an important geometric property of linear transformations in the plane and in $\mathbb{R}^3$. If T is a linear transformation and S is a set in the domain of T, let $T(S)$ denote the set of images of points in S. We are interested in how the area (or volume) of $T(S)$ compares with the area (or volume) of the original set S. For convenience, when S is a region bounded by a parallelogram, we also refer to S as a parallelogram.

THEOREM 10

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation determined by a 2×2 matrix A. If S is a parallelogram in $\mathbb{R}^2$, then

$$\{\text{area of } T(S)\} = |\det A| \cdot \{\text{area of } S\} \tag{5}$$

If T is determined by a 3×3 matrix A, and if S is a parallelepiped in $\mathbb{R}^3$, then

$$\{\text{volume of } T(S)\} = |\det A| \cdot \{\text{volume of } S\} \tag{6}$$

PROOF Consider the 2×2 case, with $A = [\,\mathbf{a}_1 \quad \mathbf{a}_2\,]$. A parallelogram at the origin in $\mathbb{R}^2$ determined by vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ has the form

$$S = \{s_1\mathbf{b}_1 + s_2\mathbf{b}_2 : 0 \le s_1 \le 1,\ 0 \le s_2 \le 1\}$$

The image of S under T consists of points of the form

$$\begin{aligned} T(s_1\mathbf{b}_1 + s_2\mathbf{b}_2) &= s_1T(\mathbf{b}_1) + s_2T(\mathbf{b}_2) \\ &= s_1A\mathbf{b}_1 + s_2A\mathbf{b}_2 \end{aligned}$$

where $0 \le s_1 \le 1$, $0 \le s_2 \le 1$. It follows that $T(S)$ is the parallelogram determined by the columns of the matrix $[\,A\mathbf{b}_1 \quad A\mathbf{b}_2\,]$. This matrix can be written as AB, where $B = [\,\mathbf{b}_1 \quad \mathbf{b}_2\,]$. By Theorem 9 and the product theorem for determinants,

$$\begin{aligned} \{\text{area of } T(S)\} &= |\det AB| = |\det A|\,|\det B| \\ &= |\det A| \cdot \{\text{area of } S\} \end{aligned} \tag{7}$$

An arbitrary parallelogram has the form $\mathbf{p} + S$, where $\mathbf{p}$ is a vector and S is a parallelogram at the origin, as seen previously. It is easy to see that T transforms $\mathbf{p} + S$

into $T(\mathbf{p}) + T(S)$. (See Exercise 26.) Since translation does not affect the area of a set,

$$\begin{aligned}\{\text{area of } T(\mathbf{p}+S)\} &= \{\text{area of } T(\mathbf{p}) + T(S)\} \\ &= \{\text{area of } T(S)\} && \text{Translation} \\ &= |\det A| \cdot \{\text{area of } S\} && \text{By equation (7)} \\ &= |\det A| \cdot \{\text{area of } (\mathbf{p}+S)\} && \text{Translation}\end{aligned}$$

This shows that (5) holds for all parallelograms in $\mathbb{R}^2$. The proof of (6) for the 3×3 case is analogous. ■

When we attempt to generalize Theorem 10 to a region in $\mathbb{R}^2$ or $\mathbb{R}^3$ that is not bounded by straight lines or planes, we must face the problem of how to define and compute its area or volume. This is a question studied in calculus, and we shall only outline the basic idea for $\mathbb{R}^2$. If R is a planar region that has a finite area, then R can be approximated by a grid of small squares that lie inside R. By making the squares sufficiently small, the area of R may be approximated as closely as desired by the sum of the areas of the small squares. See Figure 6.

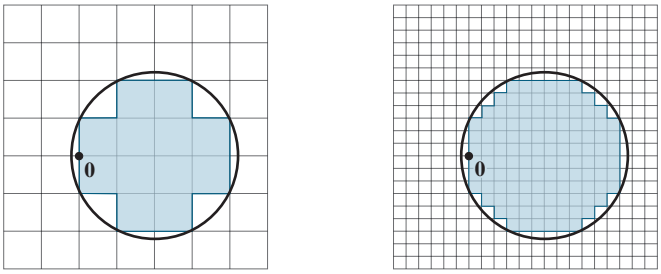


FIGURE 6 Approximating a planar region by a union of squares. The approximation improves as the grid becomes finer.

If T is a linear transformation associated with a 2×2 matrix A, then the image of a planar region R under T is approximated by the images of the small squares inside R. The proof of Theorem 10 shows that each such image is a parallelogram whose area is $|\det A|$ times the area of the square. If R' is the union of the squares inside R, then the area of $T(R')$ is $|\det A|$ times the area of R'. See Figure 7. Also, the area of $T(R')$ is close to the area of $T(R)$. An argument involving a limiting process may be given to justify the following generalization of Theorem 10.

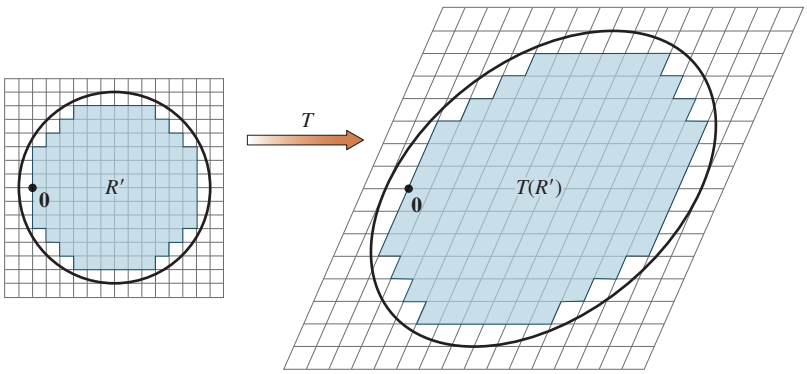
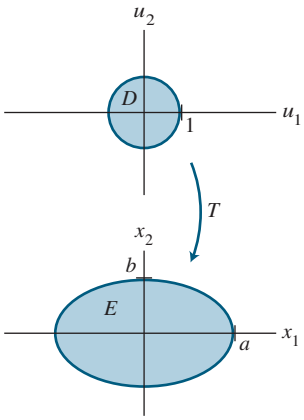


FIGURE 7 Approximating $T(R)$ by a union of parallelograms.

The conclusions of Theorem 10 hold whenever S is a region in $\mathbb{R}^2$ with finite area or a region in $\mathbb{R}^3$ with finite volume.

EXAMPLE 5 Let a and b be positive numbers. Find the area of the region E bounded by the ellipse whose equation is

$$\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1$$



SOLUTION We claim that E is the image of the unit disk D under the linear transformation T determined by the matrix $A = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$, because if $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, and $\mathbf{x} = A\mathbf{u}$, then

$$u_1 = \frac{x_1}{a} \quad \text{and} \quad u_2 = \frac{x_2}{b}$$

It follows that $\mathbf{u}$ is in the unit disk, with $u_1^2 + u_2^2 \le 1$, if and only if $\mathbf{x}$ is in E, with $(x_1/a)^2 + (x_2/b)^2 \le 1$. By the generalization of Theorem 10,

$$\begin{aligned} \{\text{area of ellipse}\} &= \{\text{area of } T(D)\} \\ &= |\det A| \cdot \{\text{area of } D\} \\ &= ab\pi(1)^2 = \pi ab \end{aligned}$$

■

Practice Problem

Let S be the parallelogram determined by the vectors $\mathbf{b}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\mathbf{b}_2 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$, and let $A = \begin{bmatrix} 1 & -.1 \\ 0 & 2 \end{bmatrix}$. Compute the area of the image of S under the mapping $\mathbf{x} \mapsto A\mathbf{x}$.

3.3 Exercises

Use Cramer's rule to compute the solutions of the systems in Exercises 1–6.

1. $5x_1 + 7x_2 = 3$
$2x_1 + 4x_2 = 1$

2. $6x_1 + x_2 = 3$
$5x_1 + 2x_2 = 4$

3. $3x_1 - 2x_2 = 3$
$-4x_1 + 6x_2 = -5$

4. $-5x_1 + 2x_2 = 9$
$3x_1 - x_2 = -4$

5. $x_1 + x_2 = 2$
$-5x_1 + 4x_3 = 0$
$x_2 - x_3 = -1$

6. $x_1 + 3x_2 + x_3 = 8$
$-x_1 + 2x_3 = 4$
$3x_1 + x_2 = 4$

In Exercises 7–10, determine the values of the parameter s for which the system has a unique solution, and describe the solution.

7. $2sx_1 + 5x_2 = 8$
$6x_1 + 3sx_2 = 4$

8. $3sx_1 + 5x_2 = 3$
$12x_1 + 5sx_2 = 2$

9. $sx_1 + 2sx_2 = -1$
$3x_1 + 6sx_2 = 4$

10. $sx_1 - 2x_2 = 1$
$4sx_1 + 4sx_2 = 2$

In Exercises 11–16, compute the adjugate of the given matrix, and then use Theorem 8 to give the inverse of the matrix.

11. $\begin{bmatrix} 0 & -2 & -1 \\ 5 & 0 & 0 \\ -1 & 1 & 1 \end{bmatrix}$

12. $\begin{bmatrix} 1 & 1 & -2 \\ -1 & 1 & 3 \\ 0 & -1 & 3 \end{bmatrix}$

13. $\begin{bmatrix} 3 & 5 & 4 \\ 1 & 0 & 1 \\ 2 & 1 & 1 \end{bmatrix}$

14. $\begin{bmatrix} 1 & -1 & 2 \\ 0 & 2 & 1 \\ 3 & 0 & 6 \end{bmatrix}$

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

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

17. Show that if A is 2×2, then Theorem 8 gives the same formula for A^{-1} as that given by Theorem 4 in Section 2.2.

18. Suppose that all the entries in A are integers and $\det A = 1$. Explain why all the entries in A^{-1} are integers.

In Exercises 19–22, find the area of the parallelogram whose vertices are listed.

19. $(0, 0), (5, 2), (6, 4), (11, 6)$

20. $(0, 0), (-3, 7), (8, -9), (5, -2)$

21. $(-6, 0), (0, 5), (4, 5), (-2, 0)$

22. $(0, -2), (5, -2), (-3, 1), (2, 1)$

23. Find the volume of the parallelepiped with one vertex at the origin and adjacent vertices at $(1, 0, -6)$, $(1, 3, 5)$, and $(6, 7, 0)$.

24. Find the volume of the parallelepiped with one vertex at the origin and adjacent vertices at $(1, 5, 0)$, $(-3, 0, 3)$, and $(-1, 4, -1)$.

25. Use the concept of volume to explain why the determinant of a 3×3 matrix A is zero if and only if A is not invertible. Do not appeal to Theorem 4 in Section 3.2. [*Hint:* Think about the columns of A.]

26. Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation, and let $\mathbf{p}$ be a vector and S a set in $\mathbb{R}^m$. Show that the image of $\mathbf{p} + S$ under T is the translated set $T(\mathbf{p}) + T(S)$ in $\mathbb{R}^n$.

27. Let S be the parallelogram determined by the vectors $\mathbf{b}_1 = \begin{bmatrix} -3 \\ 5 \end{bmatrix}$ and $\mathbf{b}_2 = \begin{bmatrix} -3 \\ 8 \end{bmatrix}$, and let $A = \begin{bmatrix} 3 & -4 \\ -4 & 6 \end{bmatrix}$. Compute the area of the image of S under the mapping $\mathbf{x} \mapsto A\mathbf{x}$.

28. Repeat Exercise 27 with $\mathbf{b}_1 = \begin{bmatrix} -3 \\ 5 \end{bmatrix}$ and $\mathbf{b}_2 = \begin{bmatrix} 0 \\ -3 \end{bmatrix}$, and $A = \begin{bmatrix} 3 & 4 \\ -2 & -2 \end{bmatrix}$.

29. Find a formula for the area of the triangle whose vertices are $\mathbf{0}$, $\mathbf{v}_1$, and $\mathbf{v}_2$ in $\mathbb{R}^2$.

30. Let R be the triangle with vertices at (x_1, y_1), (x_2, y_2), and (x_3, y_3). Show that

$$\{\text{area of triangle}\} = \frac{1}{2} \det \begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{bmatrix}$$

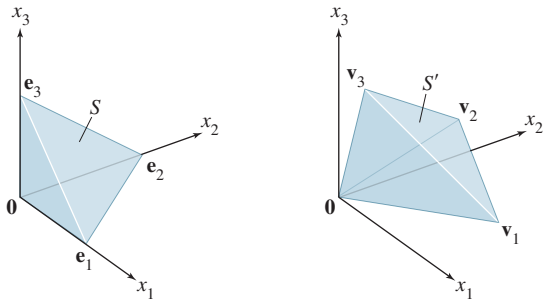
[*Hint:* Translate R to the origin by subtracting one of the vertices, and use Exercise 29.]

31. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear transformation determined by the matrix $A = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}$, where a, b, and c are positive numbers. Let S be the unit ball, whose bounding surface has the equation $x_1^2 + x_2^2 + x_3^2 = 1$.

a. Show that $T(S)$ is bounded by the ellipsoid with the equation $\dfrac{x_1^2}{a^2} + \dfrac{x_2^2}{b^2} + \dfrac{x_3^2}{c^2} = 1$.

b. Use the fact that the volume of the unit ball is $4\pi/3$ to determine the volume of the region bounded by the ellipsoid in part (a).

32. Let S be the tetrahedron in $\mathbb{R}^3$ with vertices at the vectors $\mathbf{0}$, $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$, and let S' be the tetrahedron with vertices at vectors $\mathbf{0}$, $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. See the figure.



a. Describe a linear transformation that maps S onto S'.

b. Find a formula for the volume of the tetrahedron S' using the fact that

$$\{\text{volume of } S\} = (1/3) \cdot \{\text{area of base}\} \cdot \{\text{height}\}$$

33. Let A be an $n \times n$ matrix. If $A^{-1} = \dfrac{1}{\det A} \operatorname{adj} A$ is computed, what should AA^{-1} be equal to in order to confirm that A^{-1} has been found correctly?

34. If a parallelogram fits inside a circle radius 1 and $\det A = 4$, where A is the matrix whose columns correspond to the edges of the parallelogram, does it seem like A and its determinant have been calculated correctly to correspond to the area of this parallelogram? Explain why or why not.

In Exercises 35–38, mark each statement as True or False **(T/F)**. Justify each answer.

35. **(T/F)** Two parallelograms with the same base and height have the same area.

36. **(T/F)** Applying a linear transformation to a region does not change its area.

37. **(T/F)** If A is an invertible $n \times n$ matrix, then $A^{-1} = \operatorname{adj} A$.

38. **(T/F)** Cramer's rule can only be used for invertible matrices.

T 39. Test the inverse formula of Theorem 8 for a random 4×4 matrix A. Use your matrix program to compute the cofactors of the 3×3 submatrices, construct the adjugate, and

set $B = (\text{adj } A)/(\det A)$. Then compute $B - \text{inv}(A)$, where $\text{inv}(A)$ is the inverse of A as computed by the matrix program. Use floating point arithmetic with the maximum possible number of decimal places. Report your results.

T 40. Test Cramer's rule for a random 4×4 matrix A and a random 4×1 vector $\mathbf{b}$. Compute each entry in the solution of $A\mathbf{x} = \mathbf{b}$, and compare these entries with the entries in $A^{-1}\mathbf{b}$. Write the command (or keystrokes) for your matrix program that uses Cramer's rule to produce the second entry of $\mathbf{x}$.

T 41. If your version of MATLAB has the `flops` command, use it to count the number of floating point operations to compute A^{-1} for a random 30×30 matrix. Compare this number with the number of flops needed to form $(\text{adj } A)/(\det A)$.

Solution to Practice Problem

The area of S is $\left|\det \begin{bmatrix} 1 & 5 \\ 3 & 1 \end{bmatrix}\right| = 14$, and $\det A = 2$. By Theorem 10, the area of the image of S under the mapping $\mathbf{x} \mapsto A\mathbf{x}$ is

$$|\det A| \cdot \{\text{area of } S\} = 2 \cdot 14 = 28$$

CHAPTER 3 PROJECTS

Chapter 3 projects are available online.

A. *Weighing Design*: This project develops the concept of weighing design and their corresponding matrices for use in weighing a few small, light objects.

B. *Jacobians*: This set of exercises examines how a particular determinant called the Jacobian may be used to allow us to change variables in double and triple integrals.

CHAPTER 3 SUPPLEMENTARY EXERCISES

In Exercises 1-15, mark each statement True or False **(T/F)**. Justify each answer. Assume that all matrices here are square.

1. **(T/F)** If A is a 2×2 matrix with a zero determinant, then one column of A is a multiple of the other.

2. **(T/F)** If two rows of a 3×3 matrix A are the same, then $\det A = 0$.

3. **(T/F)** If A is a 3×3 matrix, then $\det 5A = 5 \det A$.

4. **(T/F)** If A and B are $n \times n$ matrices, with $\det A = 2$ and $\det B = 3$, then $\det(A + B) = 5$.

5. **(T/F)** If A is $n \times n$ and $\det A = 2$, then $\det A^3 = 6$.

6. **(T/F)** If B is produced by interchanging two rows of A, then $\det B = \det A$.

7. **(T/F)** If B is produced by multiplying row 3 of A by 5, then $\det B = 5 \det A$.

8. **(T/F)** If B is formed by adding to one row of A a linear combination of the other rows, then $\det B = \det A$.

9. **(T/F)** $\det A^T = -\det A$.

10. **(T/F)** $\det(-A) = -\det A$.

11. **(T/F)** $\det A^T A \geq 0$.

12. **(T/F)** Any system of n linear equations in n variables can be solved by Cramer's rule.

13. **(T/F)** If $\mathbf{u}$ and $\mathbf{v}$ are in $\mathbb{R}^2$ and $\det[\,\mathbf{u} \quad \mathbf{v}\,] = 10$, then the area of the triangle in the plane with vertices at $\mathbf{0}$, $\mathbf{u}$, and $\mathbf{v}$ is 10.

14. **(T/F)** If $A^3 = 0$, then $\det A = 0$.

15. **(T/F)** If A is invertible, then $\det A^{-1} = \det A$.

Use row operations to show that the determinants in Exercises 16–18 are all zero.

16. $\begin{vmatrix} 12 & 13 & 14 \\ 15 & 16 & 17 \\ 18 & 19 & 20 \end{vmatrix}$

17. $\begin{vmatrix} 1 & a & b+c \\ 1 & b & a+c \\ 1 & c & a+b \end{vmatrix}$

18. $\begin{vmatrix} a & b & c \\ a+x & b+x & c+x \\ a+y & b+y & c+y \end{vmatrix}$

Compute the determinants in Exercises 19 and 20.

19. $\begin{vmatrix} 1 & 5 & 4 & 3 & 2 \\ 0 & 8 & 5 & 9 & 0 \\ 0 & 7 & 0 & 0 & 0 \\ 3 & 9 & 6 & 5 & 4 \\ 0 & 8 & 0 & 6 & 0 \end{vmatrix}$

20. $\begin{vmatrix} 5 & 5 & 6 & 6 & 7 \\ 4 & 0 & 4 & 4 & 0 \\ 3 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 5 & 6 & 7 & 1 & 8 \end{vmatrix}$

21. Show that the equation of the line in $\mathbb{R}^2$ through distinct points (x_1, y_1) and (x_2, y_2) can be written as

$$\det \begin{bmatrix} 1 & x & y \\ 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \end{bmatrix} = 0$$

22. Find a 3×3 determinant equation similar to that in Exercise 21 that describes the equation of the line through (x_1, y_1) with slope m.

Exercises 23 and 24 concern determinants of the following *Vandermonde matrices*.

$$T = \begin{bmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{bmatrix}, \quad V(t) = \begin{bmatrix} 1 & t & t^2 & t^3 \\ 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ 1 & x_3 & x_3^2 & x_3^3 \end{bmatrix}$$

23. Use row operations to show that

$$\det V = (x_2 - x_1)(x_3 - x_1)(x_3 - x_2)$$

24. Let x_1, x_2, and x_3 fixed numbers all distinct. Matrix V can be used to find an interpolating quadratic polynomial for the points (x_1, y_1), (x_2, y_2) and (x_3, y_3), where y_1, y_2 and y_3 are arbitrary (see Supplementary Exercise 11 in Chapter 2). Use Exercise 9 to prove the existence of an interpolating polynomial $p(t) = c_0 + c_1 t + c_2 t^2$ such that $p(x_1) = y_1$, $p(x_2) = y_2$ and $p(x_3) = y_3$.

25. Find the area of the parallelogram determined by the points $(1, 4)$, $(-1, 5)$, $(3, 9)$, and $(5, 8)$. How can you tell that the quadrilateral determined by the points is actually a parallelogram?

26. Use the concept of area of a parallelogram to write a statement about a 2×2 matrix A that is true if and only if A is invertible.

27. Show that if A is invertible, then adj A is invertible, and

$$(\text{adj } A)^{-1} = \frac{1}{\det A} A$$

[*Hint:* Given matrices B and C, what calculation(s) would show that C is the inverse of B?]

28. Let A, B, C, D, and I be $n \times n$ matrices. Use the definition or properties of a determinant to justify the following formulas. Part (c) is useful in applications of eigenvalues (Chapter 5).

a. $\det \begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix} = \det A$ b. $\det \begin{bmatrix} I & 0 \\ C & D \end{bmatrix} = \det D$

c. $\det \begin{bmatrix} A & 0 \\ C & D \end{bmatrix} = (\det A)(\det D) = \det \begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$

29. Let A, B, C, and D be $n \times n$ matrices with A invertible.

a. Find matrices X and Y to produce the block LU factorization

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ X & I \end{bmatrix} \begin{bmatrix} A & B \\ 0 & Y \end{bmatrix}$$

and then show that

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = (\det A) \cdot \det(D - CA^{-1}B)$$

b. Show that if $AC = CA$, then

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(AD - CB)$$

30. Let J be the $n \times n$ matrix of all 1's and consider $A = bI + aJ$; that is,

$$A = \begin{bmatrix} a+b & a & a & \cdots & a \\ a & a+b & a & \cdots & a \\ a & a & a+b & \cdots & a \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a & a & a & \cdots & a+b \end{bmatrix}$$

Confirm that $\det A = (na + b)b^{n-1}$ as follows:

a. Subtract row n from rows 1 to $n - 1$ and explain why this does not change the determinant of the matrix.

b. With the resulting matrix from part (a), add column 1 to column n, then add column 2 to column n and so on, and explain why this does not change the determinant of the matrix.

c. Find the determinant of the resulting matrix from (b).

31. Let A be the original matrix given in Exercise 16, and let

$$B = \begin{bmatrix} a & a & a & \cdots & a \\ a & a+b & a & \cdots & a \\ a & a & a+b & \cdots & a \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a & a & a & \cdots & a+b \end{bmatrix},$$

Notice that A, B are nearly the same except for the $(1, 1)$-entry.

a. Compute the determinant of B.

b. Use part (a) to prove the formula in Exercise 16 by induction on the size of matrix A.

T 32. Apply the result of Exercise 16 to find the determinants of the following matrices, and confirm your answers using a matrix program.

$$\begin{bmatrix} -2 & 7 & 7 & 7 \\ 7 & -2 & 7 & 7 \\ 7 & 7 & -2 & 7 \\ 7 & 7 & 7 & -2 \end{bmatrix} \qquad \begin{bmatrix} 7 & -2 & -2 & -2 & -2 \\ -2 & 7 & -2 & -2 & -2 \\ -2 & -2 & 7 & -2 & -2 \\ -2 & -2 & -2 & 7 & -2 \\ -2 & -2 & -2 & -2 & 7 \end{bmatrix}$$

T 33. Use a matrix program to compute the determinants of the following matrices.

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix} \qquad \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 & 3 \\ 1 & 2 & 3 & 4 & 4 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}$$

Use the results to guess the determinant of the matrix M, and confirm your guess by using row operations to evaluate that determinant.

$$M = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 2 & \cdots & 2 \\ 1 & 2 & 3 & \cdots & 3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & 3 & \cdots & n \end{bmatrix}$$

T 34. Use the method of Exercise 33 to guess the determinant of

$$\begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 3 & 3 & \cdots & 3 \\ 1 & 3 & 6 & \cdots & 6 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 3 & 6 & \cdots & 3(n-1) \end{bmatrix}$$

Justify your conjecture. [*Hint:* Use Exercise 28(c) and the result of Exercise 33.]

4 Vector Spaces



Introductory Example

DISCRETE-TIME SIGNALS AND DIGITAL SIGNAL PROCESSING

What is digital signal processing? Just ask Alexa, who uses signal processing to record your question and deliver the answer. In 2500 BC, the Egyptians created the first recorded discrete-time signal by carving information about the flooding of the Nile into a Palermo Stone. Despite the early beginnings of discrete-time signals, it was not until the 1940s that Claude Shannon set off the digital revolution with the ideas articulated in his paper "A Mathematical Theory of Communication."

When a person speaks into a digital processor like Alexa, it converts the sounds made by the voice into a discrete-time signal—basically a sequence of numbers $\{y_k\}$, where k represents the time at which the value y_k was recorded. Then using linear time invariant (LTI) transformations, the signal is processed to filter out unwanted noise, such as the sound of a fan running in the background. The processed signal is then compared to the signals produced by recordings of the individual sounds that make up the language of the speaker. Figure 1 shows a recording of the word "yes" and of the word "no" illustrating that the signals produced are quite distinct. Once the sounds spoken in the question are identified, machine learning is used to make a best guess at the intended question for a digital processor like Alexa. The digital processor then searches through digitized data to find the most appropriate response. Finally, the signal is processed further to produce the virtual sounds that replicate a spoken answer.

Digital signal processing (DSP) is the branch of engineering that, in the span of just a few decades, has revolutionized interpersonal communication and the entertainment industry. By reworking the principles of electronics, telecommunication, and computer science into a unifying paradigm, DSP is at the heart of the digital revolution. A smartphone fits easily in the palm of your hand, replacing numerous other devices such as cameras, video recorders, CD players, day planners, and calculators, and taking the fantasy component out of Borges's imagined Library of Babel.

The usefulness of discrete-time signals and DSP goes well beyond systems engineering. Technical analysis is employed in the investment sector. Trading opportunities are identified by applying DSP to the discrete-time signals created when the price or volume traded of a stock is recorded over time. In Example 11 of Section 4.2, price data is smoothed using a linear transformation. In the entertainment industry, audio and video are produced

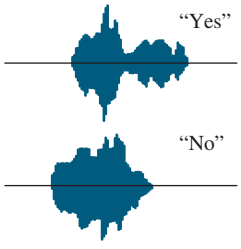


FIGURE 1

virtually and synthesized using DSP. In Example 3 of Section 4.7, we see how signal processing can be used to add richness to virtual sounds.

Discrete-time signals and DSP have become significant tools in many industries and areas of research. Mathematically speaking, discrete-time signals can be viewed as vectors that are processed using linear transformations. The operations of adding, scaling, and applying linear transformations to signals is completely analogous to the same operations for vectors in $\mathbb{R}^n$. For this reason, the set of all possible signals, $\mathbb{S}$, is treated as a *vector space*. In Sections 4.7 and 4.8, we look at the vector space of discrete-time signals in more detail.

The focus of Chapter 4 is to extend the theory of vectors in $\mathbb{R}^n$ to include signals and other mathematical structures that behave like the vectors you are already familiar with. Later on in the text, you will see how other vector spaces and their corresponding linear transformations arise in engineering, physics, biology, and statistics.

The mathematical seeds planted in Chapters 1 and 2 germinate and begin to blossom in this chapter. The beauty and power of linear algebra will be seen more clearly when you view $\mathbb{R}^n$ as only one of a variety of vector spaces that arise naturally in applied problems.

Beginning with basic definitions in Section 4.1, the general vector space framework develops gradually throughout the chapter. A goal of Sections 4.5 and 4.6 is to demonstrate how closely other vector spaces resemble $\mathbb{R}^n$. Sections 4.7 and 4.8 apply the theory of this chapter to discrete-time signals, DSP, and difference equations—the mathematics underlying the digital revolution.

4.1 Vector Spaces and Subspaces

Much of the theory in Chapters 1 and 2 rested on certain simple and obvious algebraic properties of $\mathbb{R}^n$, listed in Section 1.3. In fact, many other mathematical systems have the same properties. The specific properties of interest are listed in the following definition.

DEFINITION

A **vector space** is a nonempty set V of objects, called *vectors*, on which are defined two operations, called *addition* and *multiplication by scalars* (real numbers), subject to the ten axioms (or rules) listed below.[1] The axioms must hold for all vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V and for all scalars c and d.

1. The sum of $\mathbf{u}$ and $\mathbf{v}$, denoted by $\mathbf{u} + \mathbf{v}$, is in V.
2. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.
3. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.
4. There is a **zero** vector $\mathbf{0}$ in V such that $\mathbf{u} + \mathbf{0} = \mathbf{u}$.
5. For each $\mathbf{u}$ in V, there is a vector $-\mathbf{u}$ in V such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.

[1] Technically, V is a *real vector space*. All of the theory in this chapter also holds for a *complex vector space* in which the scalars and matrix entries are complex numbers. We will look at this briefly in Chapter 5. Until then, all scalars and matrix entries are assumed to be real.

6. The scalar multiple of $\mathbf{u}$ by c, denoted by $c\mathbf{u}$, is in V.
7. $c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$.
8. $(c + d)\mathbf{u} = c\mathbf{u} + d\mathbf{u}$.
9. $c(d\mathbf{u}) = (cd)\mathbf{u}$.
10. $1\mathbf{u} = \mathbf{u}$.

Using only these axioms, one can show that the zero vector in Axiom 4 is unique, and the vector $-\mathbf{u}$, called the **negative** of $\mathbf{u}$, in Axiom 5 is unique for each $\mathbf{u}$ in V. See Exercises 33 and 34. Proofs of the following simple facts are also outlined in the exercises:

For each $\mathbf{u}$ in V and scalar c,

$$0\mathbf{u} = \mathbf{0} \tag{1}$$

$$c\mathbf{0} = \mathbf{0} \tag{2}$$

$$-\mathbf{u} = (-1)\mathbf{u} \tag{3}$$

EXAMPLE 1 The spaces $\mathbb{R}^n$, where $n \geq 1$, are the premier examples of vector spaces. The geometric intuition developed for $\mathbb{R}^3$ will help you understand and visualize many concepts throughout the chapter. ■

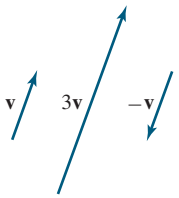


FIGURE 1

EXAMPLE 2 Let V be the set of all arrows (directed line segments) in three-dimensional space, with two arrows regarded as equal if they have the same length and point in the same direction. Define addition by the parallelogram rule (from Section 1.3), and for each $\mathbf{v}$ in V, define $c\mathbf{v}$ to be the arrow whose length is $|c|$ times the length of $\mathbf{v}$, pointing in the same direction as $\mathbf{v}$ if $c \geq 0$ and otherwise pointing in the opposite direction. (See Figure 1.) Show that V is a vector space. This space is a common model in physical problems for various forces.

SOLUTION The definition of V is geometric, using concepts of length and direction. No xyz-coordinate system is involved. An arrow of zero length is a single point and represents the zero vector. The negative of $\mathbf{v}$ is $(-1)\mathbf{v}$. So Axioms 1, 4, 5, 6, and 10 are evident. The rest are verified by geometry. For instance, see Figures 2 and 3. ■

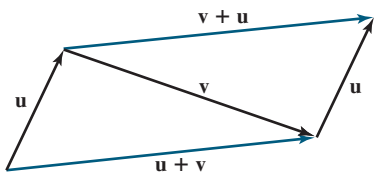


FIGURE 2 $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.

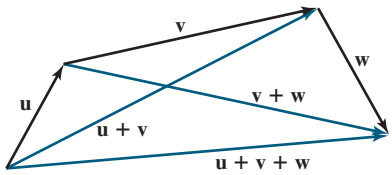


FIGURE 3 $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.

EXAMPLE 3 Let $\mathbb{S}$ be the space of all doubly infinite sequences of numbers (usually written in a row rather than a column):

$$\{y_k\} = (\ldots, y_{-2}, y_{-1}, y_0, y_1, y_2, \ldots)$$

If $\{z_k\}$ is another element of $\mathbb{S}$, then the sum $\{y_k\} + \{z_k\}$ is the sequence $\{y_k + z_k\}$ formed by adding corresponding terms of $\{y_k\}$ and $\{z_k\}$. The scalar multiple $c\{y_k\}$ is the sequence $\{cy_k\}$. The vector space axioms are verified in the same way as for $\mathbb{R}^n$.

Elements of $\mathbb{S}$ arise in engineering, for example, whenever a signal is measured (or sampled) at discrete times. A signal might be electrical, mechanical, optical, biological, audio, and so on. The digital signal processors mentioned in the chapter introduction use discrete (or digital) signals. For convenience, we will call $\mathbb{S}$ the space of (discrete-time) **signals**. A signal may be visualized by a graph as in Figure 4.

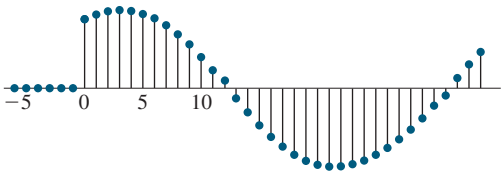


FIGURE 4 A discrete-time signal.

EXAMPLE 4 For $n \geq 0$, the set $\mathbb{P}_n$ of polynomials of degree at most n consists of all polynomials of the form

$$\mathbf{p}(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n \tag{4}$$

where the coefficients $a_0, \ldots, a_n$ and the variable t are real numbers. The *degree* of $\mathbf{p}$ is the highest power of t in (4) whose coefficient is not zero. If $\mathbf{p}(t) = a_0 \neq 0$, the degree of $\mathbf{p}$ is zero. If all the coefficients are zero, $\mathbf{p}$ is called the *zero polynomial*. The zero polynomial is included in $\mathbb{P}_n$ even though its degree, for technical reasons, is not defined.

If $\mathbf{p}$ is given by (4) and if $\mathbf{q}(t) = b_0 + b_1 t + \cdots + b_n t^n$, then the sum $\mathbf{p} + \mathbf{q}$ is defined by

$$\begin{aligned}(\mathbf{p} + \mathbf{q})(t) &= \mathbf{p}(t) + \mathbf{q}(t) \\ &= (a_0 + b_0) + (a_1 + b_1)t + \cdots + (a_n + b_n)t^n\end{aligned}$$

The scalar multiple $c\mathbf{p}$ is the polynomial defined by

$$(c\mathbf{p})(t) = c\mathbf{p}(t) = ca_0 + (ca_1)t + \cdots + (ca_n)t^n$$

These definitions satisfy Axioms 1 and 6 because $\mathbf{p} + \mathbf{q}$ and $c\mathbf{p}$ are polynomials of degree less than or equal to n. Axioms 2, 3, and 7–10 follow from properties of the real numbers. Clearly, the zero polynomial acts as the zero vector in Axiom 4. Finally, $(-1)\mathbf{p}$ acts as the negative of $\mathbf{p}$, so Axiom 5 is satisfied. Thus $\mathbb{P}_n$ is a vector space.

The vector spaces $\mathbb{P}_n$ for various n are used, for instance, in statistical trend analysis of data, discussed in Section 6.8.

EXAMPLE 5 Let V be the set of all real-valued functions defined on a set $\mathbb{D}$. (Typically, $\mathbb{D}$ is the set of real numbers or some interval on the real line.) Functions are added in the usual way: $\mathbf{f} + \mathbf{g}$ is the function whose value at t in the domain $\mathbb{D}$ is $\mathbf{f}(t) + \mathbf{g}(t)$. Likewise, for a scalar c and an $\mathbf{f}$ in V, the scalar multiple $c\mathbf{f}$ is the function whose value at t is $c\mathbf{f}(t)$. For instance, if $\mathbb{D} = \mathbb{R}$, $\mathbf{f}(t) = 1 + \sin 2t$, and $\mathbf{g}(t) = 2 + .5t$, then

$$(\mathbf{f} + \mathbf{g})(t) = 3 + \sin 2t + .5t \quad \text{and} \quad (2\mathbf{g})(t) = 4 + t$$

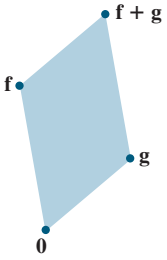


FIGURE 5
The sum of two vectors (functions).

Two functions in V are equal if and only if their values are equal for every t in $\mathbb{D}$. Hence the zero vector in V is the function that is identically zero, $\mathbf{f}(t) = 0$ for all t, and the negative of $\mathbf{f}$ is $(-1)\mathbf{f}$. Axioms 1 and 6 are obviously true, and the other axioms follow from properties of the real numbers, so V is a vector space. ■

It is important to think of each function in the vector space V of Example 5 as a single object, as just one "point" or vector in the vector space. The sum of two vectors $\mathbf{f}$ and $\mathbf{g}$ (functions in V, or elements of *any* vector space) can be visualized as in Figure 5, because this can help you carry over to a general vector space the geometric intuition you have developed while working with the vector space $\mathbb{R}^n$. See the *Study Guide* for help as you learn to adopt this more general point of view.

Subspaces

In many problems, a vector space consists of an appropriate subset of vectors from some larger vector space. In this case, only three of the ten vector space axioms need to be checked; the rest are automatically satisfied.

DEFINITION

A **subspace** of a vector space V is a subset H of V that has three properties:

a. The zero vector of V is in H.[2]

b. H is closed under vector addition. That is, for each $\mathbf{u}$ and $\mathbf{v}$ in H, the sum $\mathbf{u} + \mathbf{v}$ is in H.

c. H is closed under multiplication by scalars. That is, for each $\mathbf{u}$ in H and each scalar c, the vector $c\mathbf{u}$ is in H.

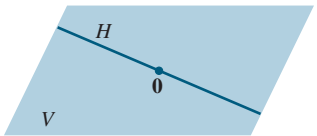


FIGURE 6
A subspace of V.

Properties (a), (b), and (c) guarantee that a subspace H of V is itself a *vector space*, under the vector space operations already defined in V. To verify this, note that properties (a), (b), and (c) are Axioms 1, 4, and 6. Axioms 2, 3, and 7–10 are automatically true in H because they apply to all elements of V, including those in H. Axiom 5 is also true in H, because if $\mathbf{u}$ is in H, then $(-1)\mathbf{u}$ is in H by property (c), and we know from equation (3) earlier in this section that $(-1)\mathbf{u}$ is the vector $-\mathbf{u}$ in Axiom 5.

So every subspace is a vector space. Conversely, every vector space is a subspace (of itself and possibly of other larger spaces). The term *subspace* is used when at least two vector spaces are in mind, with one inside the other, and the phrase *subspace of* V identifies V as the larger space. (See Figure 6.)

EXAMPLE 6 The set consisting of only the zero vector in a vector space V is a subspace of V, called the **zero subspace** and written as $\{\mathbf{0}\}$. ■

EXAMPLE 7 Let $\mathbb{P}$ be the set of all polynomials with real coefficients, with operations in $\mathbb{P}$ defined as for functions. Then $\mathbb{P}$ is a subspace of the space of all real-valued functions defined on $\mathbb{R}$. Also, for each $n \geq 0$, $\mathbb{P}_n$ is a subspace of $\mathbb{P}$, because $\mathbb{P}_n$ is a subset of $\mathbb{P}$ that contains the zero polynomial, the sum of two polynomials in $\mathbb{P}_n$ is also in $\mathbb{P}_n$, and a scalar multiple of a polynomial in $\mathbb{P}_n$ is also in $\mathbb{P}_n$. ■

[2] Some texts replace property (a) in this definition by the assumption that H is nonempty. Then (a) could be deduced from (c) and the fact that $0\mathbf{u} = \mathbf{0}$. But the best way to test for a subspace is to look first for the zero vector. If $\mathbf{0}$ is in H, then properties (b) and (c) must be checked. If $\mathbf{0}$ is *not* in H, then H cannot be a subspace and the other properties need not be checked.

EXAMPLE 8 The set of finitely supported signals $\mathbb{S}_f$ consists of the signals $\{y_k\}$, where only finitely many of the y_k are nonzero. Since the zero signal $\mathbf{0} = (\ldots, 0, 0, 0, \ldots)$ has no nonzero entries, it is clearly an element of $\mathbb{S}_f$. If two signals with finitely many nonzeros are added, the resulting signal will have finitely many nonzeros. Similarly if a signal with finitely many nonzeros is scaled, the result will still have finitely many nonzeros. Thus $\mathbb{S}_f$ is a subspace of $\mathbb{S}$, the discrete-time signals. See Figure 7.

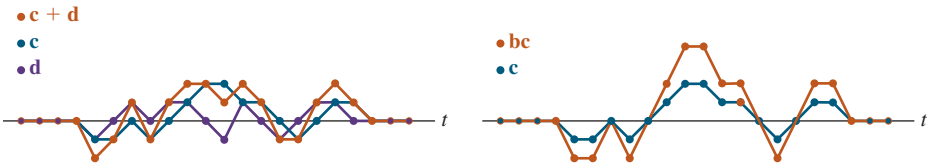


FIGURE 7

EXAMPLE 9 The vector space $\mathbb{R}^2$ is *not* a subspace of $\mathbb{R}^3$ because $\mathbb{R}^2$ is not even a subset of $\mathbb{R}^3$. (The vectors in $\mathbb{R}^3$ all have three entries, whereas the vectors in $\mathbb{R}^2$ have only two.) The set

$$H = \left\{ \begin{bmatrix} s \\ t \\ 0 \end{bmatrix} : s \text{ and } t \text{ are real} \right\}$$

is a subset of $\mathbb{R}^3$ that "looks" and "acts" like $\mathbb{R}^2$, although it is logically distinct from $\mathbb{R}^2$. See Figure 8. Show that H is a subspace of $\mathbb{R}^3$.

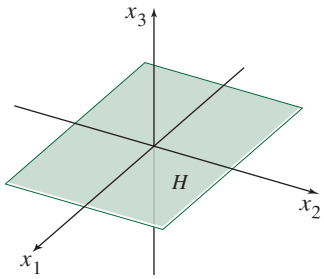


FIGURE 8
The x_1x_2-plane as a subspace of $\mathbb{R}^3$.

SOLUTION The zero vector is in H, and H is closed under vector addition and scalar multiplication because these operations on vectors in H always produce vectors whose third entries are zero (and so belong to H). Thus H is a subspace of $\mathbb{R}^3$. ■

EXAMPLE 10 A plane in $\mathbb{R}^3$ *not* through the origin is not a subspace of $\mathbb{R}^3$, because the plane does not contain the zero vector of $\mathbb{R}^3$. Similarly, a line in $\mathbb{R}^2$ *not* through the origin, such as in Figure 9, is *not* a subspace of $\mathbb{R}^2$. ■

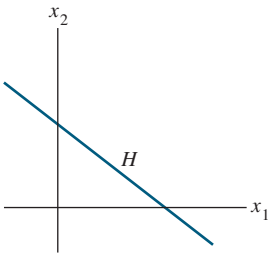


FIGURE 9
A line that is not a vector space.

A Subspace Spanned by a Set

The next example illustrates one of the most common ways of describing a subspace. As in Chapter 1, the term **linear combination** refers to any sum of scalar multiples of vectors, and Span $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ denotes the set of all vectors that can be written as linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_p$.

EXAMPLE 11 Given $\mathbf{v}_1$ and $\mathbf{v}_2$ in a vector space V, let $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$. Show that H is a subspace of V.

SOLUTION The zero vector is in H, since $\mathbf{0} = 0\mathbf{v}_1 + 0\mathbf{v}_2$. To show that H is closed under vector addition, take two arbitrary vectors in H, say,

$$\mathbf{u} = s_1\mathbf{v}_1 + s_2\mathbf{v}_2 \quad \text{and} \quad \mathbf{w} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2$$

By Axioms 2, 3, and 8 for the vector space V,

$$\begin{aligned} \mathbf{u} + \mathbf{w} &= (s_1\mathbf{v}_1 + s_2\mathbf{v}_2) + (t_1\mathbf{v}_1 + t_2\mathbf{v}_2) \\ &= (s_1 + t_1)\mathbf{v}_1 + (s_2 + t_2)\mathbf{v}_2 \end{aligned}$$

So $\mathbf{u} + \mathbf{w}$ is in H. Furthermore, if c is any scalar, then by Axioms 7 and 9,

$$c\mathbf{u} = c(s_1\mathbf{v}_1 + s_2\mathbf{v}_2) = (cs_1)\mathbf{v}_1 + (cs_2)\mathbf{v}_2$$

which shows that $c\mathbf{u}$ is in H and H is closed under scalar multiplication. Thus H is a subspace of V. ■

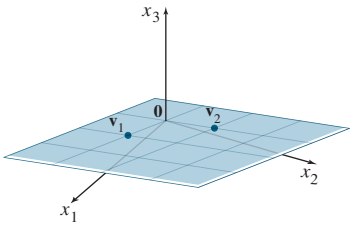


FIGURE 10
An example of a subspace.

In Section 4.5, you will see that every nonzero subspace of $\mathbb{R}^3$, other than $\mathbb{R}^3$ itself, is either Span $\{\mathbf{v}_1, \mathbf{v}_2\}$ for some linearly independent $\mathbf{v}_1$ and $\mathbf{v}_2$ or Span $\{\mathbf{v}\}$ for $\mathbf{v} \neq \mathbf{0}$. In the first case, the subspace is a plane through the origin; in the second case, it is a line through the origin. (See Figure 10.) It is helpful to keep these geometric pictures in mind, even for an abstract vector space.

The argument in Example 11 can easily be generalized to prove the following theorem.

THEOREM 1

If $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are in a vector space V, then Span $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is a subspace of V.

We call Span $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ **the subspace spanned** (or **generated**) by $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$. Given any subspace H of V, a **spanning** (or **generating**) **set** for H is a set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ in H such that $H =$ Span $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$.

The next example shows how to use Theorem 1.

EXAMPLE 12 Let H be the set of all vectors of the form $(a - 3b, b - a, a, b)$, where a and b are arbitrary scalars. That is, let $H = \{(a - 3b, b - a, a, b) : a \text{ and } b \text{ in } \mathbb{R}\}$. Show that H is a subspace of $\mathbb{R}^4$.

SOLUTION Write the vectors in H as column vectors. Then an arbitrary vector in H has the form

$$\begin{bmatrix} a - 3b \\ b - a \\ a \\ b \end{bmatrix} = a\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix} + b\begin{bmatrix} -3 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

$\mathbf{v}_1$ $\mathbf{v}_2$

This calculation shows that $H =$ Span $\{\mathbf{v}_1, \mathbf{v}_2\}$, where $\mathbf{v}_1$ and $\mathbf{v}_2$ are the vectors indicated above. Thus H is a subspace of $\mathbb{R}^4$ by Theorem 1. ■

Example 12 illustrates a useful technique of expressing a subspace H as the set of linear combinations of some small collection of vectors. If $H =$ Span $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$, we can think of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_p$ in the spanning set as "handles" that allow us to hold on to the subspace H. Calculations with the infinitely many vectors in H are often reduced to operations with the finite number of vectors in the spanning set.

EXAMPLE 13 For what value(s) of h will $\mathbf{y}$ be in the subspace of $\mathbb{R}^3$ spanned by $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, if

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 5 \\ -4 \\ -7 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -3 \\ 1 \\ 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} -4 \\ 3 \\ h \end{bmatrix}$$

SOLUTION This question is Practice Problem 2 in Section 1.3, written here with the term *subspace* rather than Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. The solution there shows that $\mathbf{y}$ is in Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ if and only if $h = 5$. That solution is worth reviewing now, along with Exercises 11–16 and 19–21 in Section 1.3. ■

Although many vector spaces in this chapter will be subspaces of $\mathbb{R}^n$, it is important to keep in mind that the abstract theory applies to other vector spaces as well. Vector spaces of functions arise in many applications, and they will receive more attention later.

Practice Problems

1. Show that the set H of all points in $\mathbb{R}^2$ of the form $(3s, 2 + 5s)$ is not a vector space, by showing that it is not closed under scalar multiplication. (Find a specific vector $\mathbf{u}$ in H and a scalar c such that $c\mathbf{u}$ is not in H.)
2. Let $W = \text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$, where $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are in a vector space V. Show that $\mathbf{v}_k$ is in W for $1 \le k \le p$. [*Hint:* First write an equation that shows that $\mathbf{v}_1$ is in W. Then adjust your notation for the general case.]
3. An $n \times n$ matrix A is said to be symmetric if $A^T = A$. Let S be the set of all 3×3 symmetric matrices. Show that S is a subspace of $M_{3\times 3}$, the vector space of 3×3 matrices.

4.1 Exercises

1. Let V be the first quadrant in the xy-plane; that is, let $V = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \ge 0, y \ge 0 \right\}$
   a. If $\mathbf{u}$ and $\mathbf{v}$ are in V, is $\mathbf{u} + \mathbf{v}$ in V? Why?
   b. Find a specific vector $\mathbf{u}$ in V and a specific scalar c such that $c\mathbf{u}$ is *not* in V. (This is enough to show that V is *not* a vector space.)
2. Let W be the union of the first and third quadrants in the xy-plane. That is, let $W = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : xy \ge 0 \right\}$.
   a. If $\mathbf{u}$ is in W and c is any scalar, is $c\mathbf{u}$ in W? Why?
   b. Find specific vectors $\mathbf{u}$ and $\mathbf{v}$ in W such that $\mathbf{u} + \mathbf{v}$ is not in W. (This is enough to show that W is *not* a vector space.)
3. Let H be the set of points inside and on the unit circle in the xy-plane. That is, let $H = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x^2 + y^2 \le 1 \right\}$. Find a specific example—two vectors or a vector and a scalar—to show that H is not a subspace of $\mathbb{R}^2$.
4. Construct a geometric figure that illustrates why a line in $\mathbb{R}^2$ *not* through the origin is not closed under vector addition.

In Exercises 5–8, determine if the given set is a subspace of $\mathbb{P}_n$ for an appropriate value of n. Justify your answers.

5. All polynomials of the form $\mathbf{p}(t) = at^2$, where a is in $\mathbb{R}$.
6. All polynomials of the form $\mathbf{p}(t) = a + t^2$, where a is in $\mathbb{R}$.
7. All polynomials of degree at most 3, with integers as coefficients.
8. All polynomials in $\mathbb{P}_n$ such that $\mathbf{p}(0) = 0$.
9. Let H be the set of all vectors of the form $\begin{bmatrix} s \\ 3s \\ 2s \end{bmatrix}$. Find a vector $\mathbf{v}$ in $\mathbb{R}^3$ such that $H = \text{Span}\{\mathbf{v}\}$. Why does this show that H is a subspace of $\mathbb{R}^3$?
10. Let H be the set of all vectors of the form $\begin{bmatrix} 2t \\ 0 \\ -t \end{bmatrix}$. Show that H is a subspace of $\mathbb{R}^3$. (Use the method of Exercise 9.)
11. Let W be the set of all vectors of the form $\begin{bmatrix} 6b + 7c \\ b \\ c \end{bmatrix}$, where b and c are arbitrary. Find vectors $\mathbf{u}$ and $\mathbf{v}$ such that $W = \text{Span}\,\mathbf{u}, \mathbf{v}$. Why does this show that W is a subspace of $\mathbb{R}^3$?
12. Let W be the set of all vectors of the form $\begin{bmatrix} s + 3t \\ s - t \\ 2s - t \\ 4t \end{bmatrix}$. Show that W is a subspace of $\mathbb{R}^4$. (Use the method of Exercise 11.)
13. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 4 \\ 2 \\ 6 \end{bmatrix}$, and $\mathbf{w} = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$.

a. Is $\mathbf{w}$ in $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$? How many vectors are in $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$?

b. How many vectors are in Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$?

c. Is $\mathbf{w}$ in the subspace spanned by $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$? Why?

14. Let $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ be as in Exercise 13, and let $\mathbf{w} = \begin{bmatrix} 8 \\ 4 \\ 7 \end{bmatrix}$. Is $\mathbf{w}$ in the subspace spanned by $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$? Why?

In Exercises 15–18, let W be the set of all vectors of the form shown, where a, b, and c represent arbitrary real numbers. In each case, either find a set S of vectors that spans W or give an example to show that W is *not* a vector space.

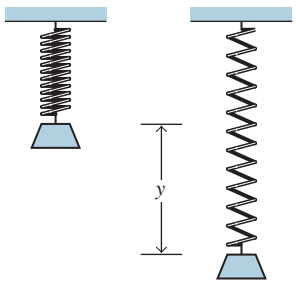
15. $\begin{bmatrix} 3a+b \\ 4 \\ a-5b \end{bmatrix}$ **16.** $\begin{bmatrix} -a+1 \\ a-6b \\ 2b+a \end{bmatrix}$

17. $\begin{bmatrix} a-b \\ b-c \\ c-a \\ b \end{bmatrix}$ **18.** $\begin{bmatrix} 4a+3b \\ 0 \\ a+b+c \\ c-2a \end{bmatrix}$

19. If a mass m is placed at the end of a spring, and if the mass is pulled downward and released, the mass–spring system will begin to oscillate. The displacement y of the mass from its resting position is given by a function of the form

$$y(t) = c_1 \cos \omega t + c_2 \sin \omega t \tag{5}$$

where ω is a constant that depends on the spring and the mass. (See the figure below.) Show that the set of all functions described in (5) (with ω fixed and c_1, c_2 arbitrary) is a vector space.



20. The set of all continuous real-valued functions defined on a closed interval $[a, b]$ in $\mathbb{R}$ is denoted by $C[a, b]$. This set is a subspace of the vector space of all real-valued functions defined on $[a, b]$.

a. What facts about continuous functions should be proved in order to demonstrate that $C[a, b]$ is indeed a subspace as claimed? (These facts are usually discussed in a calculus class.)

b. Show that $\{\mathbf{f} \text{ in } C[a, b] : \mathbf{f}(a) = \mathbf{f}(b)\}$ is a subspace of $C[a, b]$.

For fixed positive integers m and n, the set $M_{m\times n}$ of all $m \times n$ matrices is a vector space, under the usual operations of addition of matrices and multiplication by real scalars.

21. Determine if the set H of all matrices of the form $\begin{bmatrix} a & b \\ 0 & d \end{bmatrix}$ is a subspace of $M_{2\times 2}$.

22. Let F be a fixed 3×2 matrix, and let H be the set of all matrices A in $M_{2\times 4}$ with the property that $FA = 0$ (the zero matrix in $M_{3\times 4}$). Determine if H is a subspace of $M_{2\times 4}$.

In Exercises 23–32, mark each statement True or False **(T/F)**. Justify each answer.

23. **(T/F)** If $\mathbf{f}$ is a function in the vector space V of all real-valued functions on $\mathbb{R}$ and if $\mathbf{f}(t) = 0$ for some t, then $\mathbf{f}$ is the zero vector in V.

24. **(T/F)** A vector is any element of a vector space.

25. **(T/F)** An arrow in three-dimensional space can be considered to be a vector.

26. **(T/F)** If $\mathbf{u}$ is a vector in a vector space V, then $(-1)\,\mathbf{u}$ is the same as the negative of $\mathbf{u}$.

27. **(T/F)** A subset H of a vector space V is a subspace of V if the zero vector is in H.

28. **(T/F)** A vector space is also a subspace.

29. **(T/F)** A subspace is also a vector space.

30. **(T/F)** $\mathbb{R}^2$ is a subspace of $\mathbb{R}^3$.

31. **(T/F)** The polynomials of degree two or less are a subspace of the polynomials of degree three or less.

32. **(T/F)** A subset H of a vector space V is a subspace of V if the following conditions are satisfied: (i) the zero vector of V is in H, (ii) $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{u} + \mathbf{v}$ are in H, and (iii) c is a scalar and $c\mathbf{u}$ is in H.

Exercises 33–36 show how the axioms for a vector space V can be used to prove the elementary properties described after the definition of a vector space. Fill in the blanks with the appropriate axiom numbers. Because of Axiom 2, Axioms 4 and 5 imply, respectively, that $\mathbf{0} + \mathbf{u} = \mathbf{u}$ and $-\mathbf{u} + \mathbf{u} = \mathbf{0}$ for all $\mathbf{u}$.

33. Complete the following proof that the zero vector is unique. Suppose that $\mathbf{w}$ in V has the property that $\mathbf{u} + \mathbf{w} = \mathbf{w} + \mathbf{u} = \mathbf{u}$ for all $\mathbf{u}$ in V. In particular, $\mathbf{0} + \mathbf{w} = \mathbf{0}$. But $\mathbf{0} + \mathbf{w} = \mathbf{w}$, by Axiom _____. Hence $\mathbf{w} = \mathbf{0} + \mathbf{w} = \mathbf{0}$.

34. Complete the following proof that $-\mathbf{u}$ is the *unique vector* in V such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$. Suppose that $\mathbf{w}$ satisfies $\mathbf{u} + \mathbf{w} = \mathbf{0}$. Adding $-\mathbf{u}$ to both sides, we have

$$(-\mathbf{u}) + [\mathbf{u} + \mathbf{w}] = (-\mathbf{u}) + \mathbf{0}$$
$$[(-\mathbf{u}) + \mathbf{u}] + \mathbf{w} = (-\mathbf{u}) + \mathbf{0} \qquad \text{by Axiom ____ (a)}$$
$$\mathbf{0} + \mathbf{w} = (-\mathbf{u}) + \mathbf{0} \qquad \text{by Axiom ____ (b)}$$
$$\mathbf{w} = -\mathbf{u} \qquad \text{by Axiom ____ (c)}$$

35. Fill in the missing axiom numbers in the following proof that $0\mathbf{u} = \mathbf{0}$ for every $\mathbf{u}$ in V.

$$0\mathbf{u} = (0 + 0)\mathbf{u} = 0\mathbf{u} + 0\mathbf{u} \qquad \text{by Axiom ____ (a)}$$

Add the negative of $0\mathbf{u}$ to both sides:

$0\mathbf{u} + (-0\mathbf{u}) = [0\mathbf{u} + 0\mathbf{u}] + (-0\mathbf{u})$

$0\mathbf{u} + (-0\mathbf{u}) = 0\mathbf{u} + [0\mathbf{u} + (-0\mathbf{u})]$ by Axiom _____ (b)

$\mathbf{0} = 0\mathbf{u} + \mathbf{0}$ by Axiom _____ (c)

$\mathbf{0} = 0\mathbf{u}$ by Axiom _____ (d)

36. Fill in the missing axiom numbers in the following proof that $c\mathbf{0} = \mathbf{0}$ for every scalar c.

$c\mathbf{0} = c(\mathbf{0} + \mathbf{0})$ by Axiom _____ (a)

$= c\mathbf{0} + c\mathbf{0}$ by Axiom _____ (b)

Add the negative of $c\mathbf{0}$ to both sides:

$c\mathbf{0} + (-c\mathbf{0}) = [c\mathbf{0} + c\mathbf{0}] + (-c\mathbf{0})$

$c\mathbf{0} + (-c\mathbf{0}) = c\mathbf{0} + [c\mathbf{0} + (-c\mathbf{0})]$ by Axiom _____ (c)

$\mathbf{0} = c\mathbf{0} + \mathbf{0}$ by Axiom _____ (d)

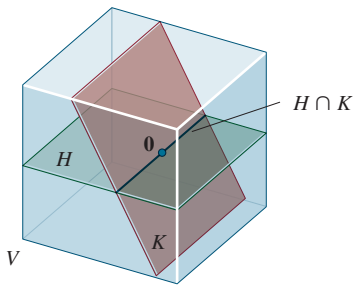
$\mathbf{0} = c\mathbf{0}$ by Axiom _____ (e)

37. Prove that $(-1)\mathbf{u} = -\mathbf{u}$. [*Hint:* Show that $\mathbf{u} + (-1)\mathbf{u} = \mathbf{0}$. Use some axioms and the results of Exercises 34 and 35.]

38. Suppose $c\mathbf{u} = \mathbf{0}$ for some nonzero scalar c. Show that $\mathbf{u} = \mathbf{0}$. Mention the axioms or properties you use.

39. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in a vector space V, and let H be any subspace of V that contains both $\mathbf{u}$ and $\mathbf{v}$. Explain why H also contains Span $\{\mathbf{u}, \mathbf{v}\}$. This shows that Span $\{\mathbf{u}, \mathbf{v}\}$ is the smallest subspace of V that contains both $\mathbf{u}$ and $\mathbf{v}$.

40. Let H and K be subspaces of a vector space V. The **intersection** of H and K, written as $H \cap K$, is the set of $\mathbf{v}$ in V that belong to both H and K. Show that $H \cap K$ is a subspace of V. (See the figure.) Give an example in $\mathbb{R}^2$ to show that the union of two subspaces is not, in general, a subspace.



41. Given subspaces H and K of a vector space V, the **sum** of H and K, written as $H + K$, is the set of all vectors in V that can be written as the sum of two vectors, one in H and the other in K; that is,

$H + K = \{\mathbf{w} : \mathbf{w} = \mathbf{u} + \mathbf{v}$ for some $\mathbf{u}$ in H and some $\mathbf{v}$ in $K\}$

a. Show that $H + K$ is a subspace of V.

b. Show that H is a subspace of $H + K$ and K is a subspace of $H + K$.

42. Suppose $\mathbf{u}_1, \ldots, \mathbf{u}_p$ and $\mathbf{v}_1, \ldots, \mathbf{v}_q$ are vectors in a vector space V, and let

$H = \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ and $K = \text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_q\}$

Show that $H + K = \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_p, \mathbf{v}_1, \ldots, \mathbf{v}_q\}$.

T 43. Show that $\mathbf{w}$ is in the subspace of $\mathbb{R}^4$ spanned by $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, where

$$\mathbf{w} = \begin{bmatrix} 6 \\ -7 \\ 8 \\ -9 \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} 7 \\ -6 \\ -5 \\ 4 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} -3 \\ 2 \\ -1 \\ -4 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} -2 \\ 1 \\ 2 \\ -5 \end{bmatrix}$$

T 44. Determine if $\mathbf{y}$ is in the subspace of $\mathbb{R}^4$ spanned by the columns of A, where

$$\mathbf{y} = \begin{bmatrix} -4 \\ -8 \\ 6 \\ -5 \end{bmatrix}, \quad A = \begin{bmatrix} 3 & -5 & -9 \\ 8 & 7 & -6 \\ -5 & -8 & 3 \\ 2 & -2 & -9 \end{bmatrix}$$

T 45. The vector space $H = \text{Span}\{1, \cos^2 t, \cos^4 t, \cos^6 t\}$ contains at least two interesting functions that will be used in a later exercise:

$\mathbf{f}(t) = 1 - 8\cos^2 t + 8\cos^4 t$

$\mathbf{g}(t) = -1 + 18\cos^2 t - 48\cos^4 t + 32\cos^6 t$

Study the graph of $\mathbf{f}$ for $0 \le t \le 2\pi$, and guess a simple formula for $\mathbf{f}(t)$. Verify your conjecture by graphing the difference between $1 + \mathbf{f}(t)$ and your formula for $\mathbf{f}(t)$. (Hopefully, you will see the constant function 1.) Repeat for $\mathbf{g}$.

T 46. Repeat Exercise 45 for the functions

$\mathbf{f}(t) = 3\sin t - 4\sin^3 t$

$\mathbf{g}(t) = 1 - 8\sin^2 t + 8\sin^4 t$

$\mathbf{h}(t) = 5\sin t - 20\sin^3 t + 16\sin^5 t$

in the vector space Span $\{1, \sin t, \sin^2 t, \ldots, \sin^5 t\}$.

Solutions to Practice Problems

1. Take any $\mathbf{u}$ in H—say, $\mathbf{u} = \begin{bmatrix} 3 \\ 7 \end{bmatrix}$—and take any $c \neq 1$—say, $c = 2$. Then $c\mathbf{u} = \begin{bmatrix} 6 \\ 14 \end{bmatrix}$. If this is in H, then there is some s such that

$$\begin{bmatrix} 3s \\ 2 + 5s \end{bmatrix} = \begin{bmatrix} 6 \\ 14 \end{bmatrix}$$

That is, $s = 2$ and $s = 12/5$, which is impossible. So $2\mathbf{u}$ is not in H and H is not a vector space.

2. $\mathbf{v}_1 = 1\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_p$. This expresses $\mathbf{v}_1$ as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_p$, so $\mathbf{v}_1$ is in W. In general, $\mathbf{v}_k$ is in W because

$$\mathbf{v}_k = 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_{k-1} + 1\mathbf{v}_k + 0\mathbf{v}_{k+1} + \cdots + 0\mathbf{v}_p$$

3. The subset S is a subspace of $M_{3\times 3}$ since it satisfies all three of the requirements listed in the definition of a subspace:
   a. Observe that the $\mathbf{0}$ in $M_{3\times 3}$ is the 3×3 zero matrix and since $\mathbf{0}^T = \mathbf{0}$, the matrix $\mathbf{0}$ is symmetric and hence $\mathbf{0}$ is in S.
   b. Let A and B in S. Notice that A and B are 3×3 symmetric matrices so $A^T = A$ and $B^T = B$. By the properties of transposes of matrices, $(A + B)^T = A^T + B^T = A + B$. Thus $A + B$ is symmetric and hence $A + B$ is in S.
   c. Let A be in S and let c be a scalar. Since A is symmetric, by the properties of symmetric matrices, $(cA)^T = c(A^T) = cA$. Thus cA is also a symmetric matrix and hence cA is in S.

4.2 Null Spaces, Column Spaces, Row Spaces, and Linear Transformations

In applications of linear algebra, subspaces of $\mathbb{R}^n$ usually arise in one of two ways: (1) as the set of all solutions to a system of homogeneous linear equations or (2) as the set of all linear combinations of certain specified vectors. In this section, we compare and contrast these two descriptions of subspaces, allowing us to practice using the concept of a subspace. Actually, as you will soon discover, we have been working with subspaces ever since Section 1.3. The main new feature here is the terminology. The section concludes with a discussion of the kernel and range of a linear transformation.

The Null Space of a Matrix

Consider the following system of homogeneous equations:

$$\begin{aligned} x_1 - 3x_2 - 2x_3 &= 0 \\ -5x_1 + 9x_2 + x_3 &= 0 \end{aligned} \tag{1}$$

In matrix form, this system is written as $A\mathbf{x} = \mathbf{0}$, where

$$A = \begin{bmatrix} 1 & -3 & -2 \\ -5 & 9 & 1 \end{bmatrix} \tag{2}$$

Recall that the set of all $\mathbf{x}$ that satisfy (1) is called the **solution set** of the system (1). Often it is convenient to relate this set directly to the matrix A and the equation $A\mathbf{x} = \mathbf{0}$. We call the set of $\mathbf{x}$ that satisfy $A\mathbf{x} = \mathbf{0}$ the **null space** of the matrix A.

DEFINITION

The **null space** of an $m \times n$ matrix A, written as Nul A, is the set of all solutions of the homogeneous equation $A\mathbf{x} = \mathbf{0}$. In set notation,

$$\text{Nul } A = \{\mathbf{x} : \mathbf{x} \text{ is in } \mathbb{R}^n \text{ and } A\mathbf{x} = \mathbf{0}\}$$

A more dynamic description of Nul A is the set of all $\mathbf{x}$ in $\mathbb{R}^n$ that are mapped into the zero vector of $\mathbb{R}^m$ via the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$. See Figure 1.

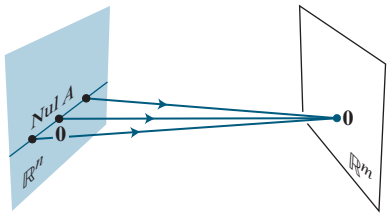


FIGURE 1

EXAMPLE 1 Let A be the matrix in (2), and let $\mathbf{u} = \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix}$. Determine if $\mathbf{u}$ belongs to the null space of A.

SOLUTION To test if $\mathbf{u}$ satisfies $A\mathbf{u} = \mathbf{0}$, simply compute

$$A\mathbf{u} = \begin{bmatrix} 1 & -3 & -2 \\ -5 & 9 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ -2 \end{bmatrix} = \begin{bmatrix} 5 - 9 + 4 \\ -25 + 27 - 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Thus $\mathbf{u}$ is in Nul A. ■

The term *space* in *null space* is appropriate because the null space of a matrix is a vector space, as shown in the next theorem.

THEOREM 2

The null space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^n$. Equivalently, the set of all solutions to a system $A\mathbf{x} = \mathbf{0}$ of m homogeneous linear equations in n unknowns is a subspace of $\mathbb{R}^n$.

PROOF Certainly Nul A is a subset of $\mathbb{R}^n$ because A has n columns. We must show that Nul A satisfies the three properties of a subspace. Of course, $\mathbf{0}$ is in Nul A. Next, let $\mathbf{u}$ and $\mathbf{v}$ represent any two vectors in Nul A. Then

$$A\mathbf{u} = \mathbf{0} \quad \text{and} \quad A\mathbf{v} = \mathbf{0}$$

To show that $\mathbf{u} + \mathbf{v}$ is in Nul A, we must show that $A(\mathbf{u} + \mathbf{v}) = \mathbf{0}$. Using a property of matrix multiplication, compute

$$A(\mathbf{u} + \mathbf{v}) = A\mathbf{u} + A\mathbf{v} = \mathbf{0} + \mathbf{0} = \mathbf{0}$$

Thus $\mathbf{u} + \mathbf{v}$ is in Nul A, and Nul A is closed under vector addition. Finally, if c is any scalar, then

$$A(c\mathbf{u}) = c(A\mathbf{u}) = c(\mathbf{0}) = \mathbf{0}$$

which shows that $c\mathbf{u}$ is in Nul A. Thus Nul A is a subspace of $\mathbb{R}^n$. ■

EXAMPLE 2 Let H be the set of all vectors in $\mathbb{R}^4$ whose coordinates a, b, c, d satisfy the equations $a - 2b + 5c = d$ and $c - a = b$. Show that H is a subspace of $\mathbb{R}^4$.

SOLUTION Rearrange the equations that describe the elements of H, and note that H is the set of all solutions of the following system of homogeneous linear equations:

$$\begin{aligned} a - 2b + 5c - d &= 0 \\ -a - b + c &= 0 \end{aligned}$$

By Theorem 2, H is a subspace of $\mathbb{R}^4$. ■

It is important that the linear equations defining the set H are homogeneous. Otherwise, the set of solutions will definitely *not* be a subspace (because the zero vector is not a solution of a nonhomogeneous system). Also, in some cases, the set of solutions could be empty.

An Explicit Description of Nul A

There is no obvious relation between vectors in Nul A and the entries in A. We say that Nul A is defined *implicitly*, because it is defined by a condition that must be checked. No explicit list or description of the elements in Nul A is given. However, *solving* the equation $A\mathbf{x} = \mathbf{0}$ amounts to producing an *explicit* description of Nul A. The next example reviews the procedure from Section 1.5.

EXAMPLE 3 Find a spanning set for the null space of the matrix

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$$

SOLUTION The first step is to find the general solution of $A\mathbf{x} = \mathbf{0}$ in terms of free variables. Row reduce the augmented matrix $[\,A \quad \mathbf{0}\,]$ to *reduced* echelon form in order to write the basic variables in terms of the free variables:

$$\begin{bmatrix} 1 & -2 & 0 & -1 & 3 & 0 \\ 0 & 0 & 1 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \begin{aligned} x_1 - 2x_2 \quad - x_4 + 3x_5 &= 0 \\ x_3 + 2x_4 - 2x_5 &= 0 \\ 0 &= 0 \end{aligned}$$

The general solution is $x_1 = 2x_2 + x_4 - 3x_5$, $x_3 = -2x_4 + 2x_5$, with x_2, x_4, and x_5 free. Next, decompose the vector giving the general solution into a linear combination of vectors where *the weights are the free variables*. That is,

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2x_2 + x_4 - 3x_5 \\ x_2 \\ -2x_4 + 2x_5 \\ x_4 \\ x_5 \end{bmatrix} = x_2 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 1 \\ 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -3 \\ 0 \\ 2 \\ 0 \\ 1 \end{bmatrix}$$

$$\uparrow \mathbf{u} \qquad \uparrow \mathbf{v} \qquad \uparrow \mathbf{w}$$

$$= x_2\mathbf{u} + x_4\mathbf{v} + x_5\mathbf{w} \tag{3}$$

Every linear combination of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ is an element of Nul A and vice versa. Thus $\{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is a spanning set for Nul A. ■

Two points should be made about the solution of Example 3 that apply to all problems of this type where Nul A contains nonzero vectors. We will use these facts later.

1. The spanning set produced by the method in Example 3 is automatically linearly independent because the free variables are the weights on the spanning vectors. For instance, look at the 2nd, 4th, and 5th entries in the solution vector in (3) and note that $x_2\mathbf{u} + x_4\mathbf{v} + x_5\mathbf{w}$ can be $\mathbf{0}$ only if the weights x_2, x_4, and x_5 are all zero.
2. When Nul A contains nonzero vectors, the number of vectors in the spanning set for Nul A equals the number of free variables in the equation $A\mathbf{x} = \mathbf{0}$.

The Column Space of a Matrix

Another important subspace associated with a matrix is its column space. Unlike the null space, the column space is defined explicitly via linear combinations.

DEFINITION

The **column space** of an $m \times n$ matrix A, written as Col A, is the set of all linear combinations of the columns of A. If $A = [\,\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_n\,]$, then

$$\text{Col } A = \text{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$$

Since Span $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is a subspace, by Theorem 1, the next theorem follows from the definition of Col A and the fact that the columns of A are in $\mathbb{R}^m$.

THEOREM 3

The column space of an $m \times n$ matrix A is a subspace of $\mathbb{R}^m$.

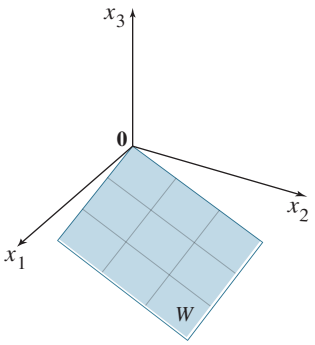
Note that a typical vector in Col A can be written as $A\mathbf{x}$ for some $\mathbf{x}$ because the notation $A\mathbf{x}$ stands for a linear combination of the columns of A. That is,

$$\text{Col } A = \{\mathbf{b} : \mathbf{b} = A\mathbf{x} \text{ for some } \mathbf{x} \text{ in } \mathbb{R}^n\}$$

The notation $A\mathbf{x}$ for vectors in Col A also shows that Col A is the *range* of the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$. We will return to this point of view at the end of the section.

EXAMPLE 4 Find a matrix A such that $W = \text{Col } A$.

$$W = \left\{ \begin{bmatrix} 6a - b \\ a + b \\ -7a \end{bmatrix} : a, b \text{ in } \mathbb{R} \right\}$$



SOLUTION First, write W as a set of linear combinations.

$$W = \left\{ a \begin{bmatrix} 6 \\ 1 \\ -7 \end{bmatrix} + b \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} : a, b \text{ in } \mathbb{R} \right\} = \text{Span} \left\{ \begin{bmatrix} 6 \\ 1 \\ -7 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

Second, use the vectors in the spanning set as the columns of A. Let $A = \begin{bmatrix} 6 & -1 \\ 1 & 1 \\ -7 & 0 \end{bmatrix}$.

Then $W = \text{Col } A$, as desired. ■

Recall from Theorem 4 in Section 1.4 that the columns of A span $\mathbb{R}^m$ if and only if the equation $A\mathbf{x} = \mathbf{b}$ has a solution for each $\mathbf{b}$. We can restate this fact as follows:

The column space of an $m \times n$ matrix A is all of $\mathbb{R}^m$ if and only if the equation $A\mathbf{x} = \mathbf{b}$ has a solution for each $\mathbf{b}$ in $\mathbb{R}^m$.

The Row Space

If A is an $m \times n$ matrix, each row of A has n entries and thus can be identified with a vector in $\mathbb{R}^n$. The set of all linear combinations of the row vectors is called the **row space** of A and is denoted by Row A. Each row has n entries, so Row A is a subspace of $\mathbb{R}^n$. Since the rows of A are identified with the columns of A^T, we could also write Col A^T in place of Row A.

EXAMPLE 5 Let

$$A = \begin{bmatrix} -2 & -5 & 8 & 0 & -17 \\ 1 & 3 & -5 & 1 & 5 \\ 3 & 11 & -19 & 7 & 1 \\ 1 & 7 & -13 & 5 & -3 \end{bmatrix} \quad \text{and} \quad \begin{aligned} \mathbf{r}_1 &= (-2, -5, 8, 0, -17) \\ \mathbf{r}_2 &= (1, 3, -5, 1, 5) \\ \mathbf{r}_3 &= (3, 11, -19, 7, 1) \\ \mathbf{r}_4 &= (1, 7, -13, 5, -3) \end{aligned}$$

The row space of A is the subspace of $\mathbb{R}^5$ spanned by $\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4\}$. That is, Row A = Span $\{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4\}$. It is natural to write row vectors horizontally; however, they may also be written as column vectors if that is more convenient. ■

The Contrast Between Nul A and Col A

It is natural to wonder how the null space and column space of a matrix are related. In fact, the two spaces are quite dissimilar, as Examples 6–8 will show. Nevertheless, a surprising connection between the null space and column space will emerge in Section 4.5, after more theory is available.

EXAMPLE 6 Let

$$A = \begin{bmatrix} 2 & 4 & -2 & 1 \\ -2 & -5 & 7 & 3 \\ 3 & 7 & -8 & 6 \end{bmatrix}$$

a. If the column space of A is a subspace of $\mathbb{R}^k$, what is k?

b. If the null space of A is a subspace of $\mathbb{R}^k$, what is k?

SOLUTION

a. The columns of A each have three entries, so Col A is a subspace of $\mathbb{R}^k$, where $k = 3$.

b. A vector $\mathbf{x}$ such that $A\mathbf{x}$ is defined must have four entries, so Nul A is a subspace of $\mathbb{R}^k$, where $k = 4$. ■

When a matrix is not square, as in Example 6, the vectors in Nul A and Col A live in entirely different "universes." For example, no linear combination of vectors in $\mathbb{R}^3$ can produce a vector in $\mathbb{R}^4$. When A is square, Nul A and Col A do have the zero vector in common, and in special cases it is possible that some nonzero vectors belong to both Nul A and Col A.

EXAMPLE 7 With A as in Example 6, find a nonzero vector in Col A and a nonzero vector in Nul A.

SOLUTION It is easy to find a vector in Col A. Any column of A will do, say, $\begin{bmatrix} 2 \\ -2 \\ 3 \end{bmatrix}$. To find a nonzero vector in Nul A, row reduce the augmented matrix $[\,A \quad \mathbf{0}\,]$ and obtain

$$[\,A \quad \mathbf{0}\,] \sim \begin{bmatrix} 1 & 0 & 9 & 0 & 0 \\ 0 & 1 & -5 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Thus, if $\mathbf{x}$ satisfies $A\mathbf{x} = \mathbf{0}$, then $x_1 = -9x_3$, $x_2 = 5x_3$, $x_4 = 0$, and x_3 is free. Assigning a nonzero value to x_3—say, $x_3 = 1$—we obtain a vector in Nul A, namely, $\mathbf{x} = (-9, 5, 1, 0)$. ■

EXAMPLE 8 With A as in Example 6, let $\mathbf{u} = \begin{bmatrix} 3 \\ -2 \\ -1 \\ 0 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 \\ -1 \\ 3 \end{bmatrix}$.

a. Determine if $\mathbf{u}$ is in Nul A. Could $\mathbf{u}$ be in Col A?

b. Determine if $\mathbf{v}$ is in Col A. Could $\mathbf{v}$ be in Nul A?

SOLUTION

a. An explicit description of Nul A is not needed here. Simply compute the product $A\mathbf{u}$.

$$A\mathbf{u} = \begin{bmatrix} 2 & 4 & -2 & 1 \\ -2 & -5 & 7 & 3 \\ 3 & 7 & -8 & 6 \end{bmatrix} \begin{bmatrix} 3 \\ -2 \\ -1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -3 \\ 3 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Obviously, $\mathbf{u}$ is *not* a solution of $A\mathbf{x} = \mathbf{0}$, so $\mathbf{u}$ is not in Nul A. Also, with four entries, $\mathbf{u}$ could not possibly be in Col A, since Col A is a subspace of $\mathbb{R}^3$.

b. Reduce $[\,A \quad \mathbf{v}\,]$ to an echelon form.

$$[\,A \quad \mathbf{v}\,] = \begin{bmatrix} 2 & 4 & -2 & 1 & 3 \\ -2 & -5 & 7 & 3 & -1 \\ 3 & 7 & -8 & 6 & 3 \end{bmatrix} \sim \begin{bmatrix} 2 & 4 & -2 & 1 & 3 \\ 0 & 1 & -5 & -4 & -2 \\ 0 & 0 & 0 & 17 & 1 \end{bmatrix}$$

At this point, it is clear that the equation $A\mathbf{x} = \mathbf{v}$ is consistent, so $\mathbf{v}$ is in Col A. With only three entries, $\mathbf{v}$ could not possibly be in Nul A, since Nul A is a subspace of $\mathbb{R}^4$. ■

The table on page 241 summarizes what we have learned about Nul A and Col A. Item 8 is a restatement of Theorems 11 and 12(a) in Section 1.9.

Kernel and Range of a Linear Transformation

Subspaces of vector spaces other than $\mathbb{R}^n$ are often described in terms of a linear transformation instead of a matrix. To make this precise, we generalize the definition given in Section 1.8.

Contrast Between Nul A and Col A for an $m \times n$ Matrix A

Nul A	Col A
1. Nul A is a subspace of $\mathbb{R}^n$.	1. Col A is a subspace of $\mathbb{R}^m$.
2. Nul A is implicitly defined; that is, you are given only a condition ($A\mathbf{x} = \mathbf{0}$) that vectors in Nul A must satisfy.	2. Col A is explicitly defined; that is, you are told how to build vectors in Col A.
3. It takes time to find vectors in Nul A. Row operations on $[\,A \quad \mathbf{0}\,]$ are required.	3. It is easy to find vectors in Col A. The columns of A are displayed; others are formed from them.
4. There is no obvious relation between Nul A and the entries in A.	4. There is an obvious relation between Col A and the entries in A, since each column of A is in Col A.
5. A typical vector $\mathbf{v}$ in Nul A has the property that $A\mathbf{v} = \mathbf{0}$.	5. A typical vector $\mathbf{v}$ in Col A has the property that the equation $A\mathbf{x} = \mathbf{v}$ is consistent.
6. Given a specific vector $\mathbf{v}$, it is easy to tell if $\mathbf{v}$ is in Nul A. Just compute $A\mathbf{v}$.	6. Given a specific vector $\mathbf{v}$, it may take time to tell if $\mathbf{v}$ is in Col A. Row operations on $[\,A \quad \mathbf{v}\,]$ are required.
7. Nul $A = \{\mathbf{0}\}$ if and only if the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.	7. Col $A = \mathbb{R}^m$ if and only if the equation $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b}$ in $\mathbb{R}^m$.
8. Nul $A = \{\mathbf{0}\}$ if and only if the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ is one-to-one.	8. Col $A = \mathbb{R}^m$ if and only if the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ maps $\mathbb{R}^n$ *onto* $\mathbb{R}^m$.

DEFINITION

A **linear transformation** T from a vector space V into a vector space W is a rule that assigns to each vector $\mathbf{x}$ in V a unique vector $T(\mathbf{x})$ in W, such that

(i) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ for all $\mathbf{u}$, $\mathbf{v}$ in V, and

(ii) $T(c\mathbf{u}) = cT(\mathbf{u})$ for all $\mathbf{u}$ in V and all scalars c.

The **kernel** (or **null space**) of such a T is the set of all $\mathbf{u}$ in V such that $T(\mathbf{u}) = \mathbf{0}$ (the zero vector in W). The **range** of T is the set of all vectors in W of the form $T(\mathbf{x})$ for some $\mathbf{x}$ in V. If T happens to arise as a matrix transformation—say, $T(\mathbf{x}) = A\mathbf{x}$ for some matrix A—then the kernel and the range of T are just the null space and the column space of A, as defined earlier.

It is not difficult to show that the kernel of T is a subspace of V. The proof is essentially the same as the one for Theorem 2. Also, the range of T is a subspace of W. See Figure 2 and Exercise 42.

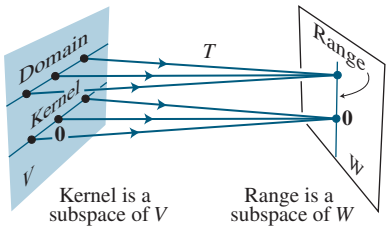


FIGURE 2 Subspaces associated with a linear transformation.

In applications, a subspace usually arises as either the kernel or the range of an appropriate linear transformation. For instance, the set of all solutions of a homogeneous linear differential equation turns out to be the kernel of a linear transformation. Typically, such a linear transformation is described in terms of one or more derivatives of a function. To explain this in any detail would take us too far afield at this point. So we consider only two examples. The first explains why the operation of differentiation is a linear transformation.

EXAMPLE 9 (Calculus required) Let V be the vector space of all real-valued functions f defined on an interval $[a, b]$ with the property that they are differentiable and their derivatives are continuous functions on $[a, b]$. Let W be the vector space $C[a, b]$ of all continuous functions on $[a, b]$, and let $D : V \to W$ be the transformation that changes f in V into its derivative f'. In calculus, two simple differentiation rules are

$$D(f + g) = D(f) + D(g) \quad \text{and} \quad D(cf) = cD(f)$$

That is, D is a linear transformation. It can be shown that the kernel of D is the set of constant functions on $[a, b]$ and the range of D is the set W of all continuous functions on $[a, b]$. ■

EXAMPLE 10 (Calculus required) The differential equation

$$y'' + \omega^2 y = 0 \tag{4}$$

where ω is a constant, is used to describe a variety of physical systems, such as the vibration of a weighted spring, the movement of a pendulum, and the voltage in an inductance-capacitance electrical circuit. The set of solutions of (4) is precisely the kernel of the linear transformation that maps a function $y = f(t)$ into the function $f''(t) + \omega^2 f(t)$. Finding an explicit description of this vector space is a problem in differential equations. The solution set turns out to be the space described in Exercise 19 in Section 4.1. ■

A common technique used in the stock market is technical analysis. Statistical trends gathered from stock-trading activity, such as price movement and volume, are analyzed. Technical analysts focus on patterns of stock-price movements, trading signals, and various other analytical charting tools to evaluate a security's strength or weakness. A moving average is a commonly used indicator in technical analysis. It smooths out price action by filtering out the effects from random price fluctuations. In the final example for this section, we examine the linear transformation that creates the two-day moving average from a "signal" of daily prices. We will look at moving average transformations that average over a longer period of time in Section 4.7.

EXAMPLE 11 Let $\{p_k\}$ in $\mathbb{S}$ represent the price of a stock that has been recorded daily over an extended period of time. Note that we can assume that $p_k = 0$ for k outside the time period under study. To create a two-day moving average, the mapping $M_2 : \mathbb{S} \to \mathbb{S}$ defined by $M_2(\{p_k\}) = \left\{\frac{p_k + p_{k-1}}{2}\right\}$ is applied to the data. Show that M_2 is a linear transformation and find its kernel.

SOLUTION To see that M_2 is a linear transformation, observe that for two signals $\{p_k\}$ and $\{q_k\}$ in $\mathbb{S}$ and any scalar c,

$$M_2(\{p_k\} + \{q_k\}) = M_2(\{p_k + q_k\}) = \left\{\frac{p_k + q_k + p_{k-1} + q_{k-1}}{2}\right\}$$
$$= \left\{\frac{p_k + p_{k-1}}{2}\right\} + \left\{\frac{q_k + q_{k-1}}{2}\right\}$$
$$= M_2(\{p_k\}) + M_2(\{q_k\})$$

and

$$M_2(c\{p_k\}) = M_2(\{cp_k\}) = \left\{\frac{cp_k + cp_{k-1}}{2}\right\} = c\left\{\frac{p_k + p_{k-1}}{2}\right\} = cM_2(\{p_k\})$$

thus M_2 is a linear transformation.

To find the kernel of M_2, notice that $\{p_k\}$ is in the kernel if and only if $\frac{p_k + p_{k-1}}{2} = 0$ for all k, and hence $p_k = -p_{k-1}$. Since this relationship is true for all integers k, it can be applied recursively resulting in $p_k = -p_{k-1} = (-1)^2 p_{k-2} = (-1)^3 p_{k-3} \ldots$. Working out from $k = 0$, any signal in the kernel can be written as $p_k = p_0(-1)^k$, a multiple of the alternating signal described by $\{(-1)^k\}$. Since the kernel of the two-day moving average function consists of all multiples of the alternating sequence, it smooths out daily fluctuations, without leveling out overall trends. (See Figure 3.)

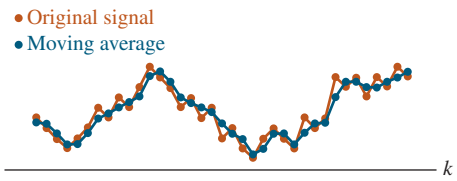


FIGURE 3

Practice Problems

1. Let $W = \left\{\begin{bmatrix} a \\ b \\ c \end{bmatrix} : a - 3b - c = 0\right\}$. Show in two different ways that W is a subspace of $\mathbb{R}^3$. (Use two theorems.)

2. Let $A = \begin{bmatrix} 7 & -3 & 5 \\ -4 & 1 & -5 \\ -5 & 2 & -4 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$, and $\mathbf{w} = \begin{bmatrix} 7 \\ 6 \\ -3 \end{bmatrix}$. Suppose you know that the equations $A\mathbf{x} = \mathbf{v}$ and $A\mathbf{x} = \mathbf{w}$ are both consistent. What can you say about the equation $A\mathbf{x} = \mathbf{v} + \mathbf{w}$?

3. Let A be an $n \times n$ matrix. If Col A = Nul A, show that Nul $A^2 = \mathbb{R}^n$.

4.2 Exercises

1. Determine if $\mathbf{w} = \begin{bmatrix} 1 \\ 3 \\ -4 \end{bmatrix}$ is in Nul A, where $A = \begin{bmatrix} 3 & -5 & -3 \\ 6 & -2 & 0 \\ -8 & 4 & 1 \end{bmatrix}$.

2. Determine if $\mathbf{w} = \begin{bmatrix} 5 \\ -3 \\ 2 \end{bmatrix}$ is in Nul A, where $A = \begin{bmatrix} 5 & 21 & 19 \\ 13 & 23 & 2 \\ 8 & 14 & 1 \end{bmatrix}$.

In Exercises 3–6, find an explicit description of Nul A by listing vectors that span the null space.

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

4. $A = \begin{bmatrix} 1 & -6 & 4 & 0 \\ 0 & 0 & 2 & 0 \end{bmatrix}$

5. $A = \begin{bmatrix} 1 & -2 & 0 & 4 & 0 \\ 0 & 0 & 1 & -9 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

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

In Exercises 7–14, either use an appropriate theorem to show that the given set, W, is a vector space, or find a specific example to the contrary.

7. $\left\{ \begin{bmatrix} a \\ b \\ c \end{bmatrix} : a + b + c = 2 \right\}$ **8.** $\left\{ \begin{bmatrix} r \\ s \\ t \end{bmatrix} : 5r - 1 = s + 2t \right\}$

9. $\left\{ \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} : \begin{array}{l} a - 2b = 4c \\ 2a = c + 3d \end{array} \right\}$ **10.** $\left\{ \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} : \begin{array}{l} a + 3b = c \\ b + c + a = d \end{array} \right\}$

11. $\left\{ \begin{bmatrix} b - 2d \\ 5 + d \\ b + 3d \\ d \end{bmatrix} : b, d \text{ real} \right\}$ **12.** $\left\{ \begin{bmatrix} b - 5d \\ 2b \\ 2d + 1 \\ d \end{bmatrix} : b, d \text{ real} \right\}$

13. $\left\{ \begin{bmatrix} c - 6d \\ d \\ c \end{bmatrix} : c, d \text{ real} \right\}$ **14.** $\left\{ \begin{bmatrix} -a + 2b \\ a - 2b \\ 3a - 6b \end{bmatrix} : a, b \text{ real} \right\}$

In Exercises 15 and 16, find A such that the given set is Col A.

15. $\left\{ \begin{bmatrix} 2s + 3t \\ r + s - 2t \\ 4r + s \\ 3r - s - t \end{bmatrix} : r, s, t \text{ real} \right\}$

16. $\left\{ \begin{bmatrix} b - c \\ 2b + c + d \\ 5c - 4d \\ d \end{bmatrix} : b, c, d \text{ real} \right\}$

For the matrices in Exercises 17–20, (a) find k such that Nul A is a subspace of $\mathbb{R}^k$, and (b) find k such that Col A is a subspace of $\mathbb{R}^k$.

17. $A = \begin{bmatrix} 2 & -8 \\ -1 & 4 \\ 1 & -4 \end{bmatrix}$ **18.** $A = \begin{bmatrix} 8 & -3 & 0 & -1 \\ -3 & 0 & -1 & 8 \\ 0 & -1 & 8 & -3 \end{bmatrix}$

19. $A = \begin{bmatrix} 4 & 5 & -2 & 6 & 0 \\ 1 & 1 & 0 & 1 & 0 \end{bmatrix}$

20. $A = \begin{bmatrix} 1 & -3 & 9 & 0 & -5 \end{bmatrix}$

21. With A as in Exercise 17, find a nonzero vector in Nul A, a nonzero vector in Col A, and a nonzero vector in Row A.

22. With A as in Exercise 3, find a nonzero vector in Nul A, a nonzero vector in Col A, and a nonzero vector in Row A.

23. Let $A = \begin{bmatrix} -6 & 12 \\ -3 & 6 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Determine if $\mathbf{w}$ is in Col A. Is $\mathbf{w}$ in Nul A?

24. Let $A = \begin{bmatrix} -8 & -2 & -9 \\ 6 & 4 & 8 \\ 4 & 0 & 4 \end{bmatrix}$ and $\mathbf{w} = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}$. Determine if $\mathbf{w}$ is in Col A. Is $\mathbf{w}$ in Nul A?

In Exercises 25–38, A denotes an $m \times n$ matrix. Mark each statement True or False **(T/F)**. Justify each answer.

25. **(T/F)** The null space of A is the solution set of the equation $A\mathbf{x} = \mathbf{0}$.

26. **(T/F)** A null space is a vector space.

27. **(T/F)** The null space of an $m \times n$ matrix is in $\mathbb{R}^m$.

28. **(T/F)** The column space of an $m \times n$ matrix is in $\mathbb{R}^m$.

29. **(T/F)** The column space of A is the range of the mapping $\mathbf{x} \mapsto A\mathbf{x}$.

30. **(T/F)** Col A is the set of all solutions of $A\mathbf{x} = \mathbf{b}$.

31. **(T/F)** If the equation $A\mathbf{x} = \mathbf{b}$ is consistent, then Col $A = \mathbb{R}^m$.

32. **(T/F)** Nul A is the kernel of the mapping $\mathbf{x} \mapsto A\mathbf{x}$.

33. **(T/F)** The kernel of a linear transformation is a vector space.

34. **(T/F)** The range of a linear transformation is a vector space.

35. **(T/F)** Col A is the set of all vectors that can be written as $A\mathbf{x}$ for some $\mathbf{x}$.

36. **(T/F)** The set of all solutions of a homogeneous linear differential equation is the kernel of a linear transformation.

37. **(T/F)** The row space of A is the same as the column space of A^T.

38. **(T/F)** The null space of A is the same as the row space of A^T.

39. It can be shown that a solution of the system below is $x_1 = 3$, $x_2 = 2$, and $x_3 = -1$. Use this fact and the theory from this section to explain why another solution is $x_1 = 30$, $x_2 = 20$, and $x_3 = -10$. (Observe how the solutions are related, but make no other calculations.)

$$\begin{aligned} x_1 - 3x_2 - 3x_3 &= 0 \\ -2x_1 + 4x_2 + 2x_3 &= 0 \\ -x_1 + 5x_2 + 7x_3 &= 0 \end{aligned}$$

40. Consider the following two systems of equations:

$$\begin{aligned} 5x_1 + x_2 - 3x_3 &= 0 \\ -9x_1 + 2x_2 + 5x_3 &= 1 \\ 4x_1 + x_2 - 6x_3 &= 9 \end{aligned} \qquad \begin{aligned} 5x_1 + x_2 - 3x_3 &= 0 \\ -9x_1 + 2x_2 + 5x_3 &= 5 \\ 4x_1 + x_2 - 6x_3 &= 45 \end{aligned}$$

It can be shown that the first system has a solution. Use this fact and the theory from this section to explain why the second system must also have a solution. (Make no row operations.)

41. Prove Theorem 3 as follows: Given an $m \times n$ matrix A, an element in Col A has the form $A\mathbf{x}$ for some $\mathbf{x}$ in $\mathbb{R}^n$. Let $A\mathbf{x}$ and $A\mathbf{w}$ represent any two vectors in Col A.

a. Explain why the zero vector is in Col A.

b. Show that the vector $A\mathbf{x} + A\mathbf{w}$ is in Col A.

c. Given a scalar c, show that $c(A\mathbf{x})$ is in Col A.

42. Let $T : V \to W$ be a linear transformation from a vector space V into a vector space W. Prove that the range of T is a subspace of W. [*Hint:* Typical elements of the range have the form $T(\mathbf{x})$ and $T(\mathbf{w})$ for some $\mathbf{x}$, $\mathbf{w}$ in V.]

43. Define $T : \mathbb{P}_2 \to \mathbb{R}^2$ by $T(\mathbf{p}) = \begin{bmatrix} \mathbf{p}(0) \\ \mathbf{p}(1) \end{bmatrix}$. For instance, if $\mathbf{p}(t) = 3 + 5t + 7t^2$, then $T(\mathbf{p}) = \begin{bmatrix} 3 \\ 15 \end{bmatrix}$.

a. Show that T is a linear transformation. [*Hint:* For arbitrary polynomials $\mathbf{p}$, $\mathbf{q}$ in $\mathbb{P}_2$, compute $T(\mathbf{p} + \mathbf{q})$ and $T(c\mathbf{p})$.]

b. Find a polynomial $\mathbf{p}$ in $\mathbb{P}_2$ that spans the kernel of T, and describe the range of T.

44. Define a linear transformation $T : \mathbb{P}_2 \to \mathbb{R}^2$ by $T(\mathbf{p}) = \begin{bmatrix} \mathbf{p}(0) \\ \mathbf{p}(0) \end{bmatrix}$. Find polynomials $\mathbf{p}_1$ and $\mathbf{p}_2$ in $\mathbb{P}_2$ that span the kernel of T, and describe the range of T.

45. Let $M_{2\times 2}$ be the vector space of all 2×2 matrices, and define $T : M_{2\times 2} \to M_{2\times 2}$ by $T(A) = A + A^T$, where $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

a. Show that T is a linear transformation.

b. Let B be any element of $M_{2\times 2}$ such that $B^T = B$. Find an A in $M_{2\times 2}$ such that $T(A) = B$.

c. Show that the range of T is the set of B in $M_{2\times 2}$ with the property that $B^T = B$.

d. Describe the kernel of T.

STUDY GUIDE offers additional resources for mastering vector spaces, subspaces, and column row, and null spaces.

46. (*Calculus required*) Define $T : C[0, 1] \to C[0, 1]$ as follows: For $\mathbf{f}$ in $C[0, 1]$, let $T(\mathbf{f})$ be the antiderivative $\mathbf{F}$ of $\mathbf{f}$ such that $\mathbf{F}(0) = 0$. Show that T is a linear transformation, and describe the kernel of T. (See the notation in Exercise 20 of Section 4.1.)

47. Let V and W be vector spaces, and let $T : V \to W$ be a linear transformation. Given a subspace U of V, let $T(U)$ denote the set of all images of the form $T(\mathbf{x})$, where $\mathbf{x}$ is in U. Show that $T(U)$ is a subspace of W.

48. Given $T : V \to W$ as in Exercise 47, and given a subspace Z of W, let U be the set of all $\mathbf{x}$ in V such that $T(\mathbf{x})$ is in Z. Show that U is a subspace of V.

T 49. Determine whether $\mathbf{w}$ is in the column space of A, the null space of A, or both, where

$$\mathbf{w} = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -3 \end{bmatrix}, \quad A = \begin{bmatrix} 7 & 6 & -4 & 1 \\ -5 & -1 & 0 & -2 \\ 9 & -11 & 7 & -3 \\ 19 & -9 & 7 & 1 \end{bmatrix}$$

T 50. Determine whether $\mathbf{w}$ is in the column space of A, the null space of A, or both, where

$$\mathbf{w} = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}, \quad A = \begin{bmatrix} -8 & 5 & -2 & 0 \\ -5 & 2 & 1 & -2 \\ 10 & -8 & 6 & -3 \\ 3 & -2 & 1 & 0 \end{bmatrix}$$

T 51. Let $\mathbf{a}_1, \ldots, \mathbf{a}_5$ denote the columns of the matrix A, where

$$A = \begin{bmatrix} 5 & 1 & 2 & 2 & 0 \\ 3 & 3 & 2 & -1 & -12 \\ 8 & 4 & 4 & -5 & 12 \\ 2 & 1 & 1 & 0 & -2 \end{bmatrix}, \quad B = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_4 \end{bmatrix}$$

a. Explain why $\mathbf{a}_3$ and $\mathbf{a}_5$ are in the column space of B.

b. Find a set of vectors that spans Nul A.

c. Let $T : \mathbb{R}^5 \to \mathbb{R}^4$ be defined by $T(\mathbf{x}) = A\mathbf{x}$. Explain why T is neither one-to-one nor onto.

T 52. Let $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ and $K = \text{Span}\{\mathbf{v}_3, \mathbf{v}_4\}$, where

$$\mathbf{v}_1 = \begin{bmatrix} 5 \\ 3 \\ 8 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 2 \\ -1 \\ 5 \end{bmatrix}, \mathbf{v}_4 = \begin{bmatrix} 0 \\ -12 \\ -28 \end{bmatrix}.$$

Then H and K are subspaces of $\mathbb{R}^3$. In fact, H and K are planes in $\mathbb{R}^3$ through the origin, and they intersect in a line through $\mathbf{0}$. Find a nonzero vector $\mathbf{w}$ that generates that line. [*Hint:* $\mathbf{w}$ can be written as $c_1\mathbf{v}_1 + c_2\mathbf{v}_2$ and also as $c_3\mathbf{v}_3 + c_4\mathbf{v}_4$. To build $\mathbf{w}$, solve the equation $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = c_3\mathbf{v}_3 + c_4\mathbf{v}_4$ for the unknown c_j's.]

Solutions to Practice Problems

1. *First method:* W is a subspace of $\mathbb{R}^3$ by Theorem 2 because W is the set of all solutions to a system of homogeneous linear equations (where the system has only one equation). Equivalently, W is the null space of the 1×3 matrix $A = [\,1 \quad -3 \quad -1\,]$.

Second method: Solve the equation $a - 3b - c = 0$ for the leading variable a in terms of the free variables b and c. Any solution has the form $\begin{bmatrix} 3b + c \\ b \\ c \end{bmatrix}$, where b and c are arbitrary, and

$$\begin{bmatrix} 3b + c \\ b \\ c \end{bmatrix} = b \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} + c \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

$\mathbf{v}_1$ $\mathbf{v}_2$

This calculation shows that $W = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$. Thus W is a subspace of $\mathbb{R}^3$ by Theorem 1. We could also solve the equation $a - 3b - c = 0$ for b or c and get alternative descriptions of W as a set of linear combinations of two vectors.

2. Both $\mathbf{v}$ and $\mathbf{w}$ are in Col A. Since Col A is a vector space, $\mathbf{v} + \mathbf{w}$ must be in Col A. That is, the equation $A\mathbf{x} = \mathbf{v} + \mathbf{w}$ is consistent.

3. Let $\mathbf{x}$ be any vector in $\mathbb{R}^n$. Notice $A\mathbf{x}$ is in Col A, since it is a linear combination of the columns of A. Since Col A = Nul A, the vector $A\mathbf{x}$ is also in Nul A. Hence $A^2\mathbf{x} = A(A\mathbf{x}) = \mathbf{0}$ establishing that every vector $\mathbf{x}$ from $\mathbb{R}^n$ is in Nul A^2.

4.3 Linearly Independent Sets; Bases

In this section we identify and study the subsets that span a vector space V or a subspace H as "efficiently" as possible. The key idea is that of linear independence, defined as in $\mathbb{R}^n$.

An indexed set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ in V is said to be **linearly independent** if the vector equation

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_p\mathbf{v}_p = \mathbf{0} \tag{1}$$

has *only* the trivial solution, $c_1 = 0, \ldots, c_p = 0$.[1]

The set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is said to be **linearly dependent** if (1) has a nontrivial solution, that is, if there are some weights, $c_1, \ldots, c_p$, *not all zero*, such that (1) holds. In such a case, (1) is called a **linear dependence relation** among $\mathbf{v}_1, \ldots, \mathbf{v}_p$.

Just as in $\mathbb{R}^n$, a set containing a single vector $\mathbf{v}$ is linearly independent if and only if $\mathbf{v} \neq \mathbf{0}$. Also, a set of two vectors is linearly dependent if and only if one of the vectors is a multiple of the other. And any set containing the zero vector is linearly dependent. The following theorem has the same proof as Theorem 7 in Section 1.7.

[1] It is convenient to use $c_1, \ldots, c_p$ in (1) for the scalars instead of $x_1, \ldots, x_p$, as we did previously.

THEOREM 4

An indexed set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ of two or more vectors, with $\mathbf{v}_1 \neq \mathbf{0}$, is linearly dependent if and only if some $\mathbf{v}_j$ (with $j > 1$) is a linear combination of the preceding vectors, $\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}$.

The main difference between linear dependence in $\mathbb{R}^n$ and in a general vector space is that when the vectors are not n-tuples, the homogeneous equation (1) usually cannot be written as a system of n linear equations. That is, the vectors cannot be made into the columns of a matrix A in order to study the equation $A\mathbf{x} = \mathbf{0}$. We must rely instead on the definition of linear dependence and on Theorem 4.

EXAMPLE 1 Let $\mathbf{p}_1(t) = 1$, $\mathbf{p}_2(t) = t$, and $\mathbf{p}_3(t) = 4 - t$. Then $\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$ is linearly dependent in $\mathbb{P}$ because $\mathbf{p}_3 = 4\mathbf{p}_1 - \mathbf{p}_2$. ■

EXAMPLE 2 The set $\{\sin t, \cos t\}$ is linearly independent in $C[0, 1]$, the space of all continuous functions on $0 \leq t \leq 1$, because $\sin t$ and $\cos t$ are not multiples of one another *as vectors in* $C[0, 1]$. That is, there is no scalar c such that $\cos t = c \cdot \sin t$ for all t in $[0, 1]$. (Look at the graphs of $\sin t$ and $\cos t$.) However, $\{\sin t \cos t, \sin 2t\}$ is linearly dependent because of the identity $\sin 2t = 2 \sin t \cos t$, for all t. ■

DEFINITION

Let H be a subspace of a vector space V. A set of vectors $\mathcal{B}$ in V is a **basis** for H if

(i) $\mathcal{B}$ is a linearly independent set, and

(ii) the subspace spanned by $\mathcal{B}$ coincides with H; that is,

$$H = \text{Span } \mathcal{B}$$

The definition of a basis applies to the case when $H = V$, because any vector space is a subspace of itself. Thus a basis of V is a linearly independent set that spans V. Observe that when $H \neq V$, condition (ii) includes the requirement that each of the vectors $\mathbf{b}$ in $\mathcal{B}$ must belong to H, because Span $\mathcal{B}$ contains every element in $\mathcal{B}$, as shown in Section 4.1.

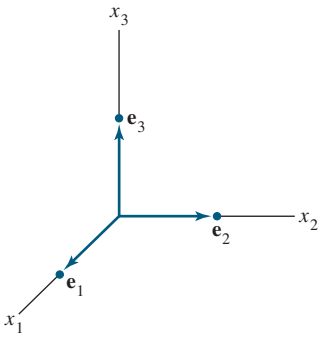


FIGURE 1
The standard basis for $\mathbb{R}^3$.

EXAMPLE 3 Let A be an invertible $n \times n$ matrix—say, $A = [\,\mathbf{a}_1 \;\cdots\; \mathbf{a}_n\,]$. Then the columns of A form a basis for $\mathbb{R}^n$ because they are linearly independent and they span $\mathbb{R}^n$, by the Invertible Matrix Theorem. ■

EXAMPLE 4 Let $\mathbf{e}_1, \ldots, \mathbf{e}_n$ be the columns of the $n \times n$ identity matrix, I_n. That is,

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

The set $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is called the **standard basis** for $\mathbb{R}^n$ (Figure 1). ■

EXAMPLE 5 Let $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 0 \\ -6 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} -4 \\ 1 \\ 7 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} -2 \\ 1 \\ 5 \end{bmatrix}$. Determine if $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a basis for $\mathbb{R}^3$.

SOLUTION Since there are exactly three vectors here in $\mathbb{R}^3$, we can use any of several methods to determine if the matrix $A = [\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \mathbf{v}_3\,]$ is invertible. For instance, two row replacements reveal that A has three pivot positions. Thus A is invertible. As in Example 3, the columns of A form a basis for $\mathbb{R}^3$. ■

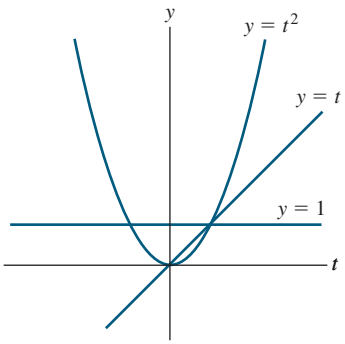


FIGURE 2
The standard basis for $\mathbb{P}_2$.

EXAMPLE 6 Let $S = \{1, t, t^2, \ldots, t^n\}$. Verify that S is a basis for $\mathbb{P}_n$. This basis is called the **standard basis** for $\mathbb{P}_n$.

SOLUTION Certainly S spans $\mathbb{P}_n$. To show that S is linearly independent, suppose that $c_0, \ldots, c_n$ satisfy

$$c_0 1 + c_1 t + c_2 t^2 + \cdots + c_n t^n = \mathbf{0}(t) \tag{2}$$

This equality means that the polynomial on the left has the same values as the zero polynomial on the right. A fundamental theorem in algebra says that the only polynomial in $\mathbb{P}_n$ with more than n zeros is the zero polynomial. That is, equation (2) holds for all t only if $c_0 = \cdots = c_n = 0$. This proves that S is linearly independent and hence is a basis for $\mathbb{P}_n$. See Figure 2. ■

Problems involving linear independence and spanning in $\mathbb{P}_n$ are handled best by a technique to be discussed in Section 4.4.

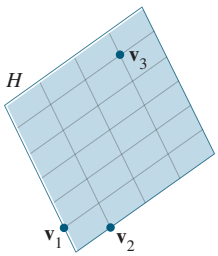
The Spanning Set Theorem

As we will see, a basis is an "efficient" spanning set that contains no unnecessary vectors. In fact, a basis can be constructed from a spanning set by discarding unneeded vectors.

EXAMPLE 7 Let

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 2 \\ -1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 6 \\ 16 \\ -5 \end{bmatrix}, \quad \text{and} \quad H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}.$$

Note that $\mathbf{v}_3 = 5\mathbf{v}_1 + 3\mathbf{v}_2$, and show that $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$. Then find a basis for the subspace H.



SOLUTION Every vector in $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ belongs to H because

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + 0\mathbf{v}_3$$

Now let $\mathbf{x}$ be any vector in H—say, $\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3$. Since $\mathbf{v}_3 = 5\mathbf{v}_1 + 3\mathbf{v}_2$, we may substitute

$$\begin{aligned} \mathbf{x} &= c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3(5\mathbf{v}_1 + 3\mathbf{v}_2) \\ &= (c_1 + 5c_3)\mathbf{v}_1 + (c_2 + 3c_3)\mathbf{v}_2 \end{aligned}$$

Thus $\mathbf{x}$ is in $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$, so every vector in H already belongs to $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$. We conclude that H and $\text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ are actually the same set of vectors. It follows that $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis of H since $\{\mathbf{v}_1, \mathbf{v}_2\}$ is obviously linearly independent. ■

The next theorem generalizes Example 7.

THEOREM 5

The Spanning Set Theorem

Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ be a set in a vector space V, and let $H = \text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$.

a. If one of the vectors in S—say, $\mathbf{v}_k$—is a linear combination of the remaining vectors in S, then the set formed from S by removing $\mathbf{v}_k$ still spans H.

b. If $H \neq \{\mathbf{0}\}$, some subset of S is a basis for H.

PROOF

a. By rearranging the list of vectors in S, if necessary, we may suppose that $\mathbf{v}_p$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_{p-1}$—say,

$$\mathbf{v}_p = a_1\mathbf{v}_1 + \cdots + a_{p-1}\mathbf{v}_{p-1} \tag{3}$$

Given any $\mathbf{x}$ in H, we may write

$$\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_{p-1}\mathbf{v}_{p-1} + c_p\mathbf{v}_p \tag{4}$$

for suitable scalars $c_1, \ldots, c_p$. Substituting the expression for $\mathbf{v}_p$ from (3) into (4), it is easy to see that $\mathbf{x}$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_{p-1}$. Thus $\{\mathbf{v}_1, \ldots, \mathbf{v}_{p-1}\}$ spans H, because $\mathbf{x}$ was an arbitrary element of H.

b. If the original spanning set S is linearly independent, then it is already a basis for H. Otherwise, one of the vectors in S depends on the others and can be deleted, by part (a). So long as there are two or more vectors in the spanning set, we can repeat this process until the spanning set is linearly independent and hence is a basis for H. If the spanning set is eventually reduced to one vector, that vector will be nonzero (and hence linearly independent) because $H \neq \{\mathbf{0}\}$. ■

Bases for Nul *A*, Col *A*, and Row *A*

We already know how to find vectors that span the null space of a matrix A. The discussion in Section 4.2 pointed out that our method always produces a linearly independent set when Nul A contains nonzero vectors. So, in this case, that method produces a *basis* for Nul A.

The next two examples describe a simple algorithm for finding a basis for the column space.

EXAMPLE 8 Find a basis for Col B, where

$$B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_5 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 0 & 2 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

SOLUTION Each nonpivot column of B is a linear combination of the pivot columns. In fact, $\mathbf{b}_2 = 4\mathbf{b}_1$ and $\mathbf{b}_4 = 2\mathbf{b}_1 - \mathbf{b}_3$. By the Spanning Set Theorem, we may discard $\mathbf{b}_2$ and $\mathbf{b}_4$, and $\{\mathbf{b}_1, \mathbf{b}_3, \mathbf{b}_5\}$ will still span Col B. Let

$$S = \{\mathbf{b}_1, \mathbf{b}_3, \mathbf{b}_5\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$

Since $\mathbf{b}_1 \neq 0$ and no vector in S is a linear combination of the vectors that precede it, S is linearly independent (Theorem 4). Thus S is a basis for Col B. ■

What about a matrix A that is *not* in reduced echelon form? Recall that any linear dependence relationship among the columns of A can be expressed in the form $A\mathbf{x} = \mathbf{0}$, where $\mathbf{x}$ is a column of weights. (If some columns are not involved in a particular dependence relation, then their weights are zero.) When A is row reduced to a matrix B, the columns of B are often totally different from the columns of A. However, the equations $A\mathbf{x} = \mathbf{0}$ and $B\mathbf{x} = \mathbf{0}$ have exactly the same set of solutions. If $A = [\,\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_n\,]$ and $B = [\,\mathbf{b}_1 \quad \cdots \quad \mathbf{b}_n\,]$, then the vector equations

$$x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \mathbf{0} \quad \text{and} \quad x_1\mathbf{b}_1 + \cdots + x_n\mathbf{b}_n = \mathbf{0}$$

also have the same set of solutions. That is, the columns of A have *exactly the same linear dependence relationships* as the columns of B.

EXAMPLE 9 It can be shown that the matrix

$$A = [\,\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_5\,] = \begin{bmatrix} 1 & 4 & 0 & 2 & -1 \\ 3 & 12 & 1 & 5 & 5 \\ 2 & 8 & 1 & 3 & 2 \\ 5 & 20 & 2 & 8 & 8 \end{bmatrix}$$

is row equivalent to the matrix B in Example 8. Find a basis for Col A.

SOLUTION In Example 8 we saw that

$$\mathbf{b}_2 = 4\mathbf{b}_1 \quad \text{and} \quad \mathbf{b}_4 = 2\mathbf{b}_1 - \mathbf{b}_3$$

so we can expect that

$$\mathbf{a}_2 = 4\mathbf{a}_1 \quad \text{and} \quad \mathbf{a}_4 = 2\mathbf{a}_1 - \mathbf{a}_3$$

Check that this is indeed the case! Thus we may discard $\mathbf{a}_2$ and $\mathbf{a}_4$ when selecting a minimal spanning set for Col A. In fact, $\{\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_5\}$ must be linearly independent because any linear dependence relationship among $\mathbf{a}_1$, $\mathbf{a}_3$, $\mathbf{a}_5$ would imply a linear dependence relationship among $\mathbf{b}_1$, $\mathbf{b}_3$, $\mathbf{b}_5$. But we know that $\{\mathbf{b}_1, \mathbf{b}_3, \mathbf{b}_5\}$ is a linearly independent set. Thus $\{\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_5\}$ is a basis for Col A. The columns we have used for this basis are the pivot columns of A. ■

Examples 8 and 9 illustrate the following useful fact.

THEOREM 6 The pivot columns of a matrix A form a basis for Col A.

PROOF The general proof uses the arguments discussed above. Let B be the reduced echelon form of A. The set of pivot columns of B is linearly independent, for no vector in the set is a linear combination of the vectors that precede it. Since A is row equivalent to B, the pivot columns of A are linearly independent as well, because any linear dependence relation among the columns of A corresponds to a linear dependence relation among the columns of B. For this same reason, every nonpivot column of A is a linear combination of the pivot columns of A. Thus the nonpivot columns of A may be discarded from the spanning set for Col A, by the Spanning Set Theorem. This leaves the pivot columns of A as a basis for Col A. ■

Warning: The pivot columns of a matrix A are evident when A has been reduced only to *echelon* form. But, be careful to use the *pivot columns of A itself* for the basis of Col A. Row operations can change the column space of a matrix. The columns of an echelon form B of A are often not in the column space of A. For instance, the columns of matrix B in Example 8 all have zeros in their last entries, so they cannot span the column space of matrix A in Example 9.

In contrast, the following theorem establishes that row reduction does not change the row space of a matrix.

THEOREM 7

If two matrices A and B are row equivalent, then their row spaces are the same. If B is in echelon form, the nonzero rows of B form a basis for the row space of A as well as for that of B.

PROOF If B is obtained from A by row operations, the rows of B are linear combinations of the rows of A. It follows that any linear combination of the rows of B is automatically a linear combination of the rows of A. Thus the row space of B is contained in the row space of A. Since row operations are reversible, the same argument shows that the row space of A is a subset of the row space of B. So the two row spaces are the same. If B is in echelon form, its nonzero rows are linearly independent because no nonzero row is a linear combination of the nonzero rows below it. (Apply Theorem 4 to the nonzero rows of B in reverse order, with the first row last.) Thus the nonzero rows of B form a basis of the (common) row space of B and A. ■

EXAMPLE 10 Find a basis for the row space of the matrix A from Example 9.

SOLUTION To find a basis for the row space, recall that matrix A from Example 9 is row equivalent to matrix B from Example 8:

$$A = \begin{bmatrix} 1 & 4 & 0 & 2 & -1 \\ 3 & 12 & 1 & 5 & 5 \\ 2 & 8 & 1 & 3 & 2 \\ 5 & 20 & 2 & 8 & 8 \end{bmatrix} \sim B = \begin{bmatrix} 1 & 4 & 0 & 2 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

By Theorem 7, the first three rows of B form a basis for the row space of A (as well as for the row space of B). Thus

$$\text{Basis for Row } A : \{(1, 4, 0, 2, 0), (0, 0, 1, -1, 0), (0, 0, 0, 0, 1)\}$$

Observe that, unlike the basis for Col A, the bases for Row A and Nul A have no simple connection with the entries in A itself.[2]

Two Views of a Basis

When the Spanning Set Theorem is used, the deletion of vectors from a spanning set must stop when the set becomes linearly independent. If an additional vector is deleted,

[2] It is possible to find a basis for the row space Row A that uses rows of A. First form A^T, and then row reduce until the pivot columns of A^T are found. These pivot columns of A^T are rows of A, and they form a basis for the row space of A.

it will not be a linear combination of the remaining vectors, and hence the smaller set will no longer span V. Thus a basis is a spanning set that is as small as possible.

A basis is also a linearly independent set that is as large as possible. If S is a basis for V, and if S is enlarged by one vector—say, $\mathbf{w}$—from V, then the new set cannot be linearly independent, because S spans V, and $\mathbf{w}$ is therefore a linear combination of the elements in S.

EXAMPLE 11 The following three sets in $\mathbb{R}^3$ show how a linearly independent set can be enlarged to a basis and how further enlargement destroys the linear independence of the set. Also, a spanning set can be shrunk to a basis, but further shrinking destroys the spanning property.

$$\left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}2\\3\\0\end{bmatrix}\right\} \qquad \left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}2\\3\\0\end{bmatrix}, \begin{bmatrix}4\\5\\6\end{bmatrix}\right\} \qquad \left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}2\\3\\0\end{bmatrix}, \begin{bmatrix}4\\5\\6\end{bmatrix}, \begin{bmatrix}7\\8\\9\end{bmatrix}\right\}$$

Linearly independent but does not span $\mathbb{R}^3$ | A basis for $\mathbb{R}^3$ | Spans $\mathbb{R}^3$ but is linearly dependent ■

Practice Problems

1. Let $\mathbf{v}_1 = \begin{bmatrix}1\\-2\\3\end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix}-2\\7\\-9\end{bmatrix}$. Determine if $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for $\mathbb{R}^3$. Is $\{\mathbf{v}_1, \mathbf{v}_2\}$ a basis for $\mathbb{R}^2$?

2. Let $\mathbf{v}_1 = \begin{bmatrix}1\\-3\\4\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}6\\2\\-1\end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix}2\\-2\\3\end{bmatrix}$, and $\mathbf{v}_4 = \begin{bmatrix}-4\\-8\\9\end{bmatrix}$. Find a basis for the subspace W spanned by $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$.

3. Let $\mathbf{v}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}$, and $H = \left\{\begin{bmatrix}s\\s\\0\end{bmatrix} : s \text{ in } \mathbb{R}\right\}$. Then every vector in H is a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$ because

$$\begin{bmatrix}s\\s\\0\end{bmatrix} = s\begin{bmatrix}1\\0\\0\end{bmatrix} + s\begin{bmatrix}0\\1\\0\end{bmatrix}$$

Is $\{\mathbf{v}_1, \mathbf{v}_2\}$ a basis for H?

4. Let V and W be vector spaces, let $T : V \to W$ and $U : V \to W$ be linear transformations, and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ be a basis for V. If $T(\mathbf{v}_j) = U(\mathbf{v}_j)$ for every value of j between 1 and p, show that $T(\mathbf{x}) = U(\mathbf{x})$ for every vector $\mathbf{x}$ in V.

STUDY GUIDE offers additional resources for mastering the concept of basis.

4.3 Exercises

Determine which sets in Exercises 1–8 are bases for $\mathbb{R}^3$. Of the sets that are *not* bases, determine which ones are linearly independent and which ones span $\mathbb{R}^3$. Justify your answers.

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

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

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

4. $\begin{bmatrix}2\\-2\\1\end{bmatrix}, \begin{bmatrix}1\\-3\\2\end{bmatrix}, \begin{bmatrix}-7\\5\\4\end{bmatrix}$

5. $\begin{bmatrix}1\\-3\\0\end{bmatrix}, \begin{bmatrix}-2\\9\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\-3\\5\end{bmatrix}$

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

7. $\begin{bmatrix} -2 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 6 \\ -1 \\ 5 \end{bmatrix}$ **8.** $\begin{bmatrix} 1 \\ -4 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ -5 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ -2 \end{bmatrix}$

Find bases for the null spaces of the matrices given in Exercises 9 and 10. Refer to the remarks that follow Example 3 in Section 4.2.

9. $\begin{bmatrix} 1 & 0 & -3 & 2 \\ 0 & 1 & -5 & 4 \\ 3 & -2 & 1 & -2 \end{bmatrix}$ **10.** $\begin{bmatrix} 1 & 0 & -5 & 1 & 4 \\ -2 & 1 & 6 & -2 & -2 \\ 0 & 2 & -8 & 1 & 9 \end{bmatrix}$

11. Find a basis for the set of vectors in $\mathbb{R}^3$ in the plane $x + 4y - 5z = 0$. [Hint: Think of the equation as a "system" of homogeneous equations.]

12. Find a basis for the set of vectors in $\mathbb{R}^2$ on the line $y = 5x$.

In Exercises 13 and 14, assume that A is row equivalent to B. Find bases for Nul A, Col A, and Row A.

13. $A = \begin{bmatrix} -2 & 4 & -2 & -4 \\ 2 & -6 & -3 & 1 \\ -3 & 8 & 2 & -3 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & 6 & 5 \\ 0 & 2 & 5 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

14. $A = \begin{bmatrix} 1 & 2 & -5 & 11 & -3 \\ 2 & 4 & -5 & 15 & 2 \\ 1 & 2 & 0 & 4 & 5 \\ 3 & 6 & -5 & 19 & -2 \end{bmatrix}$,

$B = \begin{bmatrix} 1 & 2 & 0 & 4 & 5 \\ 0 & 0 & 5 & -7 & 8 \\ 0 & 0 & 0 & 0 & -9 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

In Exercises 15–18, find a basis for the space spanned by the given vectors, $\mathbf{v}_1, \ldots, \mathbf{v}_5$.

15. $\begin{bmatrix} 1 \\ 0 \\ -3 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} -3 \\ -4 \\ 1 \\ 6 \end{bmatrix}, \begin{bmatrix} 1 \\ -3 \\ -8 \\ 7 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -6 \\ 9 \end{bmatrix}$

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

T 17. $\begin{bmatrix} 8 \\ 9 \\ -3 \\ -6 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 5 \\ 1 \\ -4 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ -4 \\ -9 \\ 6 \\ -7 \end{bmatrix}, \begin{bmatrix} 6 \\ 8 \\ 4 \\ -7 \\ 10 \end{bmatrix}, \begin{bmatrix} -1 \\ 4 \\ 11 \\ -8 \\ -7 \end{bmatrix}$

T 18. $\begin{bmatrix} -8 \\ 7 \\ 6 \\ 5 \\ -7 \end{bmatrix}, \begin{bmatrix} 8 \\ -7 \\ -9 \\ -5 \\ 7 \end{bmatrix}, \begin{bmatrix} -8 \\ 7 \\ 4 \\ 5 \\ -7 \end{bmatrix}, \begin{bmatrix} 1 \\ 4 \\ 9 \\ 6 \\ -7 \end{bmatrix}, \begin{bmatrix} -9 \\ 3 \\ -4 \\ -1 \\ 0 \end{bmatrix}$

19. Let $\mathbf{v}_1 = \begin{bmatrix} 4 \\ -3 \\ 7 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 9 \\ -2 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 7 \\ 11 \\ 6 \end{bmatrix}$, and $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$. It can be verified that $4\mathbf{v}_1 + 5\mathbf{v}_2 - 3\mathbf{v}_3 = \mathbf{0}$. Use this information to find a basis for H. There is more than one answer.

20. Let $\mathbf{v}_1 = \begin{bmatrix} 7 \\ 4 \\ -9 \\ -5 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 4 \\ -7 \\ 2 \\ 5 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 1 \\ -5 \\ 3 \\ 4 \end{bmatrix}$. It can be verified that $\mathbf{v}_1 - 3\mathbf{v}_2 + 5\mathbf{v}_3 = \mathbf{0}$. Use this information to find a basis for $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

In Exercises 21–32, mark each statement True or False **(T/F)**. Justify each answer.

21. **(T/F)** A single vector by itself is linearly dependent.

22. **(T/F)** A linearly independent set in a subspace H is a basis for H.

23. **(T/F)** If $H = \text{Span}\{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$, then $\{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$ is a basis for H.

24. **(T/F)** If a finite set S of nonzero vectors spans a vector space V, then some subset of S is a basis for V.

25. **(T/F)** The columns of an invertible $n \times n$ matrix form a basis for $\mathbb{R}^n$.

26. **(T/F)** A basis is a linearly independent set that is as large as possible.

27. **(T/F)** A basis is a spanning set that is as large as possible.

28. **(T/F)** The standard method for producing a spanning set for Nul A, described in Section 4.2, sometimes fails to produce a basis for Nul A.

29. **(T/F)** In some cases, the linear dependence relations among the columns of a matrix can be affected by certain elementary row operations on the matrix.

30. **(T/F)** If B is an echelon form of a matrix A, then the pivot columns of B form a basis for Col A.

31. **(T/F)** Row operations preserve the linear dependence relations among the rows of A.

32. **(T/F)** If A and B are row equivalent, then their row spaces are the same.

33. Suppose $\mathbb{R}^4 = \text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_4\}$. Explain why $\{\mathbf{v}_1, \ldots, \mathbf{v}_4\}$ is a basis for $\mathbb{R}^4$.

34. Let $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a linearly independent set in $\mathbb{R}^n$. Explain why $\mathcal{B}$ must be a basis for $\mathbb{R}^n$.

35. Let $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, and let H be the set of vectors in $\mathbb{R}^3$ whose second and third entries are equal. Then every vector in H has a unique expansion as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, because

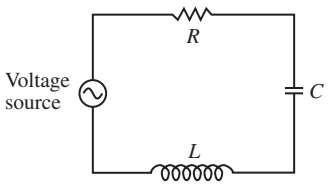
$$\begin{bmatrix} s \\ t \\ t \end{bmatrix} = s\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + (t-s)\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} + s\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

for any s and t. Is $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ a basis for H? Why or why not?

36. In the vector space of all real-valued functions, find a basis for the subspace spanned by $\{\sin t, \sin 2t, \sin t \cos t\}$.

37. Let V be the vector space of functions that describe the vibration of a mass–spring system. (Refer to Exercise 19 in Section 4.1.) Find a basis for V.

38. (*RLC circuit*) The circuit in the figure consists of a resistor (R ohms), an inductor (L henrys), a capacitor (C farads), and an initial voltage source. Let $b = R/(2L)$, and suppose R, L, and C have been selected so that b also equals $1/\sqrt{LC}$. (This is done, for instance, when the circuit is used in a voltmeter.) Let $v(t)$ be the voltage (in volts) at time t, measured across the capacitor. It can be shown that v is in the null space H of the linear transformation that maps $v(t)$ into $Lv''(t) + Rv'(t) + (1/C)v(t)$, and H consists of all functions of the form $v(t) = e^{-bt}(c_1 + c_2 t)$. Find a basis for H.



Exercises 39 and 40 show that every basis for $\mathbb{R}^n$ must contain exactly n vectors.

39. Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a set of k vectors in $\mathbb{R}^n$, with $k < n$. Use a theorem from Section 1.4 to explain why S cannot be a basis for $\mathbb{R}^n$.

40. Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a set of k vectors in $\mathbb{R}^n$, with $k > n$. Use a theorem from Chapter 1 to explain why S cannot be a basis for $\mathbb{R}^n$.

Exercises 41 and 42 reveal an important connection between linear independence and linear transformations and provide practice using the definition of linear dependence. Let V and W be vector spaces, let $T : V \to W$ be a linear transformation, and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ be a subset of V.

41. Show that if $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is linearly dependent in V, then the set of images, $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_p)\}$, is linearly dependent in W. This fact shows that if a linear transformation maps a set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ onto a linearly *independent* set $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_p)\}$, then the original set is linearly independent, too (because it cannot be linearly dependent).

42. Suppose that T is a one-to-one transformation, so that an equation $T(\mathbf{u}) = T(\mathbf{v})$ always implies $\mathbf{u} = \mathbf{v}$. Show that if the set of images $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_p)\}$ is linearly dependent, then $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is linearly dependent. This fact shows that *a one-to-one linear transformation maps a linearly independent set onto a linearly independent set* (because in this case the set of images cannot be linearly dependent).

43. Consider the polynomials $\mathbf{p}_1(t) = 1 + t^2$ and $\mathbf{p}_2(t) = 1 - t^2$. Is $\{\mathbf{p}_1, \mathbf{p}_2\}$ a linearly independent set in $\mathbb{P}_3$? Why or why not?

44. Consider the polynomials $\mathbf{p}_1(t) = 1 + t$, $\mathbf{p}_2(t) = 1 - t$, and $\mathbf{p}_3(t) = 2$ (for all t). By inspection, write a linear dependence relation among $\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_3$. Then find a basis for Span $\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3\}$.

45. Let V be a vector space that contains a linearly independent set $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$. Describe how to construct a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$ in V such that $\{\mathbf{v}_1, \mathbf{v}_3\}$ is a basis for Span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$.

T 46. Let $H = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ and $K = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, where

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 3 \\ 0 \\ -1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 0 \\ 3 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} 2 \\ -3 \\ 6 \\ -5 \end{bmatrix},$$

$$\mathbf{v}_1 = \begin{bmatrix} -4 \\ 3 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 9 \\ -4 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -1 \\ 7 \\ 6 \\ 5 \end{bmatrix}$$

Find bases for H, K, and $H + K$. (See Exercises 41 and 42 in Section 4.1.)

T 47. Show that $\{t, \sin t, \cos 2t, \sin t \cos t\}$ is a linearly independent set of functions defined on $\mathbb{R}$. Start by assuming that

$$c_1 t + c_2 \sin t + c_3 \cos 2t + c_4 \sin t \cos t = 0 \qquad (5)$$

Equation (5) must hold for all real t, so choose several specific values of t (say, $t = 0, .1, .2$) until you get a system of enough equations to determine that all the c_j must be zero.

T 48. Show that $\{1, \cos t, \cos^2 t, \ldots, \cos^6 t\}$ is a linearly independent set of functions defined on $\mathbb{R}$. Use the method of Exercise 47. (This result will be needed in Exercise 54 in Section 4.5.)

4.5 The Dimension of a Vector Space

Theorem 9 in Section 4.4 implies that a vector space V with a basis $\mathcal{B}$ containing n vectors is isomorphic to $\mathbb{R}^n$. This section shows that this number n is an intrinsic property (called the dimension) of the space V that does not depend on the particular choice of basis. The discussion of dimension will give additional insight into properties of bases.

The first theorem generalizes a well-known result about the vector space $\mathbb{R}^n$.

THEOREM 10

If a vector space V has a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$, then any set in V containing more than n vectors must be linearly dependent.

PROOF Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ be a set in V with more than n vectors. The coordinate vectors $[\mathbf{u}_1]_{\mathcal{B}}, \ldots, [\mathbf{u}_p]_{\mathcal{B}}$ form a linearly dependent set in $\mathbb{R}^n$, because there are more vectors (p) than entries (n) in each vector. So there exist scalars $c_1, \ldots, c_p$, not all zero, such that

$$c_1[\mathbf{u}_1]_{\mathcal{B}} + \cdots + c_p[\mathbf{u}_p]_{\mathcal{B}} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \text{The zero vector in } \mathbb{R}^n$$

Since the coordinate mapping is a linear transformation,

$$\left[c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p\right]_{\mathcal{B}} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

The zero vector on the right displays the n weights needed to build the vector $c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$ from the basis vectors in $\mathcal{B}$. That is, $c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p = 0\mathbf{b}_1 + \cdots + 0\mathbf{b}_n = \mathbf{0}$. Since the c_i are not all zero, $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is linearly dependent.[1] ■

Theorem 10 implies that if a vector space V has a basis $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$, then each linearly independent set in V has no more than n vectors.

THEOREM 11

If a vector space V has a basis of n vectors, then every basis of V must consist of exactly n vectors.

PROOF Let $\mathcal{B}_1$ be a basis of n vectors and $\mathcal{B}_2$ be any other basis (of V). Since $\mathcal{B}_1$ is a basis and $\mathcal{B}_2$ is linearly independent, $\mathcal{B}_2$ has no more than n vectors, by Theorem 10. Also, since $\mathcal{B}_2$ is a basis and $\mathcal{B}_1$ is linearly independent, $\mathcal{B}_2$ has at least n vectors. Thus $\mathcal{B}_2$ consists of exactly n vectors. ■

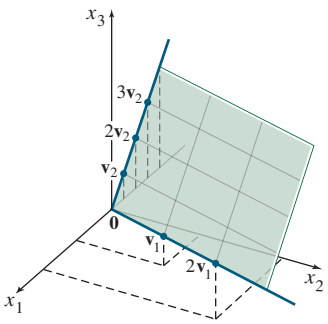
[1] Theorem 10 also applies to infinite sets in V. An infinite set is said to be linearly dependent if some finite subset is linearly dependent; otherwise, the set is linearly independent. If S is an infinite set in V, take any subset $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ of S, with $p > n$. The proof above shows that this subset is linearly dependent and hence so is S.

If a nonzero vector space V is spanned by a finite set S, then a subset of S is a basis for V, by the Spanning Set Theorem. In this case, Theorem 11 ensures that the following definition makes sense.

DEFINITION

If a vector space V is spanned by a finite set, then V is said to be **finite-dimensional**, and the **dimension** of V, written as $\dim V$, is the number of vectors in a basis for V. The dimension of the zero vector space $\{\mathbf{0}\}$ is defined to be zero. If V is not spanned by a finite set, then V is said to be **infinite-dimensional**.

EXAMPLE 1 The standard basis for $\mathbb{R}^n$ contains n vectors, so $\dim \mathbb{R}^n = n$. The standard polynomial basis $\{1, t, t^2\}$ shows that $\dim \mathbb{P}_2 = 3$. In general, $\dim \mathbb{P}_n = n + 1$. The space $\mathbb{P}$ of all polynomials is infinite-dimensional. ■



EXAMPLE 2 Let $H = \text{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$, where $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 6 \\ 2 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$. Then H is the plane studied in Example 7 in Section 4.4. A basis for H is $\{\mathbf{v}_1, \mathbf{v}_2\}$, since $\mathbf{v}_1$ and $\mathbf{v}_2$ are not multiples and hence are linearly independent. Thus $\dim H = 2$. ■

EXAMPLE 3 Find the dimension of the subspace

$$H = \left\{ \begin{bmatrix} a - 3b + 6c \\ 5a + 4d \\ b - 2c - d \\ 5d \end{bmatrix} : a, b, c, d \text{ in } \mathbb{R} \right\}$$

SOLUTION It is easy to see that H is the set of all linear combinations of the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 5 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -3 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 6 \\ 0 \\ -2 \\ 0 \end{bmatrix}, \quad \mathbf{v}_4 = \begin{bmatrix} 0 \\ 4 \\ -1 \\ 5 \end{bmatrix}$$

Clearly, $\mathbf{v}_1 \neq \mathbf{0}$, $\mathbf{v}_2$ is not a multiple of $\mathbf{v}_1$, but $\mathbf{v}_3$ is a multiple of $\mathbf{v}_2$. By the Spanning Set Theorem, we may discard $\mathbf{v}_3$ and still have a set that spans H. Finally, $\mathbf{v}_4$ is not a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_2$. So $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4\}$ is linearly independent (by Theorem 4 in Section 4.3) and hence is a basis for H. Thus $\dim H = 3$. ■

EXAMPLE 4 The subspaces of $\mathbb{R}^3$ can be classified by dimension. See Figure 1.

0-dimensional subspaces. Only the zero subspace.

1-dimensional subspaces. Any subspace spanned by a single nonzero vector. Such subspaces are lines through the origin.

2-dimensional subspaces. Any subspace spanned by two linearly independent vectors. Such subspaces are planes through the origin.

3-dimensional subspaces. Only $\mathbb{R}^3$ itself. Any three linearly independent vectors in $\mathbb{R}^3$ span all of $\mathbb{R}^3$, by the Invertible Matrix Theorem. ■

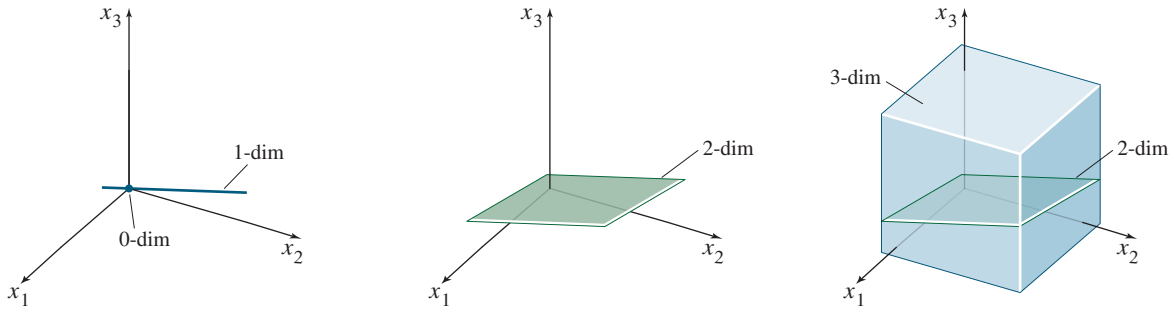


FIGURE 1 Sample subspaces of $\mathbb{R}^3$.

Subspaces of a Finite-Dimensional Space

The next theorem is a natural counterpart to the Spanning Set Theorem.

THEOREM 12

Let H be a subspace of a finite-dimensional vector space V. Any linearly independent set in H can be expanded, if necessary, to a basis for H. Also, H is finite-dimensional and

$$\dim H \leq \dim V$$

PROOF If $H = \{\mathbf{0}\}$, then certainly $\dim H = 0 \leq \dim V$. Otherwise, let $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be any linearly independent set in H. If S spans H, then S is a basis for H. Otherwise, there is some $\mathbf{u}_{k+1}$ in H that is not in Span S. But then $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}\}$ will be linearly independent, because no vector in the set can be a linear combination of vectors that precede it (by Theorem 4).

So long as the new set does not span H, we can continue this process of expanding S to a larger linearly independent set in H. But the number of vectors in a linearly independent expansion of S can never exceed the dimension of V, by Theorem 10. So eventually the expansion of S will span H and hence will be a basis for H, and $\dim H \leq \dim V$. ■

When the dimension of a vector space or subspace is known, the search for a basis is simplified by the next theorem. It says that if a set has the right number of elements, then one has only to show either that the set is linearly independent or that it spans the space. The theorem is of critical importance in numerous applied problems (involving differential equations or difference equations, for example) where linear independence is much easier to verify than spanning.

THEOREM 13

The Basis Theorem

Let V be a p-dimensional vector space, $p \geq 1$. Any linearly independent set of exactly p elements in V is automatically a basis for V. Any set of exactly p elements that spans V is automatically a basis for V.

PROOF By Theorem 12, a linearly independent set S of p elements can be extended to a basis for V. But that basis must contain exactly p elements, since $\dim V = p$. So S must already be a basis for V. Now suppose that S has p elements and spans V. Since V is nonzero, the Spanning Set Theorem implies that a subset S' of S is a basis of V. Since $\dim V = p$, S' must contain p vectors. Hence $S = S'$. ■

The Dimensions of Nul *A*, Col *A*, and Row *A*

Since the dimensions of the null space and column space of an $m \times n$ matrix are referred to frequently, they have specific names:

DEFINITION

The **rank** of an $m \times n$ matrix A is the dimension of the column space and the **nullity** of A is the dimension of the null space.

The pivot columns of a matrix A form a basis for Col A, so the rank of A is just the number of pivot columns. Since a basis for Row A can be found by taking the pivot rows from the row reduced echelon form of A, the dimension of Row A is also equal to the rank of A.

The nullity of A might seem to require more work, since finding a basis for Nul A usually takes more time than finding a basis for Col A. There is a shortcut: Let A be an $m \times n$ matrix, and suppose the equation $A\mathbf{x} = \mathbf{0}$ has k free variables. From Section 4.2, we know that the standard method of finding a spanning set for Nul A will produce exactly k linearly independent vectors—say, $\mathbf{u}_1, \ldots, \mathbf{u}_k$ – one for each free variable. So $\mathbf{u}_1, \ldots, \mathbf{u}_k$ is a basis for Nul A, and the number of free variables determines the size of the basis.

To summarize these facts for future reference:

The rank of an $m \times n$ matrix A is the number of pivot columns and the nullity of A is the number of free variables. Since the dimension of the row space is the number of pivot rows, it is also equal to the rank of A.

Putting these observations together results in the rank theorem.

THEOREM 14

The Rank Theorem

The dimensions of the column space and the null space of an $m \times n$ matrix A satisfy the equation

$$\text{rank } A + \text{nullity } A = \text{number of columns in } A$$

PROOF By Theorem 6 in Section 4.3, rank A is the number of pivot columns in A. The nullity of A equals the number of free variables in the equation $A\mathbf{x} = \mathbf{0}$. Expressed another way, the nullity of A is the number of columns of A that are *not* pivot columns. (It is the number of these columns, not the columns themselves, that is related to Nul A.) Obviously,

$$\left\{\begin{array}{c}\text{number of}\\ \text{pivot columns}\end{array}\right\} + \left\{\begin{array}{c}\text{number of}\\ \text{nonpivot columns}\end{array}\right\} = \left\{\begin{array}{c}\text{number of}\\ \text{columns}\end{array}\right\}$$

This proves the theorem. ■

EXAMPLE 5 Find the nullity and rank of

$$A = \begin{bmatrix} -3 & 6 & -1 & 1 & -7 \\ 1 & -2 & 2 & 3 & -1 \\ 2 & -4 & 5 & 8 & -4 \end{bmatrix}$$

SOLUTION Row reduce the augmented matrix $[A \ \mathbf{0}]$ to echelon form:

$$B = \begin{bmatrix} 1 & -2 & 2 & 3 & -1 & 0 \\ 0 & 0 & 1 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

There are three free variables: x_2, x_4, and x_5. Hence the nullity of A is 3. Also, the rank of A is 2 because A has two pivot columns. ■

The ideas behind Theorem 14 are visible in the calculations in Example 5. The two pivot positions in B, an echelon form of A, determine the basic variables and identify the basis vectors for Col A and those for Row A.

EXAMPLE 6

a. If A is a 7×9 matrix with nullity 2, what is the rank of A?

b. Could a 6×9 matrix have nullity 2?

SOLUTION

a. Since A has 9 columns, (rank A) $+ 2 = 9$, and hence rank $A = 7$.

b. No. If a 6×9 matrix, call it B, had a two-dimensional null space, it would have to have rank 7, by the Rank Theorem. But the columns of B are vectors in $\mathbb{R}^6$, and so the dimension of Col B cannot exceed 6; that is, rank B cannot exceed 6. ■

The next example provides a nice way to visualize the subspaces we have been studying. In Chapter 6, we will learn that Row A and Nul A have only the zero vector in common and are actually perpendicular to each other. The same fact applies to Row A^T (= Col$A$) and Nul A^T. So Figure 2, which accompanies Example 7, creates a good mental image for the general case.

EXAMPLE 7 Let $A = \begin{bmatrix} 3 & 0 & -1 \\ 3 & 0 & -1 \\ 4 & 0 & 5 \end{bmatrix}$. It is readily checked that Nul A is the x_2-axis, Row A is the x_1x_3-plane, Col A is the plane whose equation is $x_1 - x_2 = 0$, and Nul A^T is the set of all multiples of $(1, -1, 0)$. Figure 2 shows Nul A and Row A in the domain of the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$; the range of this mapping, Col A, is shown in a separate copy of $\mathbb{R}^3$, along with Nul A^T. ■

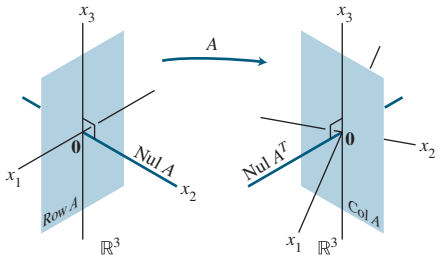


FIGURE 2 Subspaces determined by a matrix A.

Applications to Systems of Equations

The Rank Theorem is a powerful tool for processing information about systems of linear equations. The next example simulates the way a real-life problem using linear equations might be stated, without explicit mention of linear algebra terms such as matrix, subspace, and dimension.

EXAMPLE 8 A scientist has found two solutions to a homogeneous system of 40 equations in 42 variables. The two solutions are not multiples, and all other solutions can be constructed by adding together appropriate multiples of these two solutions. Can the scientist be *certain* that an associated nonhomogeneous system (with the same coefficients) has a solution?

SOLUTION Yes. Let A be the 40×42 coefficient matrix of the system. The given information implies that the two solutions are linearly independent and span Nul A. So nullity $A = 2$. By the Rank Theorem, rank $A = 42 - 2 = 40$. Since $\mathbb{R}^{40}$ is the only subspace of $\mathbb{R}^{40}$ whose dimension is 40, Col A must be all of $\mathbb{R}^{40}$. This means that every nonhomogeneous equation $A\mathbf{x} = \mathbf{b}$ has a solution. ■

Rank and the Invertible Matrix Theorem

The various vector space concepts associated with a matrix provide several more statements for the Invertible Matrix Theorem. The new statements listed here follow those in the original Invertible Matrix Theorem in Section 2.3 and other theorems in the text where statements have been added to it.

THEOREM

The Invertible Matrix Theorem (continued)

Let A be an $n \times n$ matrix. Then the following statements are each equivalent to the statement that A is an invertible matrix.

m. The columns of A form a basis of $\mathbb{R}^n$.

n. Col $A = \mathbb{R}^n$

o. rank $A = n$

p. nullity $A = 0$

q. Nul $A = \{\mathbf{0}\}$

PROOF Statement (m) is logically equivalent to statements (e) and (h) regarding linear independence and spanning. The other five statements are linked to the earlier ones of the theorem by the following chain of almost trivial implications:

$$(g) \Rightarrow (n) \Rightarrow (o) \Rightarrow (p) \Rightarrow (q) \Rightarrow (d)$$

Statement (g), which says that the equation $A\mathbf{x} = \mathbf{b}$ has at least one solution for each $\mathbf{b}$ in $\mathbb{R}^n$, implies (n), because Col A is precisely the set of all $\mathbf{b}$ such that the equation $A\mathbf{x} = \mathbf{b}$ is consistent. The implication (n) $\Rightarrow$ (o) follows from the definitions of dimension and rank. If the rank of A is n, the number of columns of A, then nullity $A = 0$, by the Rank Theorem, and so Nul $A = \{\mathbf{0}\}$. Thus (o) $\Rightarrow$ (p) $\Rightarrow$ (q). Also, (q) implies that the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, which is statement (d). Since statements (d) and (g) are already known to be equivalent to the statement that A is invertible, the proof is complete. ■

We have refrained from adding to the Invertible Matrix Theorem obvious statements about the row space of A, because the row space is the column space of A^T. Recall from statement (1) of the Invertible Matrix Theorem that A is invertible if and only if A^T is invertible. Hence every statement in the Invertible Matrix Theorem can also be stated for A^T. To do so would double the length of the theorem and produce a list of more than 30 statements!

Numerical Notes

Many algorithms discussed in this text are useful for understanding concepts and making simple computations by hand. However, the algorithms are often unsuitable for large-scale problems in real life.

Rank determination is a good example. It would seem easy to reduce a matrix to echelon form and count the pivots. But unless exact arithmetic is performed on a matrix whose entries are specified exactly, row operations can change the apparent rank of a matrix. For instance, if the value of x in the matrix $\begin{bmatrix} 5 & 7 \\ 5 & x \end{bmatrix}$ is not stored exactly as 7 in a computer, then the rank may be 1 or 2, depending on whether the computer treats $x - 7$ as zero.

In practical applications, the effective rank of a matrix A is often determined from the singular value decomposition of A, to be discussed in Section 7.4. This decomposition is also a reliable source of bases for Col A, Row A, Nul A, and Nul A^T.

Practice Problems

1. Decide whether each statement is True or False, and give a reason for each answer. Here V is a nonzero finite-dimensional vector space.

a. If $\dim V = p$ and if S is a linearly dependent subset of V, then S contains more than p vectors.

b. If S spans V and if T is a subset of V that contains more vectors than S, then T is linearly dependent.

2. Let H and K be subspaces of a vector space V. In Section 4.1, Exercise 40, it is established that $H \cap K$ is also a subspace of V. Prove $\dim (H \cap K) \le \dim H$.

4.5 Exercises

For each subspace in Exercises 1–8, (a) find a basis, and (b) state the dimension.

1. $\left\{ \begin{bmatrix} s - 2t \\ s + t \\ 3t \end{bmatrix} : s, t \text{ in } \mathbb{R} \right\}$

2. $\left\{ \begin{bmatrix} 5s \\ -t \\ -7s \end{bmatrix} : s, t \in \mathbb{R} \right\}$

3. $\left\{ \begin{bmatrix} 2c \\ a - b \\ b - 3c \\ a + 2b \end{bmatrix} : a, b, c \text{ in } \mathbb{R} \right\}$

4. $\left\{ \begin{bmatrix} a + b \\ 2a \\ 3a - b \\ -b \end{bmatrix} : a, b \text{ in } \mathbb{R} \right\}$

5. $\left\{ \begin{bmatrix} a - 4b - 2c \\ 2a + 5b - 4c \\ -a + 2c \\ -3a + 7b + 6c \end{bmatrix} : a, b, c \text{ in } \mathbb{R} \right\}$

6. $\left\{ \begin{bmatrix} 3a + 6b - c \\ 6a - 2b - 2c \\ -9a + 5b + 3c \\ -3a + b + c \end{bmatrix} : a, b, c \text{ in } \mathbb{R} \right\}$

7. $\{(a, b, c) : a - 3b + c = 0, b - 2c = 0, 2b - c = 0\}$

8. $\{(a, b, c, d) : a - 3b + c = 0\}$

In Exercises 9 and 10, find the dimension of the subspace spanned by the given vectors.

9. $\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 9 \\ 4 \\ -2 \end{bmatrix}, \begin{bmatrix} -7 \\ -3 \\ 1 \end{bmatrix}$

10. $\begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 4 \\ 1 \end{bmatrix}, \begin{bmatrix} -8 \\ 6 \\ 5 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 7 \end{bmatrix}$

Determine the dimensions of Nul A, Col A, and Row A for the matrices shown in Exercises 11–16.

11. $A = \begin{bmatrix} 1 & -6 & 9 & 0 & -2 \\ 0 & 1 & 2 & -4 & 5 \\ 0 & 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

12. $A = \begin{bmatrix} 1 & 3 & -4 & 2 & -1 & 6 \\ 0 & 0 & 1 & -3 & 7 & 0 \\ 0 & 0 & 0 & 1 & 4 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

13. $A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 0 & 0 & 1 & -6 \end{bmatrix}$

14. $A = \begin{bmatrix} 3 & 4 \\ -6 & 10 \end{bmatrix}$

15. $A = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 4 & 7 \\ 0 & 0 & 5 \end{bmatrix}$ **16.** $A = \begin{bmatrix} 1 & 4 & -1 \\ 0 & 7 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

In Exercises 17–26, V is a vector space and A is an $m \times n$ matrix. Mark each statement True or False **(T/F)**. Justify each answer.

17. **(T/F)** The number of pivot columns of a matrix equals the dimension of its column space.

18. **(T/F)** The number of variables in the equation $A\mathbf{x} = \mathbf{0}$ equals the nullity A.

19. **(T/F)** A plane in $\mathbb{R}^3$ is a two-dimensional subspace of $\mathbb{R}^3$.

20. **(T/F)** The dimension of the vector space $\mathbb{P}_4$ is 4.

21. **(T/F)** The dimension of the vector space of signals, $\mathbb{S}$, is 10.

22. **(T/F)** The dimensions of the row space and the column space of A are the same, even if A is not square.

23. **(T/F)** If B is any echelon form of A, then the pivot columns of B form a basis for the column space of A.

24. **(T/F)** The nullity of A is the number of columns of A that are not pivot columns.

25. **(T/F)** If a set $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ spans a finite-dimensional vector space V and if T is a set of more than p vectors in V, then T is linearly dependent.

26. **(T/F)** A vector space is infinite-dimensional if it is spanned by an infinite set.

27. The first four Hermite polynomials are 1, $2t$, $-2 + 4t^2$, and $-12t + 8t^3$. These polynomials arise naturally in the study of certain important differential equations in mathematical physics.[2] Show that the first four Hermite polynomials form a basis of $\mathbb{P}_3$.

28. The first four Laguerre polynomials are 1, $1 - t$, $2 - 4t + t^2$, and $6 - 18t + 9t^2 - t^3$. Show that these polynomials form a basis of $\mathbb{P}_3$.

29. Let $\mathcal{B}$ be the basis of $\mathbb{P}_3$ consisting of the Hermite polynomials in Exercise 27, and let $\mathbf{p}(t) = 7 - 12t - 8t^2 + 12t^3$. Find the coordinate vector of $\mathbf{p}$ relative to $\mathcal{B}$.

30. Let $\mathcal{B}$ be the basis of $\mathbb{P}_2$ consisting of the first three Laguerre polynomials listed in Exercise 28, and let $\mathbf{p}(t) = 7 - 8t + 3t^2$. Find the coordinate vector of $\mathbf{p}$ relative to $\mathcal{B}$.

31. Let S be a subset of an n-dimensional vector space V, and suppose S contains fewer than n vectors. Explain why S cannot span V.

32. Let H be an n-dimensional subspace of an n-dimensional vector space V. Show that $H = V$.

33. If a 4×7 matrix A has rank 4, find nullity A, rank A, and rank A^T.

34. If a 6×3 matrix A has rank 3, find nullity A, rank A, and rank A^T.

35. Suppose a 5×9 matrix A has four pivot columns. Is Col $A = \mathbb{R}^5$? Is Nul $A = \mathbb{R}^4$? Explain your answers.

36. Suppose a 5×6 matrix A has four pivot columns. What is nullity A? Is Col $A = \mathbb{R}^4$? Why or why not?

37. If the nullity of a 5×6 matrix A is 4, what are the dimensions of the column and row spaces of A?

38. If the nullity of a 7×6 matrix A is 5, what are the dimensions of the column and row spaces of A?

39. If A is a 7×5 matrix, what is the largest possible rank of A? If A is a 5×7 matrix, what is the largest possible rank of A? Explain your answers.

40. If A is a 4×3 matrix, what is the largest possible dimension of the row space of A? If A is a 3×4 matrix, what is the largest possible dimension of the row space of A? Explain.

41. Explain why the space $\mathbb{P}$ of all polynomials is an infinite-dimensional space.

42. Show that the space $C(\mathbb{R})$ of all continuous functions defined on the real line is an infinite-dimensional space.

In Exercises 43–48, V is a nonzero finite-dimensional vector space, and the vectors listed belong to V. Mark each statement True or False **(T/F)**. Justify each answer. (These questions are more difficult than those in Exercises 17–26.)

[2] See *Introduction to Functional Analysis*, 2nd ed., by A. E. Taylor and David C. Lay (New York: John Wiley & Sons, 1980), pp. 92–93. Other sets of polynomials are discussed there, too.

5 Eigenvalues and Eigenvectors



Introductory Example

DYNAMICAL SYSTEMS AND SPOTTED OWLS

In 1990, the northern spotted owl became the center of a nationwide controversy over the use and misuse of the majestic forests in the Pacific Northwest. Environmentalists convinced the federal government that the owl was threatened with extinction if logging continued in the old-growth forests (with trees more than 200 years old), where the owls prefer to live. The timber industry, anticipating the loss of 30,000–100,000 jobs as a result of new government restrictions on logging, argued that the owl should not be classified as a "threatened species" and cited a number of published scientific reports to support its case.[1]

The population of spotted owls continues to decline, and it remains a species caught in the crossfire between economic opportunities and conservation efforts. Mathematical ecologists help to analyze the effects on the spotted owl population of factors such as logging techniques, wildfires, and competition for habitat with the invasive barred owl. The life cycle of a spotted owl divides naturally into three stages: juvenile (up to 1 year old), subadult (1–2 years), and adult (older than 2 years). The owls mate for life during the subadult and adult stages, begin to breed as adults, and live for up to 20 years. Each owl pair requires about 1000 hectares (4 square miles) for its own home territory. A critical time in the life cycle is when the juveniles leave the nest. To survive and become a subadult, a juvenile must successfully find a new home range (and usually a mate).

A first step in studying the population dynamics is to model the population at yearly intervals, at times denoted by $k = 0, 1, 2, \ldots$. Usually, one assumes that there is a 1:1 ratio of males to females in each life stage and counts only the females. The population at year k can be described by a vector $\mathbf{x}_k = (j_k, s_k, a_k)$, where j_k, s_k, and a_k are the numbers of females in the juvenile, subadult, and adult stages, respectively.

Using actual field data from demographic studies, R. Lamberson and coworkers considered the following *stage-matrix model*:[2]

$$\begin{bmatrix} j_{k+1} \\ s_{k+1} \\ a_{k+1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & .33 \\ .18 & 0 & 0 \\ 0 & .71 & .94 \end{bmatrix} \begin{bmatrix} j_k \\ s_k \\ a_k \end{bmatrix}$$

Here the number of new juvenile females in year $k + 1$ is .33 times the number of adult females in year k (based on the average birth rate per owl pair). Also, 18% of the juveniles survive to become subadults, and 71% of the subadults and 94% of the adults survive to be counted as adults.

The stage-matrix model is a difference equation of the form $\mathbf{x}_{k+1} = A\mathbf{x}_k$. Such an equation is often called a **dynamical system** (or a **discrete linear dynamical**

[1] "The Great Spotted Owl War," *Reader's Digest*, November 1992, pp. 91–95.

[2] R. H. Lamberson, R. McKelvey, B. R. Noon, and C. Voss, "A Dynamic Analysis of the Viability of the Northern Spotted Owl in a Fragmented Forest Environment," *Conservation Biology* **6** (1992), 505–512. Also, a private communication from Professor Lamberson, 1993.

system) because it describes the changes in a system as time passes.

The 18% juvenile survival rate in the Lamberson stage matrix is the entry affected most by the amount of old-growth forest available. Actually, 60% of the juveniles normally survive to leave the nest, but in the Willow Creek region of California studied by Lamberson and his colleagues, only 30% of the juveniles that left the nest were able to find new home ranges. The rest perished during the search process.

A significant reason for the failure of owls to find new home ranges is the increasing fragmentation of old-growth timber stands due to clear-cutting of scattered areas on the old-growth land. When an owl leaves the protective canopy of the forest and crosses a clear-cut area, the risk of attack by predators increases dramatically. Section 5.6 will show that the model described in the chapter introduction predicts the eventual demise of the spotted owl, but that if 50% of the juveniles who survive to leave the nest also find new home ranges, then the owl population will thrive.

The goal of this chapter is to dissect the action of a linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ into elements that are easily visualized. All matrices in the chapter are square. The main applications described here are to discrete dynamical systems, differential equations, and Markov chains. However, the basic concepts—eigenvectors and eigenvalues—are useful throughout pure and applied mathematics, and they appear in settings far more general than we consider here. Eigenvalues are also used to study differential equations and *continuous* dynamical systems, they provide critical information in engineering design, and they arise naturally in fields such as physics and chemistry.

5.1 Eigenvectors and Eigenvalues

Although a transformation $\mathbf{x} \mapsto A\mathbf{x}$ may move vectors in a variety of directions, it often happens that there are special vectors on which the action of A is quite simple.

EXAMPLE 1 Let $A = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. The images of $\mathbf{u}$ and $\mathbf{v}$ under multiplication by A are shown in Figure 1. In fact, $A\mathbf{v}$ is just $2\mathbf{v}$. So A only "stretches" or dilates $\mathbf{v}$. ■

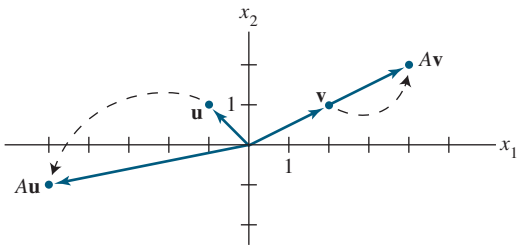


FIGURE 1 Effects of multiplication by A.

This section studies equations such as

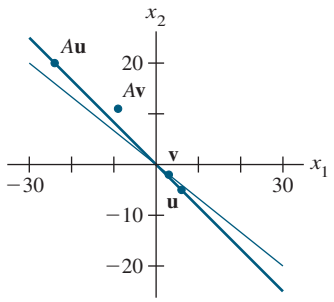
$$A\mathbf{x} = 2\mathbf{x} \quad \text{or} \quad A\mathbf{x} = -4\mathbf{x}$$

where special vectors are transformed by A into scalar multiples of themselves.

DEFINITION

An **eigenvector** of an $n \times n$ matrix A is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ. A scalar λ is called an **eigenvalue** of A if there is a nontrivial solution $\mathbf{x}$ of $A\mathbf{x} = \lambda\mathbf{x}$; such an $\mathbf{x}$ is called an *eigenvector corresponding to* λ.[1]

It is easy to determine if a given vector is an eigenvector of a matrix. See Example 2. It is also easy to decide if a specified scalar is an eigenvalue. See Example 3.



$A\mathbf{u} = -4\mathbf{u}$, but $A\mathbf{v} \neq \lambda\mathbf{v}$.

EXAMPLE 2 Let $A = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}$, $\mathbf{u} = \begin{bmatrix} 6 \\ -5 \end{bmatrix}$, and $\mathbf{v} = \begin{bmatrix} 3 \\ -2 \end{bmatrix}$. Are $\mathbf{u}$ and $\mathbf{v}$ eigenvectors of A?

SOLUTION

$$A\mathbf{u} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}\begin{bmatrix} 6 \\ -5 \end{bmatrix} = \begin{bmatrix} -24 \\ 20 \end{bmatrix} = -4\begin{bmatrix} 6 \\ -5 \end{bmatrix} = -4\mathbf{u}$$

$$A\mathbf{v} = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix}\begin{bmatrix} 3 \\ -2 \end{bmatrix} = \begin{bmatrix} -9 \\ 11 \end{bmatrix} \neq \lambda\begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

Thus $\mathbf{u}$ is an eigenvector corresponding to an eigenvalue (-4), but $\mathbf{v}$ is not an eigenvector of A, because $A\mathbf{v}$ is not a multiple of $\mathbf{v}$. ■

EXAMPLE 3 Show that 7 is an eigenvalue of matrix A in Example 2, and find the corresponding eigenvectors.

SOLUTION The scalar 7 is an eigenvalue of A if and only if the equation

$$A\mathbf{x} = 7\mathbf{x} \tag{1}$$

has a nontrivial solution. But (1) is equivalent to $A\mathbf{x} - 7\mathbf{x} = \mathbf{0}$, or

$$(A - 7I)\mathbf{x} = \mathbf{0} \tag{2}$$

To solve this homogeneous equation, form the matrix

$$A - 7I = \begin{bmatrix} 1 & 6 \\ 5 & 2 \end{bmatrix} - \begin{bmatrix} 7 & 0 \\ 0 & 7 \end{bmatrix} = \begin{bmatrix} -6 & 6 \\ 5 & -5 \end{bmatrix}$$

The columns of $A - 7I$ are obviously linearly dependent, so (2) has nontrivial solutions. Thus 7 *is* an eigenvalue of A. To find the corresponding eigenvectors, use row operations:

$$\begin{bmatrix} -6 & 6 & 0 \\ 5 & -5 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The general solution has the form $x_2\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Each vector of this form with $x_2 \neq 0$ is an eigenvector corresponding to $\lambda = 7$. ■

[1] Note that an eigenvector must be *nonzero*, by definition, but an eigenvalue may be zero. The case in which the number 0 is an eigenvalue is discussed after Example 5.

Warning: Although row reduction was used in Example 3 to find eigen*vectors*, it cannot be used to find eigen*values*. An echelon form of a matrix A usually does *not* display the eigenvalues of A.

The equivalence of equations (1) and (2) obviously holds for any λ in place of $\lambda = 7$. Thus λ is an eigenvalue of an $n \times n$ matrix A if and only if the equation

$$(A - \lambda I)\mathbf{x} = \mathbf{0} \tag{3}$$

has a nontrivial solution. The set of *all* solutions of (3) is just the null space of the matrix $A - \lambda I$. So this set is a *subspace* of $\mathbb{R}^n$ and is called the **eigenspace** of A corresponding to λ. The eigenspace consists of the zero vector and all the eigenvectors corresponding to λ.

Example 3 shows that for matrix A in Example 2, the eigenspace corresponding to $\lambda = 7$ consists of *all* multiples of $(1, 1)$, which is the line through $(1, 1)$ and the origin. From Example 2, you can check that the eigenspace corresponding to $\lambda = -4$ is the line through $(6, -5)$. These eigenspaces are shown in Figure 2, along with eigenvectors $(1, 1)$ and $(3/2, -5/4)$ and the geometric action of the transformation $\mathbf{x} \mapsto A\mathbf{x}$ on each eigenspace.

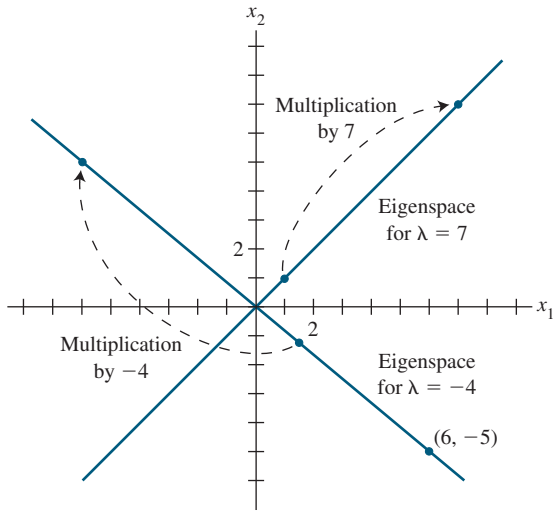


FIGURE 2 Eigenspaces for $\lambda = -4$ and $\lambda = 7$.

EXAMPLE 4 Let $A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}$. An eigenvalue of A is 2. Find a basis for the corresponding eigenspace.

SOLUTION Form

$$A - 2I = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix} - \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 6 \\ 2 & -1 & 6 \\ 2 & -1 & 6 \end{bmatrix}$$

and row reduce the augmented matrix for $(A - 2I)\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} 2 & -1 & 6 & 0 \\ 2 & -1 & 6 & 0 \\ 2 & -1 & 6 & 0 \end{bmatrix} \sim \begin{bmatrix} 2 & -1 & 6 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

At this point, it is clear that 2 is indeed an eigenvalue of A because the equation $(A - 2I)\mathbf{x} = \mathbf{0}$ has free variables. The general solution is

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x_2 \begin{bmatrix} 1/2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}, \quad x_2 \text{ and } x_3 \text{ free}$$

The eigenspace, shown in Figure 3, is a two-dimensional subspace of $\mathbb{R}^3$. A basis is

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} \right\}$$

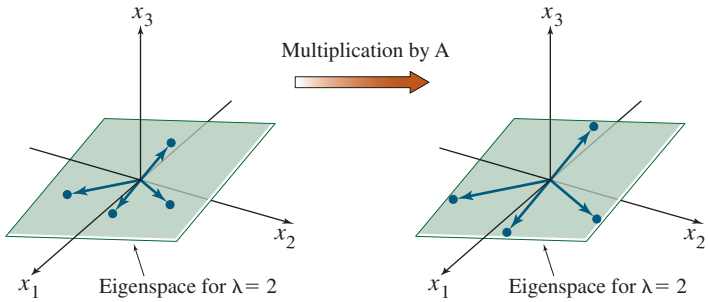


FIGURE 3 A acts as a dilation on the eigenspace.

Reasonable Answers

Remember that once you find a potential eigenvector $\mathbf{v}$, you can easily check your answer: just find $A\mathbf{v}$ and see if it is a multiple of $\mathbf{v}$. For example, to check whether $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is an eigenvector of $A = \begin{bmatrix} 1 & 2 \\ -1 & -2 \end{bmatrix}$, notice $A\mathbf{v} = \begin{bmatrix} 3 \\ -3 \end{bmatrix}$, which is not a multiple of $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, establishing that $\mathbf{v}$ is not an eigenvector. It turns out we had a sign error. The vector $\mathbf{u} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ is a correct eigenvector for A since $A\mathbf{u} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} = -1 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -1\,\mathbf{u}$.

Numerical Notes

Example 4 shows a good method for manual computation of eigenvectors in simple cases when an eigenvalue is known. Using a matrix program and row reduction to find an eigenspace (for a specified eigenvalue) usually works, too, but this is not entirely reliable. Roundoff error can lead occasionally to a reduced echelon form with the wrong number of pivots. The best computer programs compute approximations for eigenvalues and eigenvectors simultaneously, to any desired degree of accuracy, for matrices that are not too large. The size of matrices that can be analyzed increases each year as computing power and software improve.

The following theorem describes one of the few special cases in which eigenvalues can be found precisely. Calculation of eigenvalues will also be discussed in Section 5.2.

THEOREM 1 The eigenvalues of a triangular matrix are the entries on its main diagonal.

PROOF For simplicity, consider the 3×3 case. If A is upper triangular, then $A - \lambda I$ has the form

$$A - \lambda I = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix} - \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}$$
$$= \begin{bmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ 0 & a_{22} - \lambda & a_{23} \\ 0 & 0 & a_{33} - \lambda \end{bmatrix}$$

The scalar λ is an eigenvalue of A if and only if the equation $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has a nontrivial solution, that is, if and only if the equation has a free variable. Because of the zero entries in $A - \lambda I$, it is easy to see that $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has a free variable if and only if at least one of the entries on the diagonal of $A - \lambda I$ is zero. This happens if and only if λ equals one of the entries a_{11}, a_{22}, a_{33} in A. For the case in which A is lower triangular, see Exercise 36. ■

EXAMPLE 5 Let $A = \begin{bmatrix} 3 & 6 & -8 \\ 0 & 0 & 6 \\ 0 & 0 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 0 & 0 \\ -2 & 1 & 0 \\ 5 & 3 & 4 \end{bmatrix}$. The eigenvalues of A are 3, 0, and 2. The eigenvalues of B are 4 and 1. ■

What does it mean for a matrix A to have an eigenvalue of 0, such as in Example 5? This happens if and only if the equation

$$A\mathbf{x} = 0\mathbf{x} \tag{4}$$

has a nontrivial solution. But (4) is equivalent to $A\mathbf{x} = \mathbf{0}$, which has a nontrivial solution if and only if A is not invertible. Thus 0 *is an eigenvalue of A if and only if A is not invertible*. This fact will be added to the Invertible Matrix Theorem in Section 5.2.

The following important theorem will be needed later. Its proof illustrates a typical calculation with eigenvectors. One way to prove the statement "If P then Q" is to show that P and the negation of Q leads to a contradiction. This strategy is used in the proof of the theorem.

THEOREM 2 If $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are eigenvectors that correspond to distinct eigenvalues $\lambda_1, \ldots, \lambda_r$ of an $n \times n$ matrix A, then the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is linearly independent.

PROOF Suppose $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is linearly dependent. Since $\mathbf{v}_1$ is nonzero, Theorem 7 in Section 1.7 says that one of the vectors in the set is a linear combination of the preceding vectors. Let p be the least index such that $\mathbf{v}_{p+1}$ is a linear combination of the preceding (linearly independent) vectors. Then there exist scalars $c_1, \ldots, c_p$ such that

$$c_1\mathbf{v}_1 + \cdots + c_p\mathbf{v}_p = \mathbf{v}_{p+1} \tag{5}$$

Multiplying both sides of (5) by A and using the fact that $A\mathbf{v}_k = \lambda_k \mathbf{v}_k$ for each k, we obtain

$$\begin{aligned} c_1 A\mathbf{v}_1 + \cdots + c_p A\mathbf{v}_p &= A\mathbf{v}_{p+1} \\ c_1\lambda_1\mathbf{v}_1 + \cdots + c_p\lambda_p\mathbf{v}_p &= \lambda_{p+1}\mathbf{v}_{p+1} \end{aligned} \tag{6}$$

Multiplying both sides of (5) by λ_{p+1} and subtracting the result from (6), we have

$$c_1(\lambda_1 - \lambda_{p+1})\mathbf{v}_1 + \cdots + c_p(\lambda_p - \lambda_{p+1})\mathbf{v}_p = \mathbf{0} \tag{7}$$

Since $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is linearly independent, the weights in (7) are all zero. But none of the factors $\lambda_i - \lambda_{p+1}$ are zero, because the eigenvalues are distinct. Hence $c_i = 0$ for $i = 1, \ldots, p$. But then (5) says that $\mathbf{v}_{p+1} = \mathbf{0}$, which is impossible. Hence $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ cannot be linearly dependent and therefore must be linearly independent. ■

Eigenvectors and Difference Equations

This section concludes by showing how to construct solutions of the first-order difference equation discussed in the chapter introductory example:

$$\mathbf{x}_{k+1} = A\mathbf{x}_k \quad (k = 0, 1, 2, \ldots) \tag{8}$$

If A is an $n \times n$ matrix, then (8) is a *recursive* description of a sequence $\{\mathbf{x}_k\}$ in $\mathbb{R}^n$. A **solution** of (8) is an explicit description of $\{\mathbf{x}_k\}$ whose formula for each $\mathbf{x}_k$ does not depend directly on A or on the preceding terms in the sequence other than the initial term $\mathbf{x}_0$.

The simplest way to build a solution of (8) is to take an eigenvector $\mathbf{x}_0$ and its corresponding eigenvalue λ and let

$$\mathbf{x}_k = \lambda^k \mathbf{x}_0 \quad (k = 1, 2, \ldots) \tag{9}$$

This sequence is a solution because

$$A\mathbf{x}_k = A(\lambda^k \mathbf{x}_0) = \lambda^k(A\mathbf{x}_0) = \lambda^k(\lambda\mathbf{x}_0) = \lambda^{k+1}\mathbf{x}_0 = \mathbf{x}_{k+1}$$

Linear combinations of solutions in the form of equation (9) are solutions, too! See Exercise 41.

Practice Problems

1. Is 5 an eigenvalue of $A = \begin{bmatrix} 6 & -3 & 1 \\ 3 & 0 & 5 \\ 2 & 2 & 6 \end{bmatrix}$?

2. If $\mathbf{x}$ is an eigenvector of A corresponding to λ, what is $A^3\mathbf{x}$?

3. Suppose that $\mathbf{b}_1$ and $\mathbf{b}_2$ are eigenvectors corresponding to distinct eigenvalues λ_1 and λ_2, respectively, and suppose that $\mathbf{b}_3$ and $\mathbf{b}_4$ are linearly independent eigenvectors corresponding to a third distinct eigenvalue λ_3. Does it necessarily follow that $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4\}$ is a linearly independent set? [*Hint:* Consider the equation $c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + (c_3\mathbf{b}_3 + c_4\mathbf{b}_4) = \mathbf{0}$.]

4. If A is an $n \times n$ matrix and λ is an eigenvalue of A, show that 2λ is an eigenvalue of $2A$.

5.1 Exercises

1. Is $\lambda = 2$ an eigenvalue of $\begin{bmatrix} 3 & 2 \\ 3 & 8 \end{bmatrix}$? Why or why not?

2. Is $\lambda = -2$ an eigenvalue of $\begin{bmatrix} 7 & 3 \\ 3 & -1 \end{bmatrix}$? Why or why not?

3. Is $\begin{bmatrix} 1 \\ 4 \end{bmatrix}$ an eigenvector of $\begin{bmatrix} -3 & 1 \\ -3 & 8 \end{bmatrix}$? If so, find the eigenvalue.

4. Is $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ an eigenvector of $\begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$? If so, find the eigenvalue.

5. Is $\begin{bmatrix} 4 \\ -3 \\ 1 \end{bmatrix}$ an eigenvector of $\begin{bmatrix} 3 & 7 & 9 \\ -4 & -5 & 1 \\ 2 & 4 & 4 \end{bmatrix}$? If so, find the eigenvalue.

6. Is $\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$ an eigenvector of $\begin{bmatrix} 2 & 6 & 7 \\ 3 & 2 & 7 \\ 5 & 6 & 4 \end{bmatrix}$? If so, find the eigenvalue.

7. Is $\lambda = 4$ an eigenvalue of $\begin{bmatrix} 3 & 0 & -1 \\ 2 & 3 & 1 \\ -3 & 4 & 5 \end{bmatrix}$? If so, find one corresponding eigenvector.

8. Is $\lambda = 3$ an eigenvalue of $\begin{bmatrix} 1 & 2 & 2 \\ 3 & -2 & 1 \\ 0 & 1 & 1 \end{bmatrix}$? If so, find one corresponding eigenvector.

In Exercises 9–16, find a basis for the eigenspace corresponding to each listed eigenvalue.

9. $A = \begin{bmatrix} 9 & 0 \\ 2 & 3 \end{bmatrix}, \lambda = 3, 9$

10. $A = \begin{bmatrix} 14 & -4 \\ 16 & -2 \end{bmatrix}, \lambda = 6$

11. $A = \begin{bmatrix} 4 & -2 \\ -3 & 9 \end{bmatrix}, \lambda = 10$

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

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

14. $A = \begin{bmatrix} 3 & -1 & 3 \\ -1 & 3 & 3 \\ 6 & 6 & 2 \end{bmatrix}, \lambda = -4$

15. $A = \begin{bmatrix} 8 & 3 & -4 \\ -1 & 4 & 4 \\ 2 & 6 & -1 \end{bmatrix}, \lambda = 7$

16. $A = \begin{bmatrix} 3 & 0 & 2 & 0 \\ 1 & 3 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}, \lambda = 4$

Find the eigenvalues of the matrices in Exercises 17 and 18.

17. $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 5 \\ 0 & 0 & -1 \end{bmatrix}$

18. $\begin{bmatrix} 8 & 0 & 0 \\ -7 & 0 & 0 \\ 6 & -5 & -4 \end{bmatrix}$

19. For $A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}$, find one eigenvalue, with no calculation. Justify your answer.

20. Without calculation, find one eigenvalue and two linearly independent eigenvectors of $A = \begin{bmatrix} 4 & 4 & -4 \\ 4 & 4 & -4 \\ 4 & 4 & -4 \end{bmatrix}$. Justify your answer.

In Exercises 21–30, A is an $n \times n$ matrix. Mark each statement True or False **(T/F)**. Justify each answer.

21. **(T/F)** If $A\mathbf{x} = \lambda\mathbf{x}$ for some vector $\mathbf{x}$, then λ is an eigenvalue of A.

22. **(T/F)** If $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ, then $\mathbf{x}$ is an eigenvector of A.

23. **(T/F)** A matrix A is invertible if and only if 0 is an eigenvalue of A.

24. **(T/F)** A number c is an eigenvalue of A if and only if the equation $(A - cI)\mathbf{x} = 0$ has a nontrivial solution.

25. **(T/F)** Finding an eigenvector of A may be difficult, but checking whether a given vector is in fact an eigenvector is easy.

26. **(T/F)** To find the eigenvalues of A, reduce A to echelon form.

27. **(T/F)** If $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent eigenvectors, then they correspond to distinct eigenvalues.

28. **(T/F)** The eigenvalues of a matrix are on its main diagonal.

29. **(T/F)** If $\mathbf{v}$ is an eigenvector with eigenvalue 2, then $2\mathbf{v}$ is an eigenvector with eigenvalue 4.

30. **(T/F)** An eigenspace of A is a null space of a certain matrix.

31. Explain why a 2×2 matrix can have at most two distinct eigenvalues. Explain why an $n \times n$ matrix can have at most n distinct eigenvalues.

32. Construct an example of a 2×2 matrix with only one distinct eigenvalue.

33. Let λ be an eigenvalue of an invertible matrix A. Show that λ^{-1} is an eigenvalue of A^{-1}. [*Hint:* Suppose a nonzero $\mathbf{x}$ satisfies $A\mathbf{x} = \lambda\mathbf{x}$.]

34. Show that if A^2 is the zero matrix, then the only eigenvalue of A is 0.

35. Show that λ is an eigenvalue of A if and only if λ is an eigenvalue of A^T. [*Hint:* Find out how $A - \lambda I$ and $A^T - \lambda I$ are related.]

36. Use Exercise 35 to complete the proof of Theorem 1 for the case when A is lower triangular.

37. Consider an $n \times n$ matrix A with the property that the row sums all equal the same number s. Show that s is an eigenvalue of A. [*Hint:* Find an eigenvector.]

38. Consider an $n \times n$ matrix A with the property that the column sums all equal the same number s. Show that s is an eigenvalue of A. [*Hint:* Use Exercises 35 and 37.]

In Exercises 39 and 40, let A be the matrix of the linear transformation T. Without writing A, find an eigenvalue of A and describe the eigenspace.

39. T is the transformation on $\mathbb{R}^2$ that reflects points across some line through the origin.

40. T is the transformation on $\mathbb{R}^3$ that rotates points about some line through the origin.

41. Let $\mathbf{u}$ and $\mathbf{v}$ be eigenvectors of a matrix A, with corresponding eigenvalues λ and μ, and let c_1 and c_2 be scalars. Define

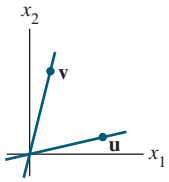
$$\mathbf{x}_k = c_1\lambda^k\mathbf{u} + c_2\mu^k\mathbf{v} \quad (k = 0, 1, 2, \ldots)$$

a. What is $\mathbf{x}_{k+1}$, by definition?

b. Compute $A\mathbf{x}_k$ from the formula for $\mathbf{x}_k$, and show that $A\mathbf{x}_k = \mathbf{x}_{k+1}$. This calculation will prove that the sequence $\{\mathbf{x}_k\}$ defined above satisfies the difference equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$ $(k = 0, 1, 2, \ldots)$.

42. Describe how you might try to build a solution of a difference equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$ $(k = 0, 1, 2, \ldots)$ if you were given the initial $\mathbf{x}_0$ and this vector did not happen to be an eigenvector of A. [*Hint:* How might you relate $\mathbf{x}_0$ to eigenvectors of A?]

43. Let $\mathbf{u}$ and $\mathbf{v}$ be the vectors shown in the figure, and suppose $\mathbf{u}$ and $\mathbf{v}$ are eigenvectors of a 2×2 matrix A that correspond to eigenvalues 2 and 3, respectively. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation given by $T(\mathbf{x}) = A\mathbf{x}$ for each $\mathbf{x}$ in $\mathbb{R}^2$, and let $\mathbf{w} = \mathbf{u} + \mathbf{v}$. Make a copy of the figure, and on the same coordinate system, carefully plot the vectors $T(\mathbf{u})$, $T(\mathbf{v})$, and $T(\mathbf{w})$.



44. Repeat Exercise 43, assuming $\mathbf{u}$ and $\mathbf{v}$ are eigenvectors of A that correspond to eigenvalues -1 and 3, respectively.

T In Exercises 45–48, use a matrix program to find the eigenvalues of the matrix. Then use the method of Example 4 with a row reduction routine to produce a basis for each eigenspace.

45. $\begin{bmatrix} 8 & -10 & -5 \\ 2 & 17 & 2 \\ -9 & -18 & 4 \end{bmatrix}$

46. $\begin{bmatrix} 9 & -4 & -2 & -4 \\ -56 & 32 & -28 & 44 \\ -14 & -14 & 6 & -14 \\ 42 & -33 & 21 & -45 \end{bmatrix}$

47. $\begin{bmatrix} 4 & -9 & -7 & 8 & 2 \\ -7 & -9 & 0 & 7 & 14 \\ 5 & 10 & 5 & -5 & -10 \\ -2 & 3 & 7 & 0 & 4 \\ -3 & -13 & -7 & 10 & 11 \end{bmatrix}$

48. $\begin{bmatrix} -4 & -4 & 20 & -8 & -1 \\ 14 & 12 & 46 & 18 & 2 \\ 6 & 4 & -18 & 8 & 1 \\ 11 & 7 & -37 & 17 & 2 \\ 18 & 12 & -60 & 24 & 5 \end{bmatrix}$

Solutions to Practice Problems

1. The number 5 is an eigenvalue of A if and only if the equation $(A - 5I)\mathbf{x} = \mathbf{0}$ has a nontrivial solution. Form

$$A - 5I = \begin{bmatrix} 6 & -3 & 1 \\ 3 & 0 & 5 \\ 2 & 2 & 6 \end{bmatrix} - \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix} = \begin{bmatrix} 1 & -3 & 1 \\ 3 & -5 & 5 \\ 2 & 2 & 1 \end{bmatrix}$$

and row reduce the augmented matrix:

$$\begin{bmatrix} 1 & -3 & 1 & 0 \\ 3 & -5 & 5 & 0 \\ 2 & 2 & 1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 1 & 0 \\ 0 & 4 & 2 & 0 \\ 0 & 8 & -1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 1 & 0 \\ 0 & 4 & 2 & 0 \\ 0 & 0 & -5 & 0 \end{bmatrix}$$

Solutions to Practice Problems (Continued)

At this point, it is clear that the homogeneous system has no free variables. Thus $A - 5I$ is an invertible matrix, which means that 5 is *not* an eigenvalue of A.

2. If $\mathbf{x}$ is an eigenvector of A corresponding to λ, then $A\mathbf{x} = \lambda\mathbf{x}$ and so

$$A^2\mathbf{x} = A(\lambda\mathbf{x}) = \lambda A\mathbf{x} = \lambda^2\mathbf{x}$$

Again, $A^3\mathbf{x} = A(A^2\mathbf{x}) = A(\lambda^2\mathbf{x}) = \lambda^2 A\mathbf{x} = \lambda^3\mathbf{x}$. The general pattern, $A^k\mathbf{x} = \lambda^k\mathbf{x}$, is proved by induction.

3. Yes. Suppose $c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + (c_3\mathbf{b}_3 + c_4\mathbf{b}_4) = \mathbf{0}$. Since any linear combination of eigenvectors corresponding to the same eigenvalue is in the eigenspace for that eigenvalue, $c_3\mathbf{b}_3 + c_4\mathbf{b}_4$ is either $\mathbf{0}$ or an eigenvector for λ_3. If $c_3\mathbf{b}_3 + c_4\mathbf{b}_4$ were an eigenvector for λ_3, then by Theorem 2, $\{\mathbf{b}_1, \mathbf{b}_2, c_3\mathbf{b}_3 + c_4\mathbf{b}_4\}$ would be a linearly independent set, which would force $c_1 = c_2 = 0$ and $c_3\mathbf{b}_3 + c_4\mathbf{b}_4 = \mathbf{0}$, contradicting that $c_3\mathbf{b}_3 + c_4\mathbf{b}_4$ is an eigenvector. Thus $c_3\mathbf{b}_3 + c_4\mathbf{b}_4$ must be $\mathbf{0}$, implying that $c_1\mathbf{b}_1 + c_2\mathbf{b}_2 = \mathbf{0}$ also. By Theorem 2, $\{\mathbf{b}_1, \mathbf{b}_2\}$ is a linearly independent set so $c_1 = c_2 = 0$. Moreover, $\{\mathbf{b}_3, \mathbf{b}_4\}$ is a linearly independent set so $c_3 = c_4 = 0$. Since all of the coefficients c_1, c_2, c_3, and c_4 must be zero, it follows that $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3, \mathbf{b}_4\}$ is a linearly independent set.

4. Since λ is an eigenvalue of A, there is a nonzero vector $\mathbf{x}$ in $\mathbb{R}^n$ such that $A\mathbf{x} = \lambda\mathbf{x}$. Multiplying both sides of this equation by 2 results in the equation $2(A\mathbf{x}) = 2(\lambda\mathbf{x})$. Thus $(2A)\mathbf{x} = (2\lambda)\mathbf{x}$ and hence 2λ is an eigenvalue of $2A$.

5.2 The Characteristic Equation

Useful information about the eigenvalues of a square matrix A is encoded in a special scalar equation called the characteristic equation of A. A simple example will lead to the general case.

EXAMPLE 1 Find the eigenvalues of $A = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}$.

SOLUTION We must find all scalars λ such that the matrix equation

$$(A - \lambda I)\mathbf{x} = \mathbf{0}$$

has a nontrivial solution. By the Invertible Matrix Theorem in Section 2.3, this problem is equivalent to finding all λ such that the matrix $A - \lambda I$ is *not* invertible, where

$$A - \lambda I = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 2-\lambda & 3 \\ 3 & -6-\lambda \end{bmatrix}$$

By Theorem 4 in Section 2.2, this matrix fails to be invertible precisely when its determinant is zero. So the eigenvalues of A are the solutions of the equation

$$\det(A - \lambda I) = \det \begin{bmatrix} 2-\lambda & 3 \\ 3 & -6-\lambda \end{bmatrix} = 0$$

Recall that

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

So

$$\begin{aligned}\det(A-\lambda I) &= (2-\lambda)(-6-\lambda)-(3)(3)\\ &= -12+6\lambda-2\lambda+\lambda^2-9\\ &= \lambda^2+4\lambda-21\\ &= (\lambda-3)(\lambda+7)\end{aligned}$$

If $\det(A-\lambda I)=0$, then $\lambda=3$ or $\lambda=-7$. So the eigenvalues of A are 3 and -7. ■

Determinants

The determinant in Example 1 transformed the matrix equation $(A-\lambda I)\mathbf{x}=\mathbf{0}$, which involves *two* unknowns λ and $\mathbf{x}$, into the scalar equation $\lambda^2+4\lambda-21=0$, which involves only *one* unknown. The same idea works for $n\times n$ matrices.

Before turning to larger matrices, recall from Section 3.1 that the matrix A_{ij} is obtained from A by deleting the ith row and jth column. The determinant of an $n\times n$ matrix A can be computed by an expansion across any row or down any column. The expansion across the ith row is given by

$$\det A=(-1)^{i+1}a_{i1}\det A_{i1}+(-1)^{i+2}a_{i2}\det A_{i2}+\cdots+(-1)^{i+n}a_{in}\det A_{in}$$

The expansion down the jth column is given by

$$\det A=(-1)^{1+j}a_{1j}\det A_{1j}+(-1)^{2+j}a_{2j}\det A_{2j}+\cdots+(-1)^{n+j}a_{nj}\det A_{nj}$$

EXAMPLE 2 Compute the determinant of

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

SOLUTION Any row or column can be chosen for the expansion. For example, expanding down the first column of A results in

$$\begin{aligned}\det A &= a_{11}\det A_{11}-a_{21}\det A_{21}+a_{31}\det A_{31}\\ &= 2\det\begin{bmatrix}0 & -1\\ 2 & 1\end{bmatrix}-4\det\begin{bmatrix}3 & 1\\ 2 & 1\end{bmatrix}+0\det\begin{bmatrix}3 & 1\\ 0 & -1\end{bmatrix}\\ &= 2(0-(-2))-4(3-2)+0(-3-0)=0\end{aligned}$$

The next theorem lists facts from Sections 3.1 and 3.2 and is included here for convenient reference. ■

THEOREM 3

Properties of Determinants

Let A and B be $n\times n$ matrices.

a. A is invertible if and only if $\det A\neq 0$.

b. $\det AB=(\det A)(\det B)$.

c. $\det A^T=\det A$.

d. If A is triangular, then $\det A$ is the product of the entries on the main diagonal of A.

e. A row replacement operation on A does not change the determinant. A row interchange changes the sign of the determinant. A row scaling also scales the determinant by the same scalar factor.

Recall that A is invertible if and only if the equation $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Notice that the number 0 is an eigenvalue of A if and only if there *is* a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = 0\mathbf{x} = \mathbf{0}$, which happens if and only if $0 = \det(A - 0I) = \det A$. Hence A is invertible if and only if 0 is *not* an eigenvalue.

THEOREM

The Invertible Matrix Theorem (continued)

Let A be an $n \times n$ matrix. Then A is invertible if and only if

r. The number 0 is *not* an eigenvalue of A.

The Characteristic Equation

Theorem 3(a) shows how to determine when a matrix of the form $A - \lambda I$ is *not* invertible. The scalar equation $\det(A - \lambda I) = 0$ is called the **characteristic equation** of A, and the argument in Example 1 justifies the following fact.

A scalar λ is an eigenvalue of an $n \times n$ matrix A if and only if λ satisfies the characteristic equation

$$\det(A - \lambda I) = 0$$

EXAMPLE 3 Find the characteristic equation of

$$A = \begin{bmatrix} 5 & -2 & 6 & -1 \\ 0 & 3 & -8 & 0 \\ 0 & 0 & 5 & 4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

SOLUTION Form $A - \lambda I$, and use Theorem 3(d):

$$\det(A - \lambda I) = \det \begin{bmatrix} 5-\lambda & -2 & 6 & -1 \\ 0 & 3-\lambda & -8 & 0 \\ 0 & 0 & 5-\lambda & 4 \\ 0 & 0 & 0 & 1-\lambda \end{bmatrix}$$
$$= (5 - \lambda)(3 - \lambda)(5 - \lambda)(1 - \lambda)$$

The characteristic equation is

$$(5 - \lambda)^2(3 - \lambda)(1 - \lambda) = 0$$

or

$$(\lambda - 5)^2(\lambda - 3)(\lambda - 1) = 0$$

Expanding the product, we can also write

$$\lambda^4 - 14\lambda^3 + 68\lambda^2 - 130\lambda + 75 = 0$$ ■

Reasonable Answers

If you want to verify λ is an eigenvalue of A, row reduce $A - \lambda I$. If you get a pivot in every column, something is amiss—the scalar λ is not an eigenvalue of A. Looking back at Example 3, notice that $A - 5I$, $A - 3I$, and $A - I$ all have at least one column without a pivot; however, if λ is chosen to be any number other than 5, 3, or 1, the matrix $A - \lambda I$ has a pivot in every column.

In Examples 1 and 3, $\det(A - \lambda I)$ is a polynomial in λ. It can be shown that if A is an $n \times n$ matrix, then $\det(A - \lambda I)$ is a polynomial of degree n called the **characteristic polynomial** of A.

The eigenvalue 5 in Example 3 is said to have *multiplicity* 2 because $(\lambda - 5)$ occurs two times as a factor of the characteristic polynomial. In general, the (**algebraic**) **multiplicity** of an eigenvalue λ is its multiplicity as a root of the characteristic equation.

EXAMPLE 4 The characteristic polynomial of a 6×6 matrix is $\lambda^6 - 4\lambda^5 - 12\lambda^4$. Find the eigenvalues and their multiplicities.

SOLUTION Factor the polynomial

$$\lambda^6 - 4\lambda^5 - 12\lambda^4 = \lambda^4(\lambda^2 - 4\lambda - 12) = \lambda^4(\lambda - 6)(\lambda + 2)$$

The eigenvalues are 0 (multiplicity 4), 6 (multiplicity 1), and -2 (multiplicity 1). ■

We could also list the eigenvalues in Example 4 as $0, 0, 0, 0, 6$, and -2, so that the eigenvalues are repeated according to their multiplicities.

Because the characteristic equation for an $n \times n$ matrix involves an nth-degree polynomial, the equation has exactly n roots, counting multiplicities, provided complex roots are allowed. Such complex roots, called *complex eigenvalues*, will be discussed in Section 5.5. Until then, we consider only real eigenvalues, and scalars will continue to be real numbers.

The characteristic equation is important for theoretical purposes. In practical work, however, eigenvalues of any matrix larger than 2×2 should be found by a computer, unless the matrix is triangular or has other special properties. Although a 3×3 characteristic polynomial is easy to compute by hand, factoring it can be difficult (unless the matrix is carefully chosen). See the Numerical Notes at the end of this section.

STUDY GUIDE has advice on how to factor a polynomial.

Similarity

The next theorem illustrates one use of the characteristic polynomial, and it provides the foundation for several iterative methods that *approximate* eigenvalues. If A and B are $n \times n$ matrices, then A **is similar to** B if there is an invertible matrix P such that $P^{-1}AP = B$, or, equivalently, $A = PBP^{-1}$. Writing Q for P^{-1}, we have $Q^{-1}BQ = A$. So B is also similar to A, and we say simply that A and B **are similar**. Changing A into $P^{-1}AP$ is called a **similarity transformation**.

THEOREM 4

If $n \times n$ matrices A and B are similar, then they have the same characteristic polynomial and hence the same eigenvalues (with the same multiplicities).

PROOF If $B = P^{-1}AP$, then

$$B - \lambda I = P^{-1}AP - \lambda P^{-1}P = P^{-1}(AP - \lambda P) = P^{-1}(A - \lambda I)P$$

Using the multiplicative property (b) in Theorem 3, we compute

$$\begin{aligned} \det(B - \lambda I) &= \det[P^{-1}(A - \lambda I)P] \\ &= \det(P^{-1}) \cdot \det(A - \lambda I) \cdot \det(P) \end{aligned} \tag{1}$$

Since $\det(P^{-1}) \cdot \det(P) = \det(P^{-1}P) = \det I = 1$, we see from equation (1) that $\det(B - \lambda I) = \det(A - \lambda I)$. ■

***Warnings*:**

1. The matrices
$$\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$
are not similar even though they have the same eigenvalues.
2. Similarity is not the same as row equivalence. (If A is row equivalent to B, then $B = EA$ for some invertible matrix E.) Row operations on a matrix usually change its eigenvalues.

Application to Dynamical Systems

Eigenvalues and eigenvectors hold the key to the discrete evolution of a dynamical system, as mentioned in the chapter introduction.

EXAMPLE 5 Let $A = \begin{bmatrix} .95 & .03 \\ .05 & .97 \end{bmatrix}$. Analyze the long-term behavior (as k increases) of the dynamical system defined by $\mathbf{x}_{k+1} = A\mathbf{x}_k$ $(k = 0, 1, 2, \ldots)$, with $\mathbf{x}_0 = \begin{bmatrix} .6 \\ .4 \end{bmatrix}$.

SOLUTION The first step is to find the eigenvalues of A and a basis for each eigenspace. The characteristic equation for A is

$$\begin{aligned} 0 &= \det \begin{bmatrix} .95 - \lambda & .03 \\ .05 & .97 - \lambda \end{bmatrix} = (.95 - \lambda)(.97 - \lambda) - (.03)(.05) \\ &= \lambda^2 - 1.92\lambda + .92 \end{aligned}$$

By the quadratic formula

$$\begin{aligned} \lambda &= \frac{1.92 \pm \sqrt{(-1.92)^2 - 4(.92)}}{2} = \frac{1.92 \pm \sqrt{.0064}}{2} \\ &= \frac{1.92 \pm .08}{2} = 1 \quad \text{or} \quad .92 \end{aligned}$$

It is readily checked that eigenvectors corresponding to $\lambda = 1$ and $\lambda = .92$ are multiples of

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 5 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

respectively.

The next step is to write the given $\mathbf{x}_0$ in terms of $\mathbf{v}_1$ and $\mathbf{v}_2$. This can be done because $\{\mathbf{v}_1, \mathbf{v}_2\}$ is obviously a basis for $\mathbb{R}^2$. (Why?) So there exist weights c_1 and c_2 such that

$$\mathbf{x}_0 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = [\,\mathbf{v}_1 \quad \mathbf{v}_2\,]\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \tag{2}$$

In fact,

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = [\,\mathbf{v}_1 \quad \mathbf{v}_2\,]^{-1}\mathbf{x}_0 = \begin{bmatrix} 3 & 1 \\ 5 & -1 \end{bmatrix}^{-1}\begin{bmatrix} .60 \\ .40 \end{bmatrix}$$
$$= \frac{1}{-8}\begin{bmatrix} -1 & -1 \\ -5 & 3 \end{bmatrix}\begin{bmatrix} .60 \\ .40 \end{bmatrix} = \begin{bmatrix} .125 \\ .225 \end{bmatrix} \tag{3}$$

Because $\mathbf{v}_1$ and $\mathbf{v}_2$ in (3) are eigenvectors of A, with $A\mathbf{v}_1 = \mathbf{v}_1$ and $A\mathbf{v}_2 = .92\mathbf{v}_2$, we easily compute each $\mathbf{x}_k$:

$$\begin{aligned} \mathbf{x}_1 = A\mathbf{x}_0 &= c_1A\mathbf{v}_1 + c_2A\mathbf{v}_2 && \text{Using linearity of } \mathbf{x} \mapsto A\mathbf{x} \\ &= c_1\mathbf{v}_1 + c_2(.92)\mathbf{v}_2 && \mathbf{v}_1 \text{ and } \mathbf{v}_2 \text{ are eigenvectors.} \\ \mathbf{x}_2 = A\mathbf{x}_1 &= c_1A\mathbf{v}_1 + c_2(.92)A\mathbf{v}_2 \\ &= c_1\mathbf{v}_1 + c_2(.92)^2\mathbf{v}_2 \end{aligned}$$

and so on. In general,

$$\mathbf{x}_k = c_1\mathbf{v}_1 + c_2(.92)^k\mathbf{v}_2 \quad (k = 0, 1, 2, \ldots)$$

Using c_1 and c_2 from (4),

$$\mathbf{x}_k = .125\begin{bmatrix} 3 \\ 5 \end{bmatrix} + .225(.92)^k\begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad (k = 0, 1, 2, \ldots) \tag{4}$$

This explicit formula for $\mathbf{x}_k$ gives the solution of the difference equation $\mathbf{x}_{k+1} = A\mathbf{x}_k$. As $k \to \infty$, $(.92)^k$ tends to zero and $\mathbf{x}_k$ tends to $\begin{bmatrix} .375 \\ .625 \end{bmatrix} = .125\mathbf{v}_1$. ■

The calculations in Example 5 have an interesting application to a Markov chain discussed in Section 5.9. Those who read that section may recognize that matrix A in Example 5 above is the same as the migration matrix M in Section 5.9, $\mathbf{x}_0$ is the initial population distribution between city and suburbs, and $\mathbf{x}_k$ represents the population distribution after k years.

Numerical Notes

1. Computer software such as Mathematica and Maple can use symbolic calculations to find the characteristic polynomial of a moderate-sized matrix. But there is no formula or finite algorithm to solve the characteristic equation of a general $n \times n$ matrix for $n \geq 5$.

Numerical Notes (Continued)

2. The best numerical methods for finding eigenvalues avoid the characteristic polynomial entirely. In fact, MATLAB finds the characteristic polynomial of a matrix A by first computing the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A and then expanding the product $(\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n)$.

3. Several common algorithms for estimating the eigenvalues of a matrix A are based on Theorem 4. The powerful *QR algorithm* is discussed in the exercises. Another technique, called *Jacobi's method*, works when $A = A^T$ and computes a sequence of matrices of the form

$$A_1 = A \quad \text{and} \quad A_{k+1} = P_k^{-1} A_k P_k \quad (k = 1, 2, \ldots)$$

Each matrix in the sequence is similar to A and so has the same eigenvalues as A. The nondiagonal entries of A_{k+1} tend to zero as k increases, and the diagonal entries tend to approach the eigenvalues of A.

4. Other methods of estimating eigenvalues are discussed in Section 5.8.

Practice Problem

Find the characteristic equation and eigenvalues of $A = \begin{bmatrix} 1 & -4 \\ 4 & 2 \end{bmatrix}$.

5.2 Exercises

Find the characteristic polynomial and the eigenvalues of the matrices in Exercises 1–8.

1. $\begin{bmatrix} 2 & 7 \\ 7 & 2 \end{bmatrix}$ **2.** $\begin{bmatrix} 8 & 4 \\ 4 & 8 \end{bmatrix}$

3. $\begin{bmatrix} 3 & -2 \\ 1 & -1 \end{bmatrix}$ **4.** $\begin{bmatrix} 5 & -5 \\ -2 & 3 \end{bmatrix}$

5. $\begin{bmatrix} 2 & 1 \\ -1 & 4 \end{bmatrix}$ **6.** $\begin{bmatrix} 1 & -4 \\ 4 & 6 \end{bmatrix}$

7. $\begin{bmatrix} 5 & 3 \\ -4 & 4 \end{bmatrix}$ **8.** $\begin{bmatrix} 7 & -2 \\ 2 & 3 \end{bmatrix}$

Exercises 9–14 require techniques from Section 3.1. Find the characteristic polynomial of each matrix using expansion across a row or down a column. [*Note:* Finding the characteristic polynomial of a 3×3 matrix is not easy to do with just row operations, because the variable λ is involved.]

9. $\begin{bmatrix} 1 & 0 & -1 \\ 2 & 3 & -1 \\ 0 & 6 & 0 \end{bmatrix}$ **10.** $\begin{bmatrix} 0 & 3 & 1 \\ 3 & 0 & 2 \\ 1 & 2 & 0 \end{bmatrix}$

11. $\begin{bmatrix} 6 & 0 & 0 \\ 5 & 4 & 3 \\ 1 & 0 & 2 \end{bmatrix}$ **12.** $\begin{bmatrix} 1 & 0 & 1 \\ -3 & 6 & 1 \\ 0 & 0 & 4 \end{bmatrix}$

13. $\begin{bmatrix} 6 & -2 & 0 \\ -2 & 9 & 0 \\ 5 & 8 & 3 \end{bmatrix}$ **14.** $\begin{bmatrix} 3 & -2 & 3 \\ 0 & -1 & 0 \\ 6 & 7 & -4 \end{bmatrix}$

For the matrices in Exercises 15–17, list the eigenvalues, repeated according to their multiplicities.

15. $\begin{bmatrix} 7 & -5 & 3 & 0 \\ 0 & 3 & 7 & -5 \\ 0 & 0 & 5 & -3 \\ 0 & 0 & 0 & 7 \end{bmatrix}$ **16.** $\begin{bmatrix} 5 & 0 & 0 & 0 \\ 8 & -4 & 0 & 0 \\ 0 & 7 & 1 & 0 \\ 1 & -5 & 2 & 1 \end{bmatrix}$

17. $\begin{bmatrix} 3 & 0 & 0 & 0 & 0 \\ -5 & 1 & 0 & 0 & 0 \\ 3 & 8 & 0 & 0 & 0 \\ 0 & -7 & 2 & 1 & 0 \\ -4 & 1 & 9 & -2 & 3 \end{bmatrix}$

18. It can be shown that the algebraic multiplicity of an eigenvalue λ is always greater than or equal to the dimension of the eigenspace corresponding to λ. Find h in the matrix A below such that the eigenspace for $\lambda = 6$ is two-dimensional:

$$A = \begin{bmatrix} 6 & 3 & 9 & -5 \\ 0 & 9 & h & 2 \\ 0 & 0 & 6 & 8 \\ 0 & 0 & 0 & 7 \end{bmatrix}$$

19. Let A be an $n \times n$ matrix, and suppose A has n real eigenvalues, $\lambda_1, \ldots, \lambda_n$, repeated according to multiplicities, so that $\det(A - \lambda I) = (\lambda_1 - \lambda)(\lambda_2 - \lambda)\cdots(\lambda_n - \lambda)$

Explain why $\det A$ is the product of the n eigenvalues of A. (This result is true for any square matrix when complex eigenvalues are considered.)

20. Use a property of determinants to show that A and A^T have the same characteristic polynomial.

In Exercises 21–30, A and B are $n \times n$ matrices. Mark each statement True or False **(T/F)**. Justify each answer.

21. **(T/F)** If 0 is an eigenvalue of A, then A is invertible.

22. **(T/F)** The zero vector is in the eigenspace of A associated with an eigenvalue λ.

23. **(T/F)** The matrix A and its transpose, A^T, have different sets of eigenvalues.

24. **(T/F)** The matrices A and $B^{-1}AB$ have the same sets of eigenvalues for every invertible matrix B.

25. **(T/F)** If 2 is an eigenvalue of A, then $A - 2I$ is not invertible.

26. **(T/F)** If two matrices have the same set of eigenvalues, then they are similar.

27. **(T/F)** If $\lambda + 5$ is a factor of the characteristic polynomial of A, then 5 is an eigenvalue of A.

28. **(T/F)** The multiplicity of a root r of the characteristic equation of A is called the algebraic multiplicity of r as an eigenvalue of A.

29. **(T/F)** The eigenvalue of the $n \times n$ identity matrix is 1 with algebraic multiplicity n.

30. **(T/F)** The matrix A can have more than n eigenvalues.

A widely used method for estimating eigenvalues of a general matrix A is the *QR algorithm*. Under suitable conditions, this algorithm produces a sequence of matrices, all similar to A, that become almost upper triangular, with diagonal entries that approach the eigenvalues of A. The main idea is to factor A (or another matrix similar to A) in the form $A = Q_1R_1$, where $Q_1^T = Q_1^{-1}$ and R_1 is upper triangular. The factors are interchanged to form $A_1 = R_1Q_1$, which is again factored as $A_1 = Q_2R_2$; then to form $A_2 = R_2Q_2$, and so on. The similarity of $A, A_1, \ldots$ follows from the more general result in Exercise 31.

31. Show that if $A = QR$ with Q invertible, then A is similar to $A_1 = RQ$.

32. Show that if A and B are similar, then $\det A = \det B$.

T **33.** Construct a random integer-valued 4×4 matrix A, and verify that A and A^T have the same characteristic polynomial (the same eigenvalues with the same multiplicities). Do A and A^T have the same eigenvectors? Make the same analysis of a 5×5 matrix. Report the matrices and your conclusions.

T **34.** Construct a random integer-valued 4×4 matrix A.

a. Reduce A to echelon form U with no row scaling, and compute $\det A$. (If A happens to be singular, start over with a new random matrix.)

b. Compute the eigenvalues of A and the product of these eigenvalues (as accurately as possible).

c. List the matrix A, and, to four decimal places, list the pivots in U and the eigenvalues of A. Compute $\det A$ with your matrix program, and compare it with the products you found in (a) and (b).

T **35.** Let $A = \begin{bmatrix} -6 & 28 & 21 \\ 4 & -15 & -12 \\ -8 & a & 25 \end{bmatrix}$. For each value of a in the set $\{32, 31.9, 31.8, 32.1, 32.2\}$, compute the characteristic polynomial of A and the eigenvalues. In each case, create a graph of the characteristic polynomial $p(t) = \det(A - tI)$ for $0 \le t \le 3$. If possible, construct all graphs on one coordinate system. Describe how the graphs reveal the changes in the eigenvalues as a changes.

Solution to Practice Problem

The characteristic equation is

$$0 = \det(A - \lambda I) = \det \begin{bmatrix} 1-\lambda & -4 \\ 4 & 2-\lambda \end{bmatrix}$$
$$= (1-\lambda)(2-\lambda) - (-4)(4) = \lambda^2 - 3\lambda + 18$$

From the quadratic formula,

$$\lambda = \frac{3 \pm \sqrt{(-3)^2 - 4(18)}}{2} = \frac{3 \pm \sqrt{-63}}{2}$$

It is clear that the characteristic equation has no real solutions, so A has no real eigenvalues. The matrix A is acting on the real vector space $\mathbb{R}^2$, and there is no nonzero vector $\mathbf{v}$ in $\mathbb{R}^2$ such that $A\mathbf{v} = \lambda\mathbf{v}$ for some scalar λ.

5.3 Diagonalization

In many cases, the eigenvalue–eigenvector information contained within a matrix A can be displayed in a useful factorization of the form $A = PDP^{-1}$ where D is a diagonal matrix. In this section, the factorization enables us to compute A^k quickly for large values of k, a fundamental idea in several applications of linear algebra. Later, in Sections 5.6 and 5.7, the factorization will be used to analyze (and *decouple*) dynamical systems.

The following example illustrates that powers of a diagonal matrix are easy to compute.

EXAMPLE 1 If $D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$, then $D^2 = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 5^2 & 0 \\ 0 & 3^2 \end{bmatrix}$ and

$$D^3 = DD^2 = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} 5^2 & 0 \\ 0 & 3^2 \end{bmatrix} = \begin{bmatrix} 5^3 & 0 \\ 0 & 3^3 \end{bmatrix}$$

In general,

$$D^k = \begin{bmatrix} 5^k & 0 \\ 0 & 3^k \end{bmatrix} \quad \text{for } k \geq 1$$

■

If $A = PDP^{-1}$ for some invertible P and diagonal D, then A^k is also easy to compute, as the next example shows.

EXAMPLE 2 Let $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$. Find a formula for A^k, given that $A = PDP^{-1}$, where

$$P = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$$

SOLUTION The standard formula for the inverse of a 2×2 matrix yields

$$P^{-1} = \begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

Then, by associativity of matrix multiplication,

$$A^2 = (PDP^{-1})(PDP^{-1}) = PD\underbrace{(P^{-1}P)}_{I}DP^{-1} = PDDP^{-1}$$

$$= PD^2P^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}\begin{bmatrix} 5^2 & 0 \\ 0 & 3^2 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

Again,

$$A^3 = (PDP^{-1})A^2 = (\underbrace{PDP^{-1}}_{I})PD^2P^{-1} = PDD^2P^{-1} = PD^3P^{-1}$$

In general, for $k \geq 1$,

$$A^k = PD^kP^{-1} = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix}\begin{bmatrix} 5^k & 0 \\ 0 & 3^k \end{bmatrix}\begin{bmatrix} 2 & 1 \\ -1 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 \cdot 5^k - 3^k & 5^k - 3^k \\ 2 \cdot 3^k - 2 \cdot 5^k & 2 \cdot 3^k - 5^k \end{bmatrix}$$

■

A square matrix A is said to be **diagonalizable** if A is similar to a diagonal matrix, that is, if $A = PDP^{-1}$ for some invertible matrix P and some diagonal matrix D. The next theorem gives a characterization of diagonalizable matrices and tells how to construct a suitable factorization.

THEOREM 5

The Diagonalization Theorem

An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.

In fact, $A = PDP^{-1}$, with D a diagonal matrix, if and only if the columns of P are n linearly independent eigenvectors of A. In this case, the diagonal entries of D are eigenvalues of A that correspond, respectively, to the eigenvectors in P.

In other words, A is diagonalizable if and only if there are enough eigenvectors to form a basis of $\mathbb{R}^n$. We call such a basis an **eigenvector basis** of $\mathbb{R}^n$.

PROOF First, observe that if P is any $n \times n$ matrix with columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$, and if D is any diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$, then

$$AP = A[\,\mathbf{v}_1 \quad \mathbf{v}_2 \quad \cdots \quad \mathbf{v}_n\,] = [\,A\mathbf{v}_1 \quad A\mathbf{v}_2 \quad \cdots \quad A\mathbf{v}_n\,] \tag{1}$$

while

$$PD = P\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = [\,\lambda_1\mathbf{v}_1 \quad \lambda_2\mathbf{v}_2 \quad \cdots \quad \lambda_n\mathbf{v}_n\,] \tag{2}$$

Now suppose A is diagonalizable and $A = PDP^{-1}$. Then right-multiplying this relation by P, we have $AP = PD$. In this case, equations (1) and (2) imply that

$$[\,A\mathbf{v}_1 \quad A\mathbf{v}_2 \quad \cdots \quad A\mathbf{v}_n\,] = [\,\lambda_1\mathbf{v}_1 \quad \lambda_2\mathbf{v}_2 \quad \cdots \quad \lambda_n\mathbf{v}_n\,] \tag{3}$$

Equating columns, we find that

$$A\mathbf{v}_1 = \lambda_1\mathbf{v}_1, \quad A\mathbf{v}_2 = \lambda_2\mathbf{v}_2, \quad \ldots, \quad A\mathbf{v}_n = \lambda_n\mathbf{v}_n \tag{4}$$

Since P is invertible, its columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$ must be linearly independent. Also, since these columns are nonzero, the equations in (4) show that $\lambda_1, \ldots, \lambda_n$ are eigenvalues and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are corresponding eigenvectors. This argument proves the "only if" parts of the first and second statements, along with the third statement, of the theorem.

Finally, given any n eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$, use them to construct the columns of P and use corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$ to construct D. By equations (1)–(3), $AP = PD$. This is true without any condition on the eigenvectors. If, in fact, the eigenvectors are linearly independent, then P is invertible (by the Invertible Matrix Theorem), and $AP = PD$ implies that $A = PDP^{-1}$. ■

Diagonalizing Matrices

EXAMPLE 3 Diagonalize the following matrix, if possible.

$$A = \begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix}$$

That is, find an invertible matrix P and a diagonal matrix D such that $A = PDP^{-1}$.

SOLUTION There are four steps to implement the description in Theorem 5.

Step 1. ***Find the eigenvalues of A.*** As mentioned in Section 5.2, the mechanics of this step are appropriate for a computer when the matrix is larger than 2×2. To avoid unnecessary distractions, the text will usually supply information needed for this step. In the present case, the characteristic equation turns out to involve a cubic polynomial that can be factored:

$$\begin{aligned} 0 = \det(A - \lambda I) &= -\lambda^3 - 3\lambda^2 + 4 \\ &= -(\lambda - 1)(\lambda + 2)^2 \end{aligned}$$

The eigenvalues are $\lambda = 1$ and $\lambda = -2$.

Step 2. ***Find three linearly independent eigenvectors of A.*** *Three* vectors are needed because A is a 3×3 matrix. This is the critical step. If it fails, then Theorem 5 says that A cannot be diagonalized. The method in Section 5.1 produces a basis for each eigenspace:

$$\text{Basis for } \lambda = 1: \quad \mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

$$\text{Basis for } \lambda = -2: \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

You can check that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a linearly independent set.

Step 3. ***Construct P from the vectors in step 2.*** The vectors may be listed in any order. Using the order chosen in step 2, form

$$P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

Step 4. ***Construct D from the corresponding eigenvalues.*** In this step, it is essential that the order of the eigenvalues matches the order chosen for the columns of P. Use the eigenvalue $\lambda = -2$ twice, once for each of the eigenvectors corresponding to $\lambda = -2$:

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

It is a good idea to check that P and D really work. To avoid computing P^{-1}, simply verify that $AP = PD$. This is equivalent to $A = PDP^{-1}$ when P is invertible. (However, be sure that P is invertible!) Compute

$$AP = \begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 \\ -1 & -2 & 0 \\ 1 & 0 & -2 \end{bmatrix}$$

$$PD = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 \\ -1 & -2 & 0 \\ 1 & 0 & -2 \end{bmatrix}$$

■

EXAMPLE 4 Diagonalize the following matrix, if possible.

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

SOLUTION The characteristic equation of A turns out to be exactly the same as that in Example 3:

$$0 = \det(A - \lambda I) = -\lambda^3 - 3\lambda^2 + 4 = -(\lambda - 1)(\lambda + 2)^2$$

The eigenvalues are $\lambda = 1$ and $\lambda = -2$. However, it is easy to verify that each eigenspace is only one-dimensional:

$$\text{Basis for } \lambda = 1: \qquad \mathbf{v}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

$$\text{Basis for } \lambda = -2: \qquad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

There are no other eigenvalues, and every eigenvector of A is a multiple of either $\mathbf{v}_1$ or $\mathbf{v}_2$. Hence it is impossible to construct a basis of $\mathbb{R}^3$ using eigenvectors of A. By Theorem 5, A is *not* diagonalizable. ■

The following theorem provides a *sufficient* condition for a matrix to be diagonalizable.

THEOREM 6

An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.

PROOF Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be eigenvectors corresponding to the n distinct eigenvalues of a matrix A. Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly independent, by Theorem 2 in Section 5.1. Hence A is diagonalizable, by Theorem 5. ■

It is not *necessary* for an $n \times n$ matrix to have n distinct eigenvalues in order to be diagonalizable. The 3×3 matrix in Example 3 is diagonalizable even though it has only two distinct eigenvalues.

EXAMPLE 5 Determine if the following matrix is diagonalizable.

$$A = \begin{bmatrix} 5 & -8 & 1 \\ 0 & 0 & 7 \\ 0 & 0 & -2 \end{bmatrix}$$

SOLUTION This is easy! Since the matrix is triangular, its eigenvalues are obviously 5, 0, and -2. Since A is a 3×3 matrix with three distinct eigenvalues, A is diagonalizable. ■

Matrices Whose Eigenvalues Are Not Distinct

If an $n \times n$ matrix A has n distinct eigenvalues, with corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$, and if $P = [\,\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_n\,]$, then P is automatically invertible because its columns

are linearly independent, by Theorem 2. When A is diagonalizable but has fewer than n distinct eigenvalues, it is still possible to build P in a way that makes P automatically invertible, as the next theorem shows.[1]

THEOREM 7

Let A be an $n \times n$ matrix whose distinct eigenvalues are $\lambda_1, \ldots, \lambda_p$.

a. For $1 \le k \le p$, the dimension of the eigenspace for λ_k is less than or equal to the multiplicity of the eigenvalue λ_k.

b. The matrix A is diagonalizable if and only if the sum of the dimensions of the eigenspaces equals n, and this happens if and only if (*i*) the characteristic polynomial factors completely into linear factors and (*ii*) the dimension of the eigenspace for each λ_k equals the multiplicity of λ_k.

c. If A is diagonalizable and $\mathcal{B}_k$ is a basis for the eigenspace corresponding to λ_k for each k, then the total collection of vectors in the sets $\mathcal{B}_1, \ldots, \mathcal{B}_p$ forms an eigenvector basis for $\mathbb{R}^n$.

EXAMPLE 6 Diagonalize the following matrix, if possible.

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

SOLUTION Since A is a triangular matrix, the eigenvalues are 5 and -3, each with multiplicity 2. Using the method in Section 5.1, we find a basis for each eigenspace.

$$\text{Basis for } \lambda = 5: \quad \mathbf{v}_1 = \begin{bmatrix} -8 \\ 4 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} -16 \\ 4 \\ 0 \\ 1 \end{bmatrix}$$

$$\text{Basis for } \lambda = -3: \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{v}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

The set $\{\mathbf{v}_1, \ldots, \mathbf{v}_4\}$ is linearly independent, by Theorem 7. So the matrix $P = [\,\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_4\,]$ is invertible, and $A = PDP^{-1}$, where

$$P = \begin{bmatrix} -8 & -16 & 0 & 0 \\ 4 & 4 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & -3 \end{bmatrix}$$

■

[1] The proof of Theorem 7 is somewhat lengthy but not difficult. For instance, see S. Friedberg, A. Insel, and L. Spence, *Linear Algebra*, 4th ed. (Englewood Cliffs, NJ: Prentice-Hall, 2002), Section 5.2.

Practice Problems

1. Compute A^8, where $A = \begin{bmatrix} 4 & -3 \\ 2 & -1 \end{bmatrix}$.
2. Let $A = \begin{bmatrix} -3 & 12 \\ -2 & 7 \end{bmatrix}$, $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$, and $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Suppose you are told that $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors of A. Use this information to diagonalize A.
3. Let A be a 4×4 matrix with eigenvalues 5, 3, and -2, and suppose you know that the eigenspace for $\lambda = 3$ is two-dimensional. Do you have enough information to determine if A is diagonalizable?

5.3 Exercises

In Exercises 1 and 2, let $A = PDP^{-1}$ and compute A^4.

1. $P = \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}$, $D = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$

2. $P = \begin{bmatrix} 2 & -3 \\ -3 & 5 \end{bmatrix}$, $D = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$

In Exercises 3 and 4, use the factorization $A = PDP^{-1}$ to compute A^k, where k represents an arbitrary positive integer.

3. $\begin{bmatrix} a & 0 \\ 3(a-b) & b \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}$

4. $\begin{bmatrix} 15 & -36 \\ 6 & -15 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -3 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} -1 & 3 \\ 1 & -2 \end{bmatrix}$

In Exercises 5 and 6, the matrix A is factored in the form PDP^{-1}. Use the Diagonalization Theorem to find the eigenvalues of A and a basis for each eigenspace.

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

6. $\begin{bmatrix} 7 & -1 & 1 \\ 6 & 2 & 3 \\ 0 & 0 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 5 & 2 & 3 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 3 & -1 & -1 \\ -2 & 1 & -1 \end{bmatrix}$

Diagonalize the matrices in Exercises 7–20, if possible. The eigenvalues for Exercises 11–16 are as follows: (11) $\lambda = 1, 2, 3$; (12) $\lambda = 1, 4$; (13) $\lambda = 5, 1$; (14) $\lambda = 3, 4$; (15) $\lambda = 3, 1$; (16) $\lambda = 2, 1$. For Exercise 18, one eigenvalue is $\lambda = 5$ and one eigenvector is $(-2, 1, 2)$.

7. $\begin{bmatrix} 1 & 0 \\ 6 & -1 \end{bmatrix}$

8. $\begin{bmatrix} 5 & 1 \\ 0 & 5 \end{bmatrix}$

9. $\begin{bmatrix} 3 & -1 \\ 1 & 5 \end{bmatrix}$

10. $\begin{bmatrix} 3 & 6 \\ 4 & 1 \end{bmatrix}$

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

12. $\begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix}$

13. $\begin{bmatrix} 2 & 2 & -1 \\ 1 & 3 & -1 \\ -1 & -2 & 2 \end{bmatrix}$

14. $\begin{bmatrix} 4 & 0 & 2 \\ 2 & 3 & 4 \\ 0 & 0 & 3 \end{bmatrix}$

15. $\begin{bmatrix} -7 & 24 & -16 \\ -2 & 7 & -4 \\ 2 & -6 & 5 \end{bmatrix}$

16. $\begin{bmatrix} 0 & -4 & -6 \\ -1 & 0 & -3 \\ 1 & 2 & 5 \end{bmatrix}$

17. $\begin{bmatrix} 4 & 0 & 0 \\ 1 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix}$

18. $\begin{bmatrix} -7 & -16 & 4 \\ 6 & 13 & -2 \\ 12 & 16 & 1 \end{bmatrix}$

19. $\begin{bmatrix} 5 & -3 & 0 & 9 \\ 0 & 3 & 1 & -2 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$

20. $\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 2 \end{bmatrix}$

In Exercises 21–28, A, P, and D are $n \times n$ matrices. Mark each statement True or False **(T/F)**. Justify each answer. (Study Theorems 5 and 6 and the examples in this section carefully before you try these exercises.)

21. **(T/F)** A is diagonalizable if $A = PDP^{-1}$ for some matrix D and some invertible matrix P.

22. **(T/F)** If $\mathbb{R}^n$ has a basis of eigenvectors of A, then A is diagonalizable.

23. **(T/F)** A is diagonalizable if and only if A has n eigenvalues, counting multiplicities.

24. **(T/F)** If A is diagonalizable, then A is invertible.

25. **(T/F)** A is diagonalizable if A has n eigenvectors.

26. **(T/F)** If A is diagonalizable, then A has n distinct eigenvalues.

27. **(T/F)** If $AP = PD$, with D diagonal, then the nonzero columns of P must be eigenvectors of A.

28. **(T/F)** If A is invertible, then A is diagonalizable.

29. A is a 5×5 matrix with two eigenvalues. One eigenspace is three-dimensional, and the other eigenspace is two-dimensional. Is A diagonalizable? Why?

30. A is a 3×3 matrix with two eigenvalues. Each eigenspace is one-dimensional. Is A diagonalizable? Why?

31. A is a 4×4 matrix with three eigenvalues. One eigenspace is one-dimensional, and one of the other eigenspaces is two-dimensional. Is it possible that A is *not* diagonalizable? Justify your answer.

32. A is a 7×7 matrix with three eigenvalues. One eigenspace is two-dimensional, and one of the other eigenspaces is three-dimensional. Is it possible that A is *not* diagonalizable? Justify your answer.

33. Show that if A is both diagonalizable and invertible, then so is A^{-1}.

34. Show that if A has n linearly independent eigenvectors, then so does A^T. [*Hint:* Use the Diagonalization Theorem.]

35. A factorization $A = PDP^{-1}$ is not unique. Demonstrate this for the matrix A in Example 2. With $D_1 = \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}$, use the information in Example 2 to find a matrix P_1 such that $A = P_1 D_1 P_1^{-1}$.

36. With A and D as in Example 2, find an invertible P_2 unequal to the P in Example 2, such that $A = P_2 D P_2^{-1}$.

37. Construct a nonzero 2×2 matrix that is invertible but not diagonalizable.

38. Construct a nondiagonal 2×2 matrix that is diagonalizable but not invertible.

Diagonalize the matrices in Exercises 39–42. Use your matrix program's eigenvalue command to find the eigenvalues, and then compute bases for the eigenspaces as in Section 5.1.

T 39. $\begin{bmatrix} -6 & 4 & 0 & 9 \\ -3 & 0 & 1 & 6 \\ -1 & -2 & 1 & 0 \\ -4 & 4 & 0 & 7 \end{bmatrix}$

T 40. $\begin{bmatrix} 0 & 13 & 8 & 4 \\ 4 & 9 & 8 & 4 \\ 8 & 6 & 12 & 8 \\ 0 & 5 & 0 & -4 \end{bmatrix}$

T 41. $\begin{bmatrix} 11 & -6 & 4 & -10 & -4 \\ -3 & 5 & -2 & 4 & 1 \\ -8 & 12 & -3 & 12 & 4 \\ 1 & 6 & -2 & 3 & -1 \\ 8 & -18 & 8 & -14 & -1 \end{bmatrix}$

T 42. $\begin{bmatrix} 4 & 4 & 2 & 3 & -2 \\ 0 & 1 & -2 & -2 & 2 \\ 6 & 12 & 11 & 2 & -4 \\ 9 & 20 & 10 & 10 & -6 \\ 15 & 28 & 14 & 5 & -3 \end{bmatrix}$

Solutions to Practice Problems

1. $\det(A - \lambda I) = \lambda^2 - 3\lambda + 2 = (\lambda - 2)(\lambda - 1)$. The eigenvalues are 2 and 1, and the corresponding eigenvectors are $\mathbf{v}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Next, form

$$P = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}, \qquad D = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad \text{and} \quad P^{-1} = \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix}$$

Since $A = PDP^{-1}$,

$$\begin{aligned} A^8 = PD^8P^{-1} &= \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2^8 & 0 \\ 0 & 1^8 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 256 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -2 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 766 & -765 \\ 510 & -509 \end{bmatrix} \end{aligned}$$

2. Compute $A\mathbf{v}_1 = \begin{bmatrix} -3 & 12 \\ -2 & 7 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 1 \cdot \mathbf{v}_1$, and

$$A\mathbf{v}_2 = \begin{bmatrix} -3 & 12 \\ -2 & 7 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix} = 3 \cdot \mathbf{v}_2$$

So, $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors for the eigenvalues 1 and 3, respectively. Thus

$$A = PDP^{-1}, \quad \text{where} \quad P = \begin{bmatrix} 3 & 2 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$$

This page is intentionally left blank

6 Orthogonality and Least Squares



Introductory Example

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Can you tell a puppy from a kitten regardless of breed or coloring? Of course! Both may be furry bundles of joy, but one is clearly a feline and one is clearly a canine to the human eye. Simple. But, as blatantly obvious as it may seem to our trained eyes that are capable of interpreting meaning in the pixels of an image, this turns out to be a significant challenge for a machine. With the advancements in artificial intelligence (AI) and machine learning, computers are rapidly improving their capability to identify the creature on the left in the picture above as a puppy and the one on the right as a cat.

Many industries are now using AI technology to speed up the process of what once took hours of mindless work, such as post office scanners that can read bar codes and handwriting on envelopes to sort the mail with precision and speed. Nordstrom is using machine learning to design, display, organize, and recommend clothing ensembles to customers, exemplifying how even in creative aesthetic fields machine learning can be used to interpret color and shape patterns in pixels and organize visual possibilities into that which our eyes register as pleasing.

When calling a service desk number, one is often greeted by a machine that asks a series of questions and provides suggestions. Only persistence in interacting with this machine gets the caller through to a real person. More and more service calls are answered by machines, making it easier for customers' simple questions to be answered succinctly without waiting time and the obnoxious jingles of hold music. Google has designed an AI assistant that will handle making service calls for you too—booking a restaurant or hair appointment on your behalf.

AI and machine learning comprise developing systems that interpret external data correctly, learn from such data, and use that learning to achieve specific goals and tasks through flexibility and adaptation. Often, the driving engine behind these techniques is linear algebra. In Section 6.2, we see a simple way to design a matrix so that matrix multiplication can identify the correct pattern of blue and white squares. In Sections 6.5, 6.6, and 6.8, we explore techniques used in machine learning.

In order to find an approximate solution to an inconsistent system of equations that has no actual solution, a well-defined notion of nearness is needed. Section 6.1 introduces the concepts of distance and orthogonality in a vector space. Sections 6.2 and 6.3 show how orthogonality can be used to identify the point within a subspace W that is nearest to a point $\mathbf{y}$ lying outside of W. By taking W to be the column space of a matrix, Section 6.5 develops a method for producing approximate ("least-squares") solutions for inconsistent linear systems, an important technique in machine learning, which is discussed in Sections 6.6 and 6.8.

Section 6.4 provides another opportunity to see orthogonal projections at work, creating a matrix factorization widely used in numerical linear algebra. The remaining sections examine some of the many least-squares problems that arise in applications, including those in vector spaces more general than $\mathbb{R}^n$.

6.1 Inner Product, Length, and Orthogonality

Geometric concepts of length, distance, and perpendicularity, which are well known for $\mathbb{R}^2$ and $\mathbb{R}^3$, are defined here for $\mathbb{R}^n$. These concepts provide powerful geometric tools for solving many applied problems, including the least-squares problems mentioned above. All three notions are defined in terms of the inner product of two vectors.

The Inner Product

If $\mathbf{u}$ and $\mathbf{v}$ are vectors in $\mathbb{R}^n$, then we regard $\mathbf{u}$ and $\mathbf{v}$ as $n \times 1$ matrices. The transpose $\mathbf{u}^T$ is a $1 \times n$ matrix, and the matrix product $\mathbf{u}^T\mathbf{v}$ is a 1×1 matrix, which we write as a single real number (a scalar) without brackets. The number $\mathbf{u}^T\mathbf{v}$ is called the **inner product** of $\mathbf{u}$ and $\mathbf{v}$, and often it is written as $\mathbf{u}\cdot\mathbf{v}$. This inner product, mentioned in the exercises for Section 2.1, is also referred to as a **dot product**. If

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

then the inner product of $\mathbf{u}$ and $\mathbf{v}$ is

$$\begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1v_1 + u_2v_2 + \cdots + u_nv_n$$

EXAMPLE 1 Compute $\mathbf{u}\cdot\mathbf{v}$ and $\mathbf{v}\cdot\mathbf{u}$ for $\mathbf{u} = \begin{bmatrix} 2 \\ -5 \\ -1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 3 \\ 2 \\ -3 \end{bmatrix}$.

SOLUTION

$$\mathbf{u}\cdot\mathbf{v} = \mathbf{u}^T\mathbf{v} = \begin{bmatrix} 2 & -5 & -1 \end{bmatrix}\begin{bmatrix} 3 \\ 2 \\ -3 \end{bmatrix} = (2)(3) + (-5)(2) + (-1)(-3) = -1$$

$$\mathbf{v}\cdot\mathbf{u} = \mathbf{v}^T\mathbf{u} = \begin{bmatrix} 3 & 2 & -3 \end{bmatrix}\begin{bmatrix} 2 \\ -5 \\ -1 \end{bmatrix} = (3)(2) + (2)(-5) + (-3)(-1) = -1$$ ■

It is clear from the calculations in Example 1 why $\mathbf{u}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{u}$. This commutativity of the inner product holds in general. The following properties of the inner product are easily deduced from properties of the transpose operation in Section 2.1. (See Exercises 29 and 30 at the end of this section.)

THEOREM 1

Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be vectors in $\mathbb{R}^n$, and let c be a scalar. Then

a. $\mathbf{u}\cdot\mathbf{v} = \mathbf{v}\cdot\mathbf{u}$

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

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

d. $\mathbf{u}\cdot\mathbf{u} \geq 0$, and $\mathbf{u}\cdot\mathbf{u} = 0$ if and only if $\mathbf{u} = \mathbf{0}$

Properties (b) and (c) can be combined several times to produce the following useful rule:

$$(c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p)\cdot\mathbf{w} = c_1(\mathbf{u}_1\cdot\mathbf{w}) + \cdots + c_p(\mathbf{u}_p\cdot\mathbf{w})$$

The Length of a Vector

If $\mathbf{v}$ is in $\mathbb{R}^n$, with entries $v_1, \ldots, v_n$, then the square root of $\mathbf{v}\cdot\mathbf{v}$ is defined because $\mathbf{v}\cdot\mathbf{v}$ is nonnegative.

DEFINITION

The **length** (or **norm**) of $\mathbf{v}$ is the nonnegative scalar $\|\mathbf{v}\|$ defined by

$$\|\mathbf{v}\| = \sqrt{\mathbf{v}\cdot\mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}, \quad \text{and} \quad \|\mathbf{v}\|^2 = \mathbf{v}\cdot\mathbf{v}$$

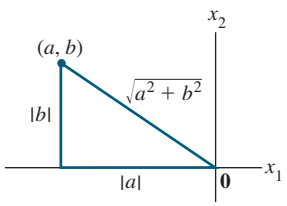


FIGURE 1
Interpretation of $\|\mathbf{v}\|$ as length.

Suppose $\mathbf{v}$ is in $\mathbb{R}^2$, say, $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$. If we identify $\mathbf{v}$ with a geometric point in the plane, as usual, then $\|\mathbf{v}\|$ coincides with the standard notion of the length of the line segment from the origin to $\mathbf{v}$. This follows from the Pythagorean Theorem applied to a triangle such as the one in Figure 1.

A similar calculation with the diagonal of a rectangular box shows that the definition of length of a vector $\mathbf{v}$ in $\mathbb{R}^3$ coincides with the usual notion of length.

For any scalar c, the length of $c\mathbf{v}$ is $|c|$ times the length of $\mathbf{v}$. That is,

$$\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$$

(To see this, compute $\|c\mathbf{v}\|^2 = (c\mathbf{v})\cdot(c\mathbf{v}) = c^2\mathbf{v}\cdot\mathbf{v} = c^2\|\mathbf{v}\|^2$ and take square roots.)

A vector whose length is 1 is called a **unit vector**. If we *divide* a nonzero vector $\mathbf{v}$ by its length—that is, multiply by $1/\|\mathbf{v}\|$—we obtain a unit vector $\mathbf{u}$ because the length of $\mathbf{u}$ is $(1/\|\mathbf{v}\|)\|\mathbf{v}\|$. The process of creating $\mathbf{u}$ from $\mathbf{v}$ is sometimes called **normalizing** $\mathbf{v}$, and we say that $\mathbf{u}$ is *in the same direction* as $\mathbf{v}$.

Several examples that follow use the space-saving notation for (column) vectors.

EXAMPLE 2 Let $\mathbf{v} = (1, -2, 2, 0)$. Find a unit vector $\mathbf{u}$ in the same direction as $\mathbf{v}$.

SOLUTION First, compute the length of $\mathbf{v}$:

$$\|\mathbf{v}\|^2 = \mathbf{v}\cdot\mathbf{v} = (1)^2 + (-2)^2 + (2)^2 + (0)^2 = 9$$
$$\|\mathbf{v}\| = \sqrt{9} = 3$$

Then, multiply $\mathbf{v}$ by $1/\|\mathbf{v}\|$ to obtain

$$\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v} = \frac{1}{3}\mathbf{v} = \frac{1}{3}\begin{bmatrix} 1 \\ -2 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \\ 0 \end{bmatrix}$$

To check that $\|\mathbf{u}\| = 1$, it suffices to show that $\|\mathbf{u}\|^2 = 1$.

$$\|\mathbf{u}\|^2 = \mathbf{u}\cdot\mathbf{u} = \left(\tfrac{1}{3}\right)^2 + \left(-\tfrac{2}{3}\right)^2 + \left(\tfrac{2}{3}\right)^2 + (0)^2$$
$$= \tfrac{1}{9} + \tfrac{4}{9} + \tfrac{4}{9} + 0 = 1$$

■

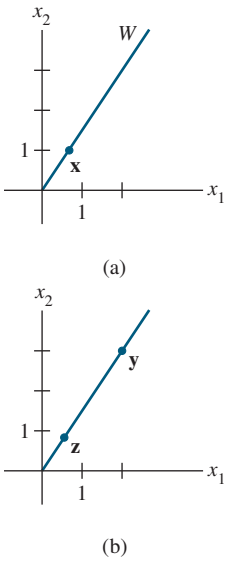


FIGURE 2
Normalizing a vector to produce a unit vector.

EXAMPLE 3 Let W be the subspace of $\mathbb{R}^2$ spanned by $\mathbf{x} = (\frac{2}{3}, 1)$. Find a unit vector $\mathbf{z}$ that is a basis for W.

SOLUTION W consists of all multiples of $\mathbf{x}$, as in Figure 2(a). Any nonzero vector in W is a basis for W. To simplify the calculation, "scale" $\mathbf{x}$ to eliminate fractions. That is, multiply $\mathbf{x}$ by 3 to get

$$\mathbf{y} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

Now compute $\|\mathbf{y}\|^2 = 2^2 + 3^2 = 13$, $\|\mathbf{y}\| = \sqrt{13}$, and normalize $\mathbf{y}$ to get

$$\mathbf{z} = \frac{1}{\sqrt{13}}\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 2/\sqrt{13} \\ 3/\sqrt{13} \end{bmatrix}$$

See Figure 2(b). Another unit vector is $(-2/\sqrt{13}, -3/\sqrt{13})$. ■

Distance in $\mathbb{R}^n$

We are ready now to describe how close one vector is to another. Recall that if a and b are real numbers, the distance on the number line between a and b is the number $|a - b|$. Two examples are shown in Figure 3. This definition of distance in $\mathbb{R}$ has a direct analogue in $\mathbb{R}^n$.

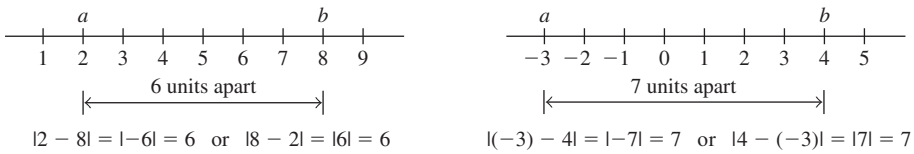


FIGURE 3 Distances in $\mathbb{R}$.

DEFINITION

For $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, the **distance between u and v**, written as $\text{dist}(\mathbf{u}, \mathbf{v})$, is the length of the vector $\mathbf{u} - \mathbf{v}$. That is,

$$\text{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

In $\mathbb{R}^2$ and $\mathbb{R}^3$, this definition of distance coincides with the usual formulas for the Euclidean distance between two points, as the next two examples show.

EXAMPLE 4 Compute the distance between the vectors $\mathbf{u} = (7, 1)$ and $\mathbf{v} = (3, 2)$.

SOLUTION Calculate

$$\mathbf{u} - \mathbf{v} = \begin{bmatrix} 7 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$$

$$\|\mathbf{u} - \mathbf{v}\| = \sqrt{4^2 + (-1)^2} = \sqrt{17}$$

The vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{u} - \mathbf{v}$ are shown in Figure 4. When the vector $\mathbf{u} - \mathbf{v}$ is added to $\mathbf{v}$, the result is $\mathbf{u}$. Notice that the parallelogram in Figure 4 shows that the distance from $\mathbf{u}$ to $\mathbf{v}$ is the same as the distance from $\mathbf{u} - \mathbf{v}$ to $\mathbf{0}$. ■

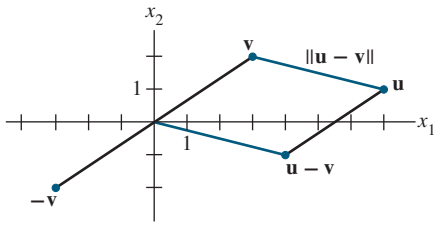


FIGURE 4 The distance between $\mathbf{u}$ and $\mathbf{v}$ is the length of $\mathbf{u} - \mathbf{v}$.

EXAMPLE 5 If $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$, then

$$\begin{aligned} \text{dist}(\mathbf{u}, \mathbf{v}) &= \|\mathbf{u} - \mathbf{v}\| = \sqrt{(\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v})} \\ &= \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + (u_3 - v_3)^2} \end{aligned}$$

■

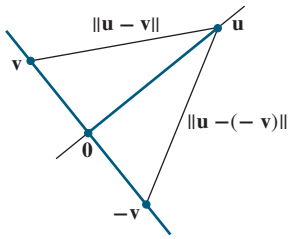


FIGURE 5

Orthogonal Vectors

The rest of this chapter depends on the fact that the concept of perpendicular lines in ordinary Euclidean geometry has an analogue in $\mathbb{R}^n$.

Consider $\mathbb{R}^2$ or $\mathbb{R}^3$ and two lines through the origin determined by vectors $\mathbf{u}$ and $\mathbf{v}$. The two lines shown in Figure 5 are geometrically perpendicular if and only if the distance from $\mathbf{u}$ to $\mathbf{v}$ is the same as the distance from $\mathbf{u}$ to $-\mathbf{v}$. This is the same as requiring the squares of the distances to be the same. Now

$$\begin{aligned} [\,\text{dist}(\mathbf{u}, -\mathbf{v})\,]^2 &= \|\mathbf{u} - (-\mathbf{v})\|^2 = \|\mathbf{u} + \mathbf{v}\|^2 \\ &= (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) \\ &= \mathbf{u} \cdot (\mathbf{u} + \mathbf{v}) + \mathbf{v} \cdot (\mathbf{u} + \mathbf{v}) \qquad \text{Theorem 1(b)} \\ &= \mathbf{u} \cdot \mathbf{u} + \mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{v} \qquad \text{Theorem 1(a), (b)} \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\mathbf{u} \cdot \mathbf{v} \qquad \text{Theorem 1(a)} \end{aligned} \tag{1}$$

The same calculations with $\mathbf{v}$ and $-\mathbf{v}$ interchanged show that

$$[\text{dist}\,(\mathbf{u}, \mathbf{v})]^2 = \|\mathbf{u}\|^2 + \|-\mathbf{v}\|^2 + 2\mathbf{u}\cdot(-\mathbf{v})$$
$$= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\mathbf{u}\cdot\mathbf{v}$$

The two squared distances are equal if and only if $2\mathbf{u}\cdot\mathbf{v} = -2\mathbf{u}\cdot\mathbf{v}$, which happens if and only if $\mathbf{u}\cdot\mathbf{v} = 0$.

This calculation shows that when vectors $\mathbf{u}$ and $\mathbf{v}$ are identified with geometric points, the corresponding lines through the points and the origin are perpendicular if and only if $\mathbf{u}\cdot\mathbf{v} = 0$. The following definition generalizes to $\mathbb{R}^n$ this notion of perpendicularity (or *orthogonality*, as it is commonly called in linear algebra).

DEFINITION

Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are **orthogonal** (to each other) if $\mathbf{u}\cdot\mathbf{v} = 0$.

Observe that the zero vector is orthogonal to every vector in $\mathbb{R}^n$ because $\mathbf{0}^T\mathbf{v} = 0$ for all $\mathbf{v}$.

The next theorem provides a useful fact about orthogonal vectors. The proof follows immediately from the calculation in (1) and the definition of orthogonality. The right triangle shown in Figure 6 provides a visualization of the lengths that appear in the theorem.

THEOREM 2

The Pythagorean Theorem

Two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if and only if $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

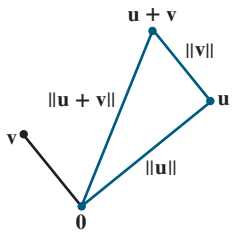


FIGURE 6

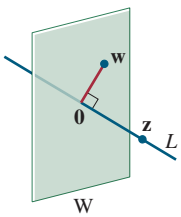


FIGURE 7

A plane and line through $\mathbf{0}$ as orthogonal complements.

Orthogonal Complements

To provide practice using inner products, we introduce a concept here that will be of use in Section 6.3 and elsewhere in the chapter. If a vector $\mathbf{z}$ is orthogonal to every vector in a subspace W of $\mathbb{R}^n$, then $\mathbf{z}$ is said to be **orthogonal to** W. The set of all vectors $\mathbf{z}$ that are orthogonal to W is called the **orthogonal complement** of W and is denoted by $W^\perp$ (and read as "W perpendicular" or simply "W perp").

EXAMPLE 6 Let W be a plane through the origin in $\mathbb{R}^3$, and let L be the line through the origin and perpendicular to W. If $\mathbf{z}$ and $\mathbf{w}$ are nonzero, $\mathbf{z}$ is on L, and $\mathbf{w}$ is in W, then the line segment from $\mathbf{0}$ to $\mathbf{z}$ is perpendicular to the line segment from $\mathbf{0}$ to $\mathbf{w}$; that is, $\mathbf{z}\cdot\mathbf{w} = 0$. See Figure 7. So each vector on L is orthogonal to every $\mathbf{w}$ in W. In fact, L consists of *all* vectors that are orthogonal to the $\mathbf{w}$'s in W, and W consists of all vectors orthogonal to the $\mathbf{z}$'s in L. That is,

$$L = W^\perp \quad \text{and} \quad W = L^\perp$$

The following two facts about $W^\perp$, with W a subspace of $\mathbb{R}^n$, are needed later in the chapter. Proofs are suggested in Exercises 37 and 38. Exercises 35–39 provide excellent practice using properties of the inner product.

1. A vector $\mathbf{x}$ is in $W^\perp$ if and only if $\mathbf{x}$ is orthogonal to every vector in a set that spans W.
2. $W^\perp$ is a subspace of $\mathbb{R}^n$.

The next theorem and Exercise 39 verify the claims made in Section 4.5 concerning the subspaces shown in Figure 8.

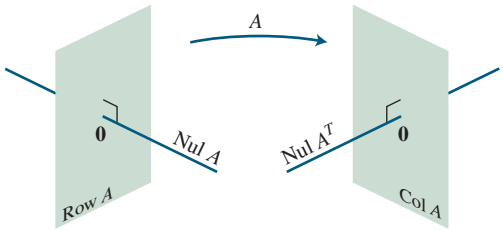


FIGURE 8 The fundamental subspaces determined by an $m \times n$ matrix A.

Remark: A common way to prove that two sets, say S and T, are equal is to show that S is a subset of T and T is a subset of S. The proof of the next theorem that Nul A = (Row A)$^\perp$ is established by showing that Nul A is a subset of (Row A)$^\perp$ and (Row A)$^\perp$ is a subset of Nul A. That is, an arbitrary element $\mathbf{x}$ in Nul A is shown to be in (Row A)$^\perp$, and then an arbitrary element $\mathbf{x}$ in (Row A)$^\perp$ is shown to be in Nul A.

THEOREM 3

Let A be an $m \times n$ matrix. The orthogonal complement of the row space of A is the null space of A, and the orthogonal complement of the column space of A is the null space of A^T:

$$(\text{Row } A)^\perp = \text{Nul } A \quad \text{and} \quad (\text{Col } A)^\perp = \text{Nul } A^T$$

PROOF The row–column rule for computing $A\mathbf{x}$ shows that if $\mathbf{x}$ is in Nul A, then $\mathbf{x}$ is orthogonal to each row of A (with the rows treated as vectors in $\mathbb{R}^n$). Since the rows of A span the row space, $\mathbf{x}$ is orthogonal to Row A. Conversely, if $\mathbf{x}$ is orthogonal to Row A, then $\mathbf{x}$ is certainly orthogonal to each row of A, and hence $A\mathbf{x} = \mathbf{0}$. This proves the first statement of the theorem. Since this statement is true for any matrix, it is true for A^T. That is, the orthogonal complement of the row space of A^T is the null space of A^T. This proves the second statement, because Row A^T = Col A.

Angles in $\mathbb{R}^2$ and $\mathbb{R}^3$ (Optional)

If $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors in either $\mathbb{R}^2$ or $\mathbb{R}^3$, then there is a nice connection between their inner product and the angle ϑ between the two line segments from the origin to the points identified with $\mathbf{u}$ and $\mathbf{v}$. The formula is

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \, \|\mathbf{v}\| \cos \vartheta \tag{2}$$

To verify this formula for vectors in $\mathbb{R}^2$, consider the triangle shown in Figure 9, with sides of lengths $\|\mathbf{u}\|$, $\|\mathbf{v}\|$, and $\|\mathbf{u} - \mathbf{v}\|$. By the law of cosines,

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\| \, \|\mathbf{v}\| \cos \vartheta$$

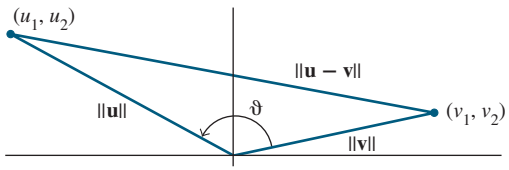


FIGURE 9 The angle between two vectors.

which can be rearranged to produce

$$\begin{aligned}\|\mathbf{u}\|\,\|\mathbf{v}\|\cos\vartheta &= \frac{1}{2}\left[\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - \|\mathbf{u}-\mathbf{v}\|^2\right]\\ &= \frac{1}{2}\left[u_1^2 + u_2^2 + v_1^2 + v_2^2 - (u_1-v_1)^2 - (u_2-v_2)^2\right]\\ &= u_1v_1 + u_2v_2\\ &= \mathbf{u}\cdot\mathbf{v}\end{aligned}$$

The verification for $\mathbb{R}^3$ is similar. When $n > 3$, formula (2) may be used to *define* the angle between two vectors in $\mathbb{R}^n$. In statistics, for instance, the value of $\cos\vartheta$ defined by (2) for suitable vectors $\mathbf{u}$ and $\mathbf{v}$ is what statisticians call a *correlation coefficient*.

Practice Problems

1. Let $\mathbf{a} = \begin{bmatrix}-2\\1\end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix}-3\\1\end{bmatrix}$. Compute $\dfrac{\mathbf{a}\cdot\mathbf{b}}{\mathbf{a}\cdot\mathbf{a}}$ and $\left(\dfrac{\mathbf{a}\cdot\mathbf{b}}{\mathbf{a}\cdot\mathbf{a}}\right)\mathbf{a}$.

2. Let $\mathbf{c} = \begin{bmatrix}4/3\\-1\\2/3\end{bmatrix}$ and $\mathbf{d} = \begin{bmatrix}5\\6\\-1\end{bmatrix}$.

a. Find a unit vector $\mathbf{u}$ in the direction of $\mathbf{c}$.

b. Show that $\mathbf{d}$ is orthogonal to $\mathbf{c}$.

c. Use the results of (a) and (b) to explain why $\mathbf{d}$ must be orthogonal to the unit vector $\mathbf{u}$.

3. Let W be a subspace of $\mathbb{R}^n$. Exercise 38 establishes that $W^\perp$ is also a subspace of $\mathbb{R}^n$. Prove that $\dim W + \dim W^\perp = n$.

6.1 Exercises

Compute the quantities in Exercises 1–8 using the vectors

$$\mathbf{u} = \begin{bmatrix}-1\\2\end{bmatrix},\quad \mathbf{v} = \begin{bmatrix}2\\3\end{bmatrix},\quad \mathbf{w} = \begin{bmatrix}3\\-1\\-5\end{bmatrix},\quad \mathbf{x} = \begin{bmatrix}6\\-2\\3\end{bmatrix}$$

1. $\mathbf{u}\cdot\mathbf{u}$, $\mathbf{v}\cdot\mathbf{u}$, and $\dfrac{\mathbf{v}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}$

2. $\mathbf{w}\cdot\mathbf{w}$, $\mathbf{x}\cdot\mathbf{w}$, and $\dfrac{\mathbf{x}\cdot\mathbf{w}}{\mathbf{w}\cdot\mathbf{w}}$

3. $\dfrac{1}{\mathbf{w}\cdot\mathbf{w}}\mathbf{w}$

4. $\dfrac{1}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$

5. $\left(\dfrac{\mathbf{u}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v}$

6. $\left(\dfrac{\mathbf{x}\cdot\mathbf{w}}{\mathbf{x}\cdot\mathbf{x}}\right)\mathbf{x}$

7. $\|\mathbf{w}\|$

8. $\|\mathbf{x}\|$

In Exercises 9–12, find a unit vector in the direction of the given vector.

9. $\begin{bmatrix}-30\\40\end{bmatrix}$

10. $\begin{bmatrix}3\\6\\-3\end{bmatrix}$

11. $\begin{bmatrix}2/9\\1/3\\1\end{bmatrix}$

12. $\begin{bmatrix}8/3\\1\end{bmatrix}$

13. Find the distance between $\mathbf{x} = \begin{bmatrix}10\\-3\end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix}-1\\-5\end{bmatrix}$.

14. Find the distance between $\mathbf{u} = \begin{bmatrix}0\\-1\\3\end{bmatrix}$ and $\mathbf{z} = \begin{bmatrix}-7\\-5\\7\end{bmatrix}$.

Determine which pairs of vectors in Exercises 15–18 are orthogonal.

15. $\mathbf{a} = \begin{bmatrix}8\\-5\end{bmatrix}$, $\mathbf{b} = \begin{bmatrix}-2\\-3\end{bmatrix}$

16. $\mathbf{x} = \begin{bmatrix}4\\-2\\5\end{bmatrix}$, $\mathbf{y} = \begin{bmatrix}11\\-1\\-9\end{bmatrix}$

17. $\mathbf{u} = \begin{bmatrix}3\\2\\-5\\0\end{bmatrix}$, $\mathbf{v} = \begin{bmatrix}-4\\1\\-2\\6\end{bmatrix}$

18. $\mathbf{w} = \begin{bmatrix}3\\-6\\7\\8\end{bmatrix}$, $\mathbf{z} = \begin{bmatrix}-9\\6\\17\\-7\end{bmatrix}$

In Exercises 19–28, all vectors are in $\mathbb{R}^n$. Mark each statement True or False **(T/F)**. Justify each answer.

19. **(T/F)** $\mathbf{v}\cdot\mathbf{v} = \|\mathbf{v}\|^2$.

20. **(T/F)** $\mathbf{u}\cdot\mathbf{v} - \mathbf{v}\cdot\mathbf{u} = 0$.

21. **(T/F)** If the distance from $\mathbf{u}$ to $\mathbf{v}$ equals the distance from $\mathbf{u}$ to $-\mathbf{v}$, then $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

22. **(T/F)** If $\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 = \|\mathbf{u}+\mathbf{v}\|^2$, then $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

23. **(T/F)** If vectors $\mathbf{v}_1, \ldots, \mathbf{v}_p$ span a subspace W and if $\mathbf{x}$ is orthogonal to each $\mathbf{v}_j$ for $j = 1, \ldots, p$, then $\mathbf{x}$ is in $W^\perp$.

24. **(T/F)** If $\mathbf{x}$ is orthogonal to every vector in a subspace W then $\mathbf{x}$ is in $W^\perp$.

25. **(T/F)** For any scalar c, $\|c\mathbf{v}\| = c\|\mathbf{v}\|$.

26. **(T/F)** For any scalar c, $\mathbf{u}\cdot(c\mathbf{v}) = c(\mathbf{u}\cdot\mathbf{v})$.

27. **(T/F)** For a square matrix A, vectors in Col A are orthogonal to vectors in Nul A.

28. **(T/F)** For an $m \times n$ matrix A, vectors in the null space of A are orthogonal to vectors in the row space of A.

29. Use the transpose definition of the inner product to verify parts (b) and (c) of Theorem 1. Mention the appropriate facts from Chapter 2.

30. Let $\mathbf{u} = (u_1, u_2, u_3)$. Explain why $\mathbf{u}\cdot\mathbf{u} \geq 0$. When is $\mathbf{u}\cdot\mathbf{u} = 0$?

31. Let $\mathbf{u} = \begin{bmatrix} 3 \\ -4 \\ -1 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} -8 \\ -7 \\ 4 \end{bmatrix}$. Compute and compare $\mathbf{u}\cdot\mathbf{v}$, $\|\mathbf{u}\|^2$, $\|\mathbf{v}\|^2$, and $\|\mathbf{u}+\mathbf{v}\|^2$. Do not use the Pythagorean Theorem.

32. Verify the *parallelogram law* for vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$:

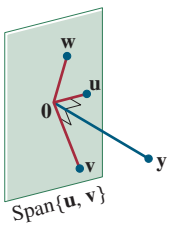
$$\|\mathbf{u}+\mathbf{v}\|^2 + \|\mathbf{u}-\mathbf{v}\|^2 = 2\|\mathbf{u}\|^2 + 2\|\mathbf{v}\|^2$$

33. Let $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$. Describe the set H of vectors $\begin{bmatrix} x \\ y \end{bmatrix}$ that are orthogonal to $\mathbf{v}$. [*Hint:* Consider $\mathbf{v} = \mathbf{0}$ and $\mathbf{v} \neq \mathbf{0}$.]

34. Let $\mathbf{u} = \begin{bmatrix} 5 \\ -6 \\ 7 \end{bmatrix}$, and let W be the set of all $\mathbf{x}$ in $\mathbb{R}^3$ such that $\mathbf{u}\cdot\mathbf{x} = 0$. What theorem in Chapter 4 can be used to show that W is a subspace of $\mathbb{R}^3$? Describe W in geometric language.

35. Suppose a vector $\mathbf{y}$ is orthogonal to vectors $\mathbf{u}$ and $\mathbf{v}$. Show that $\mathbf{y}$ is orthogonal to the vector $\mathbf{u}+\mathbf{v}$.

36. Suppose $\mathbf{y}$ is orthogonal to $\mathbf{u}$ and $\mathbf{v}$. Show that $\mathbf{y}$ is orthogonal to every $\mathbf{w}$ in Span $\{\mathbf{u}, \mathbf{v}\}$. [*Hint:* An arbitrary $\mathbf{w}$ in Span $\{\mathbf{u}, \mathbf{v}\}$ has the form $\mathbf{w} = c_1\mathbf{u} + c_2\mathbf{v}$. Show that $\mathbf{y}$ is orthogonal to such a vector $\mathbf{w}$.]



37. Let $W = \text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$. Show that if $\mathbf{x}$ is orthogonal to each $\mathbf{v}_j$, for $1 \leq j \leq p$, then $\mathbf{x}$ is orthogonal to every vector in W.

38. Let W be a subspace of $\mathbb{R}^n$, and let $W^\perp$ be the set of all vectors orthogonal to W. Show that $W^\perp$ is a subspace of $\mathbb{R}^n$ using the following steps.

a. Take $\mathbf{z}$ in $W^\perp$, and let $\mathbf{u}$ represent any element of W. Then $\mathbf{z}\cdot\mathbf{u} = 0$. Take any scalar c and show that $c\mathbf{z}$ is orthogonal to $\mathbf{u}$. (Since $\mathbf{u}$ was an arbitrary element of W, this will show that $c\mathbf{z}$ is in $W^\perp$.)

b. Take $\mathbf{z}_1$ and $\mathbf{z}_2$ in $W^\perp$, and let $\mathbf{u}$ be any element of W. Show that $\mathbf{z}_1 + \mathbf{z}_2$ is orthogonal to $\mathbf{u}$. What can you conclude about $\mathbf{z}_1 + \mathbf{z}_2$? Why?

c. Finish the proof that $W^\perp$ is a subspace of $\mathbb{R}^n$.

39. Show that if $\mathbf{x}$ is in both W and $W^\perp$, then $\mathbf{x} = \mathbf{0}$.

T 40. Construct a pair $\mathbf{u}$, $\mathbf{v}$ of random vectors in $\mathbb{R}^4$, and let

$$A = \begin{bmatrix} .5 & .5 & .5 & .5 \\ .5 & .5 & -.5 & -.5 \\ .5 & -.5 & .5 & -.5 \\ .5 & -.5 & -.5 & .5 \end{bmatrix}$$

a. Denote the columns of A by $\mathbf{a}_1, \ldots, \mathbf{a}_4$. Compute the length of each column, and compute $\mathbf{a}_1\cdot\mathbf{a}_2$, $\mathbf{a}_1\cdot\mathbf{a}_3$, $\mathbf{a}_1\cdot\mathbf{a}_4$, $\mathbf{a}_2\cdot\mathbf{a}_3$, $\mathbf{a}_2\cdot\mathbf{a}_4$, and $\mathbf{a}_3\cdot\mathbf{a}_4$.

b. Compute and compare the lengths of $\mathbf{u}$, $A\mathbf{u}$, $\mathbf{v}$, and $A\mathbf{v}$.

c. Use equation (2) in this section to compute the cosine of the angle between $\mathbf{u}$ and $\mathbf{v}$. Compare this with the cosine of the angle between $A\mathbf{u}$ and $A\mathbf{v}$.

d. Repeat parts (b) and (c) for two other pairs of random vectors. What do you conjecture about the effect of A on vectors?

T 41. Generate random vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{v}$ in $\mathbb{R}^4$ with integer entries (and $\mathbf{v} \neq \mathbf{0}$), and compute the quantities

$$\left(\frac{\mathbf{x}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v}, \left(\frac{\mathbf{y}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v}, \frac{(\mathbf{x}+\mathbf{y})\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v}, \frac{(10\mathbf{x})\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v}$$

Repeat the computations with new random vectors $\mathbf{x}$ and $\mathbf{y}$. What do you conjecture about the mapping $\mathbf{x} \mapsto T(\mathbf{x}) = \left(\frac{\mathbf{x}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\right)\mathbf{v}$ (for $\mathbf{v} \neq \mathbf{0}$)? Verify your conjecture algebraically.

T 42. Let $A = \begin{bmatrix} -6 & 3 & -27 & -33 & -13 \\ 6 & -5 & 25 & 28 & 14 \\ 8 & -6 & 34 & 38 & 18 \\ 12 & -10 & 50 & 41 & 23 \\ 14 & -21 & 49 & 29 & 33 \end{bmatrix}$. Construct a matrix N whose columns form a basis for Nul A, and construct a matrix R whose *rows* form a basis for Row A (see Section 4.6 for details). Perform a matrix computation with N and R that illustrates a fact from Theorem 3.

Solutions to Practice Problems

1. $\mathbf{a}\cdot\mathbf{b} = 7$, $\mathbf{a}\cdot\mathbf{a} = 5$. Hence $\dfrac{\mathbf{a}\cdot\mathbf{b}}{\mathbf{a}\cdot\mathbf{a}} = \dfrac{7}{5}$, and $\left(\dfrac{\mathbf{a}\cdot\mathbf{b}}{\mathbf{a}\cdot\mathbf{a}}\right)\mathbf{a} = \dfrac{7}{5}\mathbf{a} = \begin{bmatrix} -14/5 \\ 7/5 \end{bmatrix}$.

2. a. Scale $\mathbf{c}$, multiplying by 3 to get $\mathbf{y} = \begin{bmatrix} 4 \\ -3 \\ 2 \end{bmatrix}$. Compute $\|\mathbf{y}\|^2 = 29$ and $\|\mathbf{y}\| = \sqrt{29}$. The unit vector in the direction of both $\mathbf{c}$ and $\mathbf{y}$ is $\mathbf{u} = \dfrac{1}{\|\mathbf{y}\|}\mathbf{y} = \begin{bmatrix} 4/\sqrt{29} \\ -3/\sqrt{29} \\ 2/\sqrt{29} \end{bmatrix}$.

b. $\mathbf{d}$ is orthogonal to $\mathbf{c}$, because

$$\mathbf{d}\cdot\mathbf{c} = \begin{bmatrix} 5 \\ 6 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 4/3 \\ -1 \\ 2/3 \end{bmatrix} = \frac{20}{3} - 6 - \frac{2}{3} = 0$$

c. $\mathbf{d}$ is orthogonal to $\mathbf{u}$, because $\mathbf{u}$ has the form $k\mathbf{c}$ for some k, and

$$\mathbf{d}\cdot\mathbf{u} = \mathbf{d}\cdot(k\mathbf{c}) = k(\mathbf{d}\cdot\mathbf{c}) = k(0) = 0$$

3. If $W \neq \{\mathbf{0}\}$, let $\{\mathbf{b}_1, \ldots, \mathbf{b}_p\}$ be a basis for W, where $1 \leq p \leq n$. Let A be the $p \times n$ matrix having rows $\mathbf{b}_1^T, \ldots, \mathbf{b}_p^T$. It follows that W is the row space of A. Theorem 3 implies that $W^\perp = (\text{Row } A)^\perp = \text{Nul } A$ and hence $\dim W^\perp = \dim \text{Nul } A$. Thus, $\dim W + \dim W^\perp = \dim \text{Row } A + \dim \text{Nul } A = \text{rank } A + \dim \text{Nul } A = n$, by the Rank Theorem. If $W = \{\mathbf{0}\}$, then $W^\perp = \mathbb{R}^n$, and the result follows.

6.2 Orthogonal Sets

A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ in $\mathbb{R}^n$ is said to be an **orthogonal set** if each pair of distinct vectors from the set is orthogonal, that is, if $\mathbf{u}_i\cdot\mathbf{u}_j = 0$ whenever $i \neq j$.

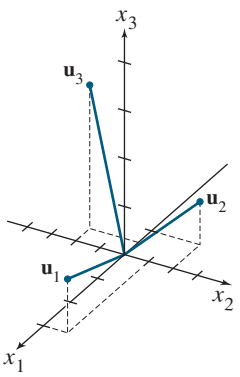


FIGURE 1

EXAMPLE 1 Show that $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set, where

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix}$$

SOLUTION Consider the three possible pairs of distinct vectors, namely $\{\mathbf{u}_1, \mathbf{u}_2\}$, $\{\mathbf{u}_1, \mathbf{u}_3\}$, and $\{\mathbf{u}_2, \mathbf{u}_3\}$.

$$\mathbf{u}_1\cdot\mathbf{u}_2 = 3(-1) + 1(2) + 1(1) = 0$$
$$\mathbf{u}_1\cdot\mathbf{u}_3 = 3\left(-\tfrac{1}{2}\right) + 1(-2) + 1\left(\tfrac{7}{2}\right) = 0$$
$$\mathbf{u}_2\cdot\mathbf{u}_3 = -1\left(-\tfrac{1}{2}\right) + 2(-2) + 1\left(\tfrac{7}{2}\right) = 0$$

Each pair of distinct vectors is orthogonal, and so $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set. See Figure 1; the three line segments are mutually perpendicular. ■

THEOREM 4

If $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent and hence is a basis for the subspace spanned by S.

PROOF If $\mathbf{0} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$ for some scalars $c_1, \ldots, c_p$, then

$$\begin{aligned} 0 = \mathbf{0}\cdot\mathbf{u}_1 &= (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p)\cdot\mathbf{u}_1 \\ &= (c_1\mathbf{u}_1)\cdot\mathbf{u}_1 + (c_2\mathbf{u}_2)\cdot\mathbf{u}_1 + \cdots + (c_p\mathbf{u}_p)\cdot\mathbf{u}_1 \\ &= c_1(\mathbf{u}_1\cdot\mathbf{u}_1) + c_2(\mathbf{u}_2\cdot\mathbf{u}_1) + \cdots + c_p(\mathbf{u}_p\cdot\mathbf{u}_1) \\ &= c_1(\mathbf{u}_1\cdot\mathbf{u}_1) \end{aligned}$$

because $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_2, \ldots, \mathbf{u}_p$. Since $\mathbf{u}_1$ is nonzero, $\mathbf{u}_1\cdot\mathbf{u}_1$ is not zero and so $c_1 = 0$. Similarly, $c_2, \ldots, c_p$ must be zero. Thus S is linearly independent. ■

DEFINITION

An **orthogonal basis** for a subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.

The next theorem suggests why an orthogonal basis is much nicer than other bases. The weights in a linear combination can be computed easily.

THEOREM 5

Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each $\mathbf{y}$ in W, the weights in the linear combination

$$\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$$

are given by

$$c_j = \frac{\mathbf{y}\cdot\mathbf{u}_j}{\mathbf{u}_j\cdot\mathbf{u}_j} \qquad (j = 1, \ldots, p)$$

PROOF As in the preceding proof, the orthogonality of $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ shows that

$$\mathbf{y}\cdot\mathbf{u}_1 = (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p)\cdot\mathbf{u}_1 = c_1(\mathbf{u}_1\cdot\mathbf{u}_1)$$

Since $\mathbf{u}_1\cdot\mathbf{u}_1$ is not zero, the equation above can be solved for c_1. To find c_j for $j = 2, \ldots, p$, compute $\mathbf{y}\cdot\mathbf{u}_j$ and solve for c_j. ■

EXAMPLE 2 The set $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ in Example 1 is an orthogonal basis for $\mathbb{R}^3$. Express the vector $\mathbf{y} = \begin{bmatrix} 6 \\ 1 \\ -8 \end{bmatrix}$ as a linear combination of the vectors in S.

SOLUTION Compute

$$\begin{aligned} \mathbf{y}\cdot\mathbf{u}_1 &= 11, & \mathbf{y}\cdot\mathbf{u}_2 &= -12, & \mathbf{y}\cdot\mathbf{u}_3 &= -33 \\ \mathbf{u}_1\cdot\mathbf{u}_1 &= 11, & \mathbf{u}_2\cdot\mathbf{u}_2 &= 6, & \mathbf{u}_3\cdot\mathbf{u}_3 &= 33/2 \end{aligned}$$

By Theorem 5,

$$\begin{aligned} \mathbf{y} &= \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 + \frac{\mathbf{y}\cdot\mathbf{u}_3}{\mathbf{u}_3\cdot\mathbf{u}_3}\mathbf{u}_3 \\ &= \frac{11}{11}\mathbf{u}_1 + \frac{-12}{6}\mathbf{u}_2 + \frac{-33}{33/2}\mathbf{u}_3 \\ &= \mathbf{u}_1 - 2\mathbf{u}_2 - 2\mathbf{u}_3 \end{aligned}$$

■

Notice how easy it is to compute the weights needed to build $\mathbf{y}$ from an orthogonal basis. If the basis were not orthogonal, it would be necessary to solve a system of linear equations in order to find the weights, as in Chapter 1.

We turn next to a construction that will become a key step in many calculations involving orthogonality, and it will lead to a geometric interpretation of Theorem 5.

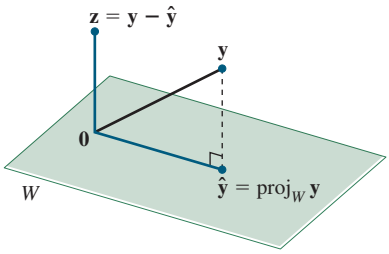


FIGURE 2
Finding α to make $\mathbf{y} - \hat{\mathbf{y}}$ orthogonal to $\mathbf{u}$.

An Orthogonal Projection

Given a nonzero vector $\mathbf{u}$ in $\mathbb{R}^n$, consider the problem of decomposing a vector $\mathbf{y}$ in $\mathbb{R}^n$ into the sum of two vectors, one a multiple of $\mathbf{u}$ and the other orthogonal to $\mathbf{u}$. We wish to write

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \tag{1}$$

where $\hat{\mathbf{y}} = \alpha\mathbf{u}$ for some scalar α and $\mathbf{z}$ is some vector orthogonal to $\mathbf{u}$. See Figure 2. Given any scalar α, let $\mathbf{z} = \mathbf{y} - \alpha\mathbf{u}$, so that (1) is satisfied. Then $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to $\mathbf{u}$ if and only if

$$0 = (\mathbf{y} - \alpha\mathbf{u})\cdot\mathbf{u} = \mathbf{y}\cdot\mathbf{u} - (\alpha\mathbf{u})\cdot\mathbf{u} = \mathbf{y}\cdot\mathbf{u} - \alpha(\mathbf{u}\cdot\mathbf{u})$$

That is, (1) is satisfied with $\mathbf{z}$ orthogonal to $\mathbf{u}$ if and only if $\alpha = \dfrac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}$ and $\hat{\mathbf{y}} = \dfrac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u}$. The vector $\hat{\mathbf{y}}$ is called the **orthogonal projection of y onto u**, and the vector $\mathbf{z}$ is called the **component of y orthogonal to u**.

If c is any nonzero scalar and if $\mathbf{u}$ is replaced by $c\mathbf{u}$ in the definition of $\hat{\mathbf{y}}$, then the orthogonal projection of $\mathbf{y}$ onto $c\mathbf{u}$ is exactly the same as the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$ (Exercise 39). Hence this projection is determined by the *subspace* L spanned by $\mathbf{u}$ (the line through $\mathbf{u}$ and $\mathbf{0}$). Sometimes $\hat{\mathbf{y}}$ is denoted by $\text{proj}_L\,\mathbf{y}$ and is called the **orthogonal projection of y onto** L. That is,

$$\hat{\mathbf{y}} = \text{proj}_L\,\mathbf{y} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u} \tag{2}$$

EXAMPLE 3 Let $\mathbf{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$. Find the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$. Then write $\mathbf{y}$ as the sum of two orthogonal vectors, one in Span $\{\mathbf{u}\}$ and one orthogonal to $\mathbf{u}$.

SOLUTION Compute

$$\mathbf{y}\cdot\mathbf{u} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}\cdot\begin{bmatrix} 4 \\ 2 \end{bmatrix} = 40$$

$$\mathbf{u}\cdot\mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}\cdot\begin{bmatrix} 4 \\ 2 \end{bmatrix} = 20$$

The orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$ is

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\mathbf{u} = \frac{40}{20}\mathbf{u} = 2\begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

and the component of $\mathbf{y}$ orthogonal to $\mathbf{u}$ is

$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 7 \\ 6 \end{bmatrix} - \begin{bmatrix} 8 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

The sum of these two vectors is $\mathbf{y}$. That is,

$$\begin{bmatrix} 7 \\ 6 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix} + \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

$$\uparrow \quad \uparrow \quad \uparrow$$

$$\mathbf{y} \quad \hat{\mathbf{y}} \quad (\mathbf{y} - \hat{\mathbf{y}})$$

This decomposition of $\mathbf{y}$ is illustrated in Figure 3. *Note:* If the calculations above are correct, then $\{\hat{\mathbf{y}}, \mathbf{y} - \hat{\mathbf{y}}\}$ will be an orthogonal set. As a check, compute

$$\hat{\mathbf{y}}\cdot(\mathbf{y} - \hat{\mathbf{y}}) = \begin{bmatrix} 8 \\ 4 \end{bmatrix}\cdot\begin{bmatrix} -1 \\ 2 \end{bmatrix} = -8 + 8 = 0$$

■

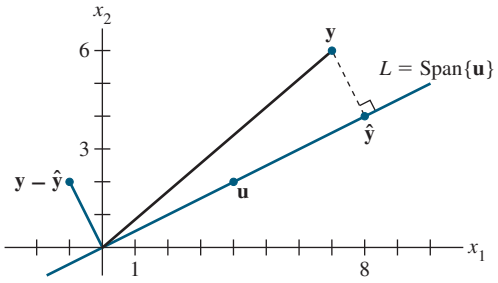


FIGURE 3 The orthogonal projection of $\mathbf{y}$ onto a line L through the origin.

Since the line segment in Figure 3 between $\mathbf{y}$ and $\hat{\mathbf{y}}$ is perpendicular to L, by construction of $\hat{\mathbf{y}}$, the point identified with $\hat{\mathbf{y}}$ is the closest point of L to $\mathbf{y}$. (This can be proved from geometry. We will assume this for $\mathbb{R}^2$ now and prove it for $\mathbb{R}^n$ in Section 6.3.)

EXAMPLE 4 Find the distance in Figure 3 from $\mathbf{y}$ to L.

SOLUTION The distance from $\mathbf{y}$ to L is the length of the perpendicular line segment from $\mathbf{y}$ to the orthogonal projection $\hat{\mathbf{y}}$. This length equals the length of $\mathbf{y} - \hat{\mathbf{y}}$. Thus the distance is

$$\|\mathbf{y} - \hat{\mathbf{y}}\| = \sqrt{(-1)^2 + 2^2} = \sqrt{5}$$

■

A Geometric Interpretation of Theorem 5

The formula for the orthogonal projection $\hat{\mathbf{y}}$ in (2) has the same appearance as each of the terms in Theorem 5. Thus Theorem 5 decomposes a vector $\mathbf{y}$ into a sum of orthogonal projections onto one-dimensional subspaces.

It is easy to visualize the case in which $W = \mathbb{R}^2 = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, with $\mathbf{u}_1$ and $\mathbf{u}_2$ orthogonal. Any $\mathbf{y}$ in $\mathbb{R}^2$ can be written in the form

$$\mathbf{y} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 \tag{3}$$

The first term in (3) is the projection of $\mathbf{y}$ onto the subspace spanned by $\mathbf{u}_1$ (the line through $\mathbf{u}_1$ and the origin), and the second term is the projection of $\mathbf{y}$ onto the subspace spanned by $\mathbf{u}_2$. Thus (3) expresses $\mathbf{y}$ as the sum of its projections onto the (orthogonal) axes determined by $\mathbf{u}_1$ and $\mathbf{u}_2$. See Figure 4.

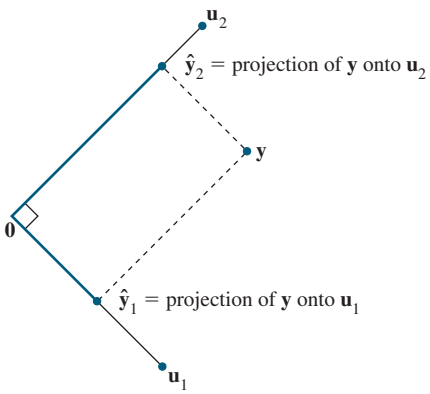


FIGURE 4 A vector decomposed into the sum of two projections.

Theorem 5 decomposes each $\mathbf{y}$ in Span $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ into the sum of p projections onto one-dimensional subspaces that are mutually orthogonal.

Decomposing a Force into Component Forces

The decomposition in Figure 4 can occur in physics when some sort of force is applied to an object. Choosing an appropriate coordinate system allows the force to be represented by a vector $\mathbf{y}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$. Often the problem involves some particular direction of interest, which is represented by another vector $\mathbf{u}$. For instance, if the object is moving in a straight line when the force is applied, the vector $\mathbf{u}$ might point in the direction of movement, as in Figure 5. A key step in the problem is to decompose the force into a component in the direction of $\mathbf{u}$ and a component orthogonal to $\mathbf{u}$. The calculations would be analogous to those previously made in Example 3.

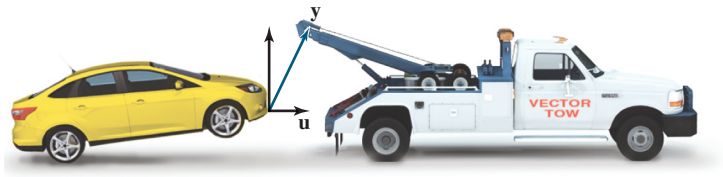


FIGURE 5

Orthonormal Sets

A set $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an **orthonormal set** if it is an orthogonal set of unit vectors. If W is the subspace spanned by such a set, then $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an **orthonormal basis** for W, since the set is automatically linearly independent, by Theorem 4.

The simplest example of an orthonormal set is the standard basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ for $\mathbb{R}^n$. Any nonempty subset of $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is orthonormal, too. Here is a more complicated example.

EXAMPLE 5 Show that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal basis of $\mathbb{R}^3$, where

$$\mathbf{v}_1 = \begin{bmatrix} 3/\sqrt{11} \\ 1/\sqrt{11} \\ 1/\sqrt{11} \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -1/\sqrt{66} \\ -4/\sqrt{66} \\ 7/\sqrt{66} \end{bmatrix}$$

SOLUTION Compute

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = -3/\sqrt{66} + 2/\sqrt{66} + 1/\sqrt{66} = 0$$
$$\mathbf{v}_1 \cdot \mathbf{v}_3 = -3/\sqrt{726} - 4/\sqrt{726} + 7/\sqrt{726} = 0$$
$$\mathbf{v}_2 \cdot \mathbf{v}_3 = 1/\sqrt{396} - 8/\sqrt{396} + 7/\sqrt{396} = 0$$

Thus $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal set. Also,

$$\mathbf{v}_1 \cdot \mathbf{v}_1 = 9/11 + 1/11 + 1/11 = 1$$
$$\mathbf{v}_2 \cdot \mathbf{v}_2 = 1/6 + 4/6 + 1/6 = 1$$
$$\mathbf{v}_3 \cdot \mathbf{v}_3 = 1/66 + 16/66 + 49/66 = 1$$

which shows that $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are unit vectors. Thus $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal set. Since the set is linearly independent, its three vectors form a basis for $\mathbb{R}^3$. See Figure 6.

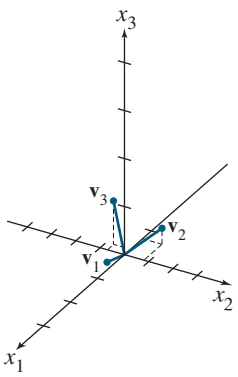


FIGURE 6

When the vectors in an orthogonal set of nonzero vectors are *normalized* to have unit length, the new vectors will still be orthogonal, and hence the new set will be an orthonormal set. See Exercise 40. It is easy to check that the vectors in Figure 6 (Example 5) are simply the unit vectors in the directions of the vectors in Figure 1 (Example 1).

Matrices whose columns form an orthonormal set are important in applications and in computer algorithms for matrix computations. Their main properties are given in Theorems 6 and 7.

THEOREM 6

An $m \times n$ matrix U has orthonormal columns if and only if $U^TU = I$.

PROOF To simplify notation, we suppose that U has only three columns, each a vector in $\mathbb{R}^m$. The proof of the general case is essentially the same. Let $U = [\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3\,]$ and compute

$$U^TU = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \mathbf{u}_3^T \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 & \mathbf{u}_1^T\mathbf{u}_3 \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 & \mathbf{u}_2^T\mathbf{u}_3 \\ \mathbf{u}_3^T\mathbf{u}_1 & \mathbf{u}_3^T\mathbf{u}_2 & \mathbf{u}_3^T\mathbf{u}_3 \end{bmatrix} \tag{4}$$

The entries in the matrix at the right are inner products, using transpose notation. The columns of U are orthogonal if and only if

$$\mathbf{u}_1^T\mathbf{u}_2 = \mathbf{u}_2^T\mathbf{u}_1 = 0, \quad \mathbf{u}_1^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_1 = 0, \quad \mathbf{u}_2^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_2 = 0 \tag{5}$$

The columns of U all have unit length if and only if

$$\mathbf{u}_1^T\mathbf{u}_1 = 1, \quad \mathbf{u}_2^T\mathbf{u}_2 = 1, \quad \mathbf{u}_3^T\mathbf{u}_3 = 1 \tag{6}$$

The theorem follows immediately from (4)–(6). ■

THEOREM 7

Let U be an $m \times n$ matrix with orthonormal columns, and let $\mathbf{x}$ and $\mathbf{y}$ be in $\mathbb{R}^n$. Then

a. $\|U\mathbf{x}\| = \|\mathbf{x}\|$

b. $(U\mathbf{x})\cdot(U\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$

c. $(U\mathbf{x})\cdot(U\mathbf{y}) = 0$ if and only if $\mathbf{x}\cdot\mathbf{y} = 0$

Properties (a) and (c) say that the linear mapping $\mathbf{x} \mapsto U\mathbf{x}$ preserves lengths and orthogonality. These properties are crucial for many computer algorithms. See Exercise 33 for the proof of Theorem 7.

EXAMPLE 6 Let $U = \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix}$. Notice that U has orthonormal columns and

$$U^TU = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 2/3 & -2/3 & 1/3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Verify that $\|U\mathbf{x}\| = \|\mathbf{x}\|$.

SOLUTION

$$U\mathbf{x} = \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix} \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$$

$$\|U\mathbf{x}\| = \sqrt{9+1+1} = \sqrt{11}$$

$$\|\mathbf{x}\| = \sqrt{2+9} = \sqrt{11}$$

■

Theorems 6 and 7 are particularly useful when applied to *square* matrices. An **orthogonal matrix** is a square invertible matrix U such that $U^{-1} = U^T$. By Theorem 6, such a matrix has orthonormal columns.[1] It is easy to see that any *square* matrix with orthonormal columns is an orthogonal matrix. Surprisingly, such a matrix must have orthonormal *rows*, too. See Exercises 35 and 36. Orthogonal matrices will appear frequently in Chapter 7.

EXAMPLE 7 The matrix

$$U = \begin{bmatrix} 3/\sqrt{11} & -1/\sqrt{6} & -1/\sqrt{66} \\ 1/\sqrt{11} & 2/\sqrt{6} & -4/\sqrt{66} \\ 1/\sqrt{11} & 1/\sqrt{6} & 7/\sqrt{66} \end{bmatrix}$$

is an orthogonal matrix because it is square and because its columns are orthonormal, by Example 5. Verify that the rows are orthonormal, too! ■

Practice Problems

1. Let $\mathbf{u}_1 = \begin{bmatrix} -1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$. Show that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthonormal basis for $\mathbb{R}^2$.

2. Let $\mathbf{y}$ and L be as in Example 3 and Figure 3. Compute the orthogonal projection $\hat{\mathbf{y}}$ of $\mathbf{y}$ onto L using $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ instead of the $\mathbf{u}$ in Example 3.

3. Let U and $\mathbf{x}$ be as in Example 6, and let $\mathbf{y} = \begin{bmatrix} -3\sqrt{2} \\ 6 \end{bmatrix}$. Verify that $U\mathbf{x} \cdot U\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$.

4. Let U be an $n \times n$ matrix with orthonormal columns. Show that $\det U = \pm 1$.

6.2 Exercises

In Exercises 1–6, determine which sets of vectors are orthogonal.

1. $\begin{bmatrix} -1 \\ 4 \\ -3 \end{bmatrix}, \begin{bmatrix} 5 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ -4 \\ -7 \end{bmatrix}$

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

3. $\begin{bmatrix} 2 \\ -7 \\ -1 \end{bmatrix}, \begin{bmatrix} -6 \\ -3 \\ 9 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}$

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

5. $\begin{bmatrix} 3 \\ -2 \\ 1 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \\ -3 \\ 4 \end{bmatrix}, \begin{bmatrix} 3 \\ 8 \\ 7 \\ 0 \end{bmatrix}$

6. $\begin{bmatrix} 5 \\ -4 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} -4 \\ 1 \\ -3 \\ 8 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \\ 5 \\ -1 \end{bmatrix}$

In Exercises 7–10, show that $\{\mathbf{u}_1, \mathbf{u}_2\}$ or $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal basis for $\mathbb{R}^2$ or $\mathbb{R}^3$, respectively. Then express $\mathbf{x}$ as a linear combination of the $\mathbf{u}$'s.

7. $\mathbf{u}_1 = \begin{bmatrix} 2 \\ -3 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 6 \\ 4 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} 9 \\ -7 \end{bmatrix}$

[1] A better name might be *orthonormal matrix*, and this term is found in some statistics texts. However, *orthogonal matrix* is the standard term in linear algebra.

8. $\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 6 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$

9. $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 1 \\ -4 \\ 1 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} 4 \\ 2 \\ 4 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} 6 \\ 4 \\ -2 \end{bmatrix}$

10. $\mathbf{u}_1 = \begin{bmatrix} 4 \\ -4 \\ 0 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} 3 \\ -4 \\ 7 \end{bmatrix}$

11. Compute the orthogonal projection of $\begin{bmatrix} 1 \\ 7 \end{bmatrix}$ onto the line through $\begin{bmatrix} -4 \\ 2 \end{bmatrix}$ and the origin.

12. Compute the orthogonal projection of $\begin{bmatrix} -3 \\ 4 \end{bmatrix}$ onto the line through $\begin{bmatrix} 1 \\ -3 \end{bmatrix}$ and the origin.

13. Let $\mathbf{y} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 4 \\ -7 \end{bmatrix}$. Write $\mathbf{y}$ as the sum of two orthogonal vectors, one in Span $\{\mathbf{u}\}$ and one orthogonal to $\mathbf{u}$.

14. Let $\mathbf{y} = \begin{bmatrix} 2 \\ 6 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 6 \\ 1 \end{bmatrix}$. Write $\mathbf{y}$ as the sum of a vector in Span $\{\mathbf{u}\}$ and a vector orthogonal to $\mathbf{u}$.

15. Let $\mathbf{y} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 8 \\ 6 \end{bmatrix}$. Compute the distance from $\mathbf{y}$ to the line through $\mathbf{u}$ and the origin.

16. Let $\mathbf{y} = \begin{bmatrix} -1 \\ 7 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$. Compute the distance from $\mathbf{y}$ to the line through $\mathbf{u}$ and the origin.

In Exercises 17–22, determine which sets of vectors are orthonormal. If a set is only orthogonal, normalize the vectors to produce an orthonormal set.

17. $\begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}, \begin{bmatrix} -1/2 \\ 0 \\ 1/2 \end{bmatrix}$

18. $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$

19. $\begin{bmatrix} -.6 \\ .8 \end{bmatrix}, \begin{bmatrix} .8 \\ .6 \end{bmatrix}$

20. $\begin{bmatrix} 4/3 \\ 7/3 \\ 4/3 \end{bmatrix}, \begin{bmatrix} 7/3 \\ -4/3 \\ 0 \end{bmatrix}$

21. $\begin{bmatrix} 1/\sqrt{10} \\ 3/\sqrt{20} \\ 3/\sqrt{20} \end{bmatrix}, \begin{bmatrix} 3/\sqrt{10} \\ -1/\sqrt{20} \\ -1/\sqrt{20} \end{bmatrix}, \begin{bmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$

22. $\begin{bmatrix} 1/\sqrt{18} \\ 4/\sqrt{18} \\ 1/\sqrt{18} \end{bmatrix}, \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ -1/\sqrt{2} \end{bmatrix}, \begin{bmatrix} -2/3 \\ 1/3 \\ -2/3 \end{bmatrix}$

In Exercises 23–32, all vectors are in $\mathbb{R}^n$. Mark each statement True or False **(T/F)**. Justify each answer.

23. (T/F) Not every linearly independent set in $\mathbb{R}^n$ is an orthogonal set.

24. (T/F) Not every orthogonal set in $\mathbb{R}^n$ is linearly independent.

25. (T/F) If $\mathbf{y}$ is a linear combination of nonzero vectors from an orthogonal set, then the weights in the linear combination can be computed without row operations on a matrix.

26. (T/F) If a set $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ has the property that $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ whenever $i \neq j$, then S is an orthonormal set.

27. (T/F) If the vectors in an orthogonal set of nonzero vectors are normalized, then some of the new vectors may not be orthogonal.

28. (T/F) If the columns of an $m \times n$ matrix A are orthonormal, then the linear mapping $\mathbf{x} \mapsto A\mathbf{x}$ preserves lengths.

29. (T/F) A matrix with orthonormal columns is an orthogonal matrix.

30. (T/F) The orthogonal projection of $\mathbf{y}$ onto $\mathbf{v}$ is the same as the orthogonal projection of $\mathbf{y}$ onto $c\mathbf{v}$ whenever $c \neq 0$.

31. (T/F) If L is a line through $\mathbf{0}$ and if $\hat{\mathbf{y}}$ is the orthogonal projection of $\mathbf{y}$ onto L, then $\|\hat{\mathbf{y}}\|$ gives the distance from $\mathbf{y}$ to L.

32. (T/F) An orthogonal matrix is invertible.

33. Prove Theorem 7. [*Hint:* For (a), compute $\|U\mathbf{x}\|^2$, or prove (b) first.]

34. Suppose W is a subspace of $\mathbb{R}^n$ spanned by n nonzero orthogonal vectors. Explain why $W = \mathbb{R}^n$.

35. Let U be a square matrix with orthonormal columns. Explain why U is invertible. (Mention the theorems you use.)

36. Let U be an $n \times n$ orthogonal matrix. Show that the rows of U form an orthonormal basis of $\mathbb{R}^n$.

37. Let U and V be $n \times n$ orthogonal matrices. Explain why UV is an orthogonal matrix. [That is, explain why UV is invertible and its inverse is $(UV)^T$.]

38. Let U be an orthogonal matrix, and construct V by interchanging some of the columns of U. Explain why V is an orthogonal matrix.

39. Show that the orthogonal projection of a vector $\mathbf{y}$ onto a line L through the origin in $\mathbb{R}^2$ does not depend on the choice of the nonzero $\mathbf{u}$ in L used in the formula for $\hat{\mathbf{y}}$. To do this, suppose $\mathbf{y}$ and $\mathbf{u}$ are given and $\hat{\mathbf{y}}$ has been computed by formula (2) in this section. Replace $\mathbf{u}$ in that formula by $c\mathbf{u}$, where c is an unspecified nonzero scalar. Show that the new formula gives the same $\hat{\mathbf{y}}$.

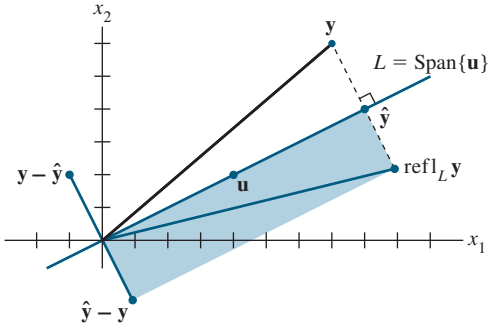
40. Let $\{\mathbf{v}_1, \mathbf{v}_2\}$ be an orthogonal set of nonzero vectors, and let c_1, c_2 be any nonzero scalars. Show that $\{c_1\mathbf{v}_1, c_2\mathbf{v}_2\}$ is also an orthogonal set. Since orthogonality of a set is defined in terms of pairs of vectors, this shows that if the vectors in an orthogonal set are normalized, the new set will still be orthogonal.

41. Given $\mathbf{u} \neq \mathbf{0}$ in $\mathbb{R}^n$, let $L = \text{Span}\{\mathbf{u}\}$. Show that the mapping $\mathbf{x} \mapsto \text{proj}_L \mathbf{x}$ is a linear transformation.

42. Given $\mathbf{u} \neq \mathbf{0}$ in $\mathbb{R}^n$, let $L = \text{Span}\{\mathbf{u}\}$. For $\mathbf{y}$ in $\mathbb{R}^n$, the **reflection of y in** L is the point $\text{refl}_L \mathbf{y}$ defined by

$$\text{refl}_L \mathbf{y} = 2 \text{ proj}_L \mathbf{y} - \mathbf{y}$$

See the figure, which shows that $\text{refl}_L \mathbf{y}$ is the sum of $\hat{\mathbf{y}} = \text{proj}_L \mathbf{y}$ and $\hat{\mathbf{y}} - \mathbf{y}$. Show that the mapping $\mathbf{y} \mapsto \text{refl}_L \mathbf{y}$ is a linear transformation.



The reflection of **y** in a line through the origin.

T 43. Show that the columns of the matrix A are orthogonal by making an appropriate matrix calculation. State the calculation you use.

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

T 44. In parts (a)–(d), let U be the matrix formed by normalizing each column of the matrix A in Exercise 43.

a. Compute $U^T U$ and $U U^T$. How do they differ?

b. Generate a random vector $\mathbf{y}$ in $\mathbb{R}^8$, and compute $\mathbf{p} = U U^T \mathbf{y}$ and $\mathbf{z} = \mathbf{y} - \mathbf{p}$. Explain why $\mathbf{p}$ is in Col A. Verify that $\mathbf{z}$ is orthogonal to $\mathbf{p}$.

c. Verify that $\mathbf{z}$ is orthogonal to each column of U.

d. Notice that $\mathbf{y} = \mathbf{p} + \mathbf{z}$, with $\mathbf{p}$ in Col A. Explain why $\mathbf{z}$ is in $(\text{Col } A)^{\perp}$. (The significance of this decomposition of $\mathbf{y}$ will be explained in the next section.)

Solutions to Practice Problems

1. The vectors are orthogonal because

$$\mathbf{u}_1 \cdot \mathbf{u}_2 = -2/5 + 2/5 = 0$$

They are unit vectors because

$$\|\mathbf{u}_1\|^2 = (-1/\sqrt{5})^2 + (2/\sqrt{5})^2 = 1/5 + 4/5 = 1$$
$$\|\mathbf{u}_2\|^2 = (2/\sqrt{5})^2 + (1/\sqrt{5})^2 = 4/5 + 1/5 = 1$$

In particular, the set $\{\mathbf{u}_1, \mathbf{u}_2\}$ is linearly independent, and hence is a basis for $\mathbb{R}^2$ since there are two vectors in the set.

2. When $\mathbf{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$,

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \mathbf{u} = \frac{20}{5} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 4 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

This is the same $\hat{\mathbf{y}}$ found in Example 3. The orthogonal projection does not depend on the **u** chosen on the line. See Exercise 39.

3. $U\mathbf{y} = \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix} \begin{bmatrix} -3\sqrt{2} \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \\ -7 \\ 2 \end{bmatrix}$

Also, from Example 6, $\mathbf{x} = \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix}$ and $U\mathbf{x} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$. Hence

$$U\mathbf{x} \cdot U\mathbf{y} = 3 + 7 + 2 = 12, \quad \text{and} \quad \mathbf{x} \cdot \mathbf{y} = -6 + 18 = 12$$

STUDY GUIDE offers additional resources for mastering the concepts around an orthogonal basis.

4. Since U is an $n \times n$ matrix with orthonormal columns, by Theorem 6, $U^T U = I$. Taking the determinant of the left side of this equation, and applying Theorems 5 and 6 from Section 3.2 results in $\det U^T U = (\det U^T)(\det U) = (\det U)(\det U) = (\det U)^2$. Recall $\det I = 1$. Putting the two sides of the equation back together results in $(\det U)^2 = 1$ and hence $\det U = \pm 1$.

6.3 Orthogonal Projections

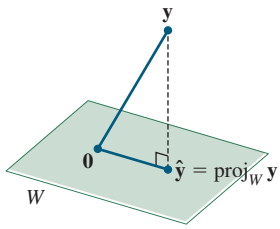


FIGURE 1

The orthogonal projection of a point in $\mathbb{R}^2$ onto a line through the origin has an important analogue in $\mathbb{R}^n$. Given a vector $\mathbf{y}$ and a subspace W in $\mathbb{R}^n$, there is a vector $\hat{\mathbf{y}}$ in W such that (1) $\hat{\mathbf{y}}$ is the unique vector in W for which $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to W, and (2) $\hat{\mathbf{y}}$ is the unique vector in W closest to $\mathbf{y}$. See Figure 1. These two properties of $\hat{\mathbf{y}}$ provide the key to finding least-squares solutions of linear systems.

To prepare for the first theorem, observe that whenever a vector $\mathbf{y}$ is written as a linear combination of vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ in $\mathbb{R}^n$, the terms in the sum for $\mathbf{y}$ can be grouped into two parts so that $\mathbf{y}$ can be written as

$$\mathbf{y} = \mathbf{z}_1 + \mathbf{z}_2$$

where $\mathbf{z}_1$ is a linear combination of some of the $\mathbf{u}_i$ and $\mathbf{z}_2$ is a linear combination of the rest of the $\mathbf{u}_i$. This idea is particularly useful when $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is an orthogonal basis. Recall from Section 6.1 that $W^\perp$ denotes the set of all vectors orthogonal to a subspace W.

EXAMPLE 1 Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_5\}$ be an orthogonal basis for $\mathbb{R}^5$ and let

$$\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_5\mathbf{u}_5$$

Consider the subspace $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, and write $\mathbf{y}$ as the sum of a vector $\mathbf{z}_1$ in W and a vector $\mathbf{z}_2$ in $W^\perp$.

SOLUTION Write

$$\mathbf{y} = \underbrace{c_1\mathbf{u}_1 + c_2\mathbf{u}_2}_{\mathbf{z}_1} + \underbrace{c_3\mathbf{u}_3 + c_4\mathbf{u}_4 + c_5\mathbf{u}_5}_{\mathbf{z}_2}$$

where $\mathbf{z}_1 = c_1\mathbf{u}_1 + c_2\mathbf{u}_2$ is in $\text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$

and $\mathbf{z}_2 = c_3\mathbf{u}_3 + c_4\mathbf{u}_4 + c_5\mathbf{u}_5$ is in $\text{Span}\{\mathbf{u}_3, \mathbf{u}_4, \mathbf{u}_5\}$.

To show that $\mathbf{z}_2$ is in $W^\perp$, it suffices to show that $\mathbf{z}_2$ is orthogonal to the vectors in the basis $\{\mathbf{u}_1, \mathbf{u}_2\}$ for W. (See Section 6.1.) Using properties of the inner product, compute

$$\begin{aligned}\mathbf{z}_2 \cdot \mathbf{u}_1 &= (c_3\mathbf{u}_3 + c_4\mathbf{u}_4 + c_5\mathbf{u}_5) \cdot \mathbf{u}_1 \\ &= c_3\mathbf{u}_3 \cdot \mathbf{u}_1 + c_4\mathbf{u}_4 \cdot \mathbf{u}_1 + c_5\mathbf{u}_5 \cdot \mathbf{u}_1 \\ &= 0\end{aligned}$$

because $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_3$, $\mathbf{u}_4$, and $\mathbf{u}_5$. A similar calculation shows that $\mathbf{z}_2 \cdot \mathbf{u}_2 = 0$. Thus $\mathbf{z}_2$ is in $W^\perp$. ■

The next theorem shows that the decomposition $\mathbf{y} = \mathbf{z}_1 + \mathbf{z}_2$ in Example 1 can be computed without having an orthogonal basis for $\mathbb{R}^n$. It is enough to have an orthogonal basis only for W.

THEOREM 8

The Orthogonal Decomposition Theorem

Let W be a subspace of $\mathbb{R}^n$. Then each $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \tag{1}$$

where $\hat{\mathbf{y}}$ is in W and $\mathbf{z}$ is in $W^{\perp}$. In fact, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is any orthogonal basis of W, then

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y}\cdot\mathbf{u}_p}{\mathbf{u}_p\cdot\mathbf{u}_p}\mathbf{u}_p \tag{2}$$

and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.

The vector $\hat{\mathbf{y}}$ in (2) is called the **orthogonal projection of y onto** W and often is written as $\text{proj}_W \mathbf{y}$. See Figure 2. When W is a one-dimensional subspace, the formula for $\hat{\mathbf{y}}$ matches the formula given in Section 6.2.

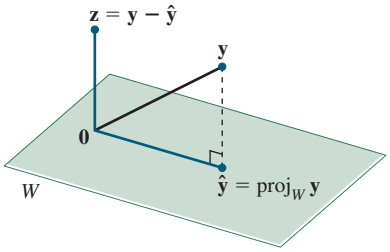


FIGURE 2 The orthogonal projection of $\mathbf{y}$ onto W.

PROOF Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ be any orthogonal basis for W, and define $\hat{\mathbf{y}}$ by (2).[1] Then $\hat{\mathbf{y}}$ is in W because $\hat{\mathbf{y}}$ is a linear combination of the basis $\mathbf{u}_1, \ldots, \mathbf{u}_p$. Let $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$. Since $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_2, \ldots, \mathbf{u}_p$, it follows from (2) that

$$\begin{aligned}\mathbf{z}\cdot\mathbf{u}_1 = (\mathbf{y} - \hat{\mathbf{y}})\cdot\mathbf{u}_1 &= \mathbf{y}\cdot\mathbf{u}_1 - \left(\frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\right)\mathbf{u}_1\cdot\mathbf{u}_1 - 0 - \cdots - 0 \\ &= \mathbf{y}\cdot\mathbf{u}_1 - \mathbf{y}\cdot\mathbf{u}_1 = 0\end{aligned}$$

Thus $\mathbf{z}$ is orthogonal to $\mathbf{u}_1$. Similarly, $\mathbf{z}$ is orthogonal to each $\mathbf{u}_j$ in the basis for W. Hence $\mathbf{z}$ is orthogonal to every vector in W. That is, $\mathbf{z}$ is in $W^{\perp}$.

To show that the decomposition in (1) is unique, suppose $\mathbf{y}$ can also be written as $\mathbf{y} = \hat{\mathbf{y}}_1 + \mathbf{z}_1$, with $\hat{\mathbf{y}}_1$ in W and $\mathbf{z}_1$ in $W^{\perp}$. Then $\hat{\mathbf{y}} + \mathbf{z} = \hat{\mathbf{y}}_1 + \mathbf{z}_1$ (since both sides equal $\mathbf{y}$), and so

$$\hat{\mathbf{y}} - \hat{\mathbf{y}}_1 = \mathbf{z}_1 - \mathbf{z}$$

This equality shows that the vector $\mathbf{v} = \hat{\mathbf{y}} - \hat{\mathbf{y}}_1$ is in W and in $W^{\perp}$ (because $\mathbf{z}_1$ and $\mathbf{z}$ are both in $W^{\perp}$, and $W^{\perp}$ is a subspace). Hence $\mathbf{v}\cdot\mathbf{v} = 0$, which shows that $\mathbf{v} = \mathbf{0}$. This proves that $\hat{\mathbf{y}} = \hat{\mathbf{y}}_1$ and also $\mathbf{z}_1 = \mathbf{z}$. ■

The uniqueness of the decomposition (1) shows that the orthogonal projection $\hat{\mathbf{y}}$ depends only on W and not on the particular basis used in (2).

[1] We may assume that W is not the zero subspace, for otherwise $W^{\perp} = \mathbb{R}^n$ and (1) is simply $\mathbf{y} = \mathbf{0} + \mathbf{y}$. The next section will show that any nonzero subspace of $\mathbb{R}^n$ has an orthogonal basis.

EXAMPLE 2 Let $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Observe that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Write $\mathbf{y}$ as the sum of a vector in W and a vector orthogonal to W.

SOLUTION The orthogonal projection of $\mathbf{y}$ onto W is

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2$$

$$= \frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{15}{30}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix}$$

Also

$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} = \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$$

Theorem 8 ensures that $\mathbf{y} - \hat{\mathbf{y}}$ is in $W^{\perp}$. To check the calculations, however, it is a good idea to verify that $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to both $\mathbf{u}_1$ and $\mathbf{u}_2$ and hence to all of W. The desired decomposition of $\mathbf{y}$ is

$$\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} + \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$$

■

A Geometric Interpretation of the Orthogonal Projection

When W is a one-dimensional subspace, the formula (2) for $\text{proj}_W \mathbf{y}$ contains just one term. Thus, when $\dim W > 1$, each term in (2) is itself an orthogonal projection of $\mathbf{y}$ onto a one-dimensional subspace spanned by one of the $\mathbf{u}$'s in the basis for W. Figure 3 illustrates this when W is a subspace of $\mathbb{R}^3$ spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$. Here $\hat{\mathbf{y}}_1$ and $\hat{\mathbf{y}}_2$ denote the projections of $\mathbf{y}$ onto the lines spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$, respectively. The orthogonal projection $\hat{\mathbf{y}}$ of $\mathbf{y}$ onto W is the sum of the projections of $\mathbf{y}$ onto one-dimensional subspaces that are orthogonal to each other. The vector $\hat{\mathbf{y}}$ in Figure 3 corresponds to the vector $\mathbf{y}$ in Figure 4 of Section 6.2, because now it is $\hat{\mathbf{y}}$ that is in W.

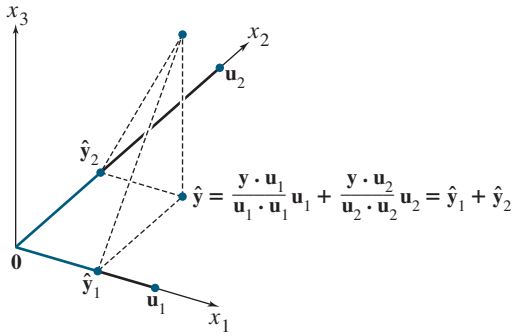


FIGURE 3 The orthogonal projection of $\mathbf{y}$ is the sum of its projections onto one-dimensional subspaces that are mutually orthogonal.

Properties of Orthogonal Projections

If $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthogonal basis for W and if $\mathbf{y}$ happens to be in W, then the formula for $\text{proj}_W \mathbf{y}$ is exactly the same as the representation of $\mathbf{y}$ given in Theorem 5 in Section 6.2. In this case, $\text{proj}_W \mathbf{y} = \mathbf{y}$.

If $\mathbf{y}$ is in $W = \text{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$, then $\text{proj}_W \mathbf{y} = \mathbf{y}$.

This fact also follows from the next theorem.

THEOREM 9

The Best Approximation Theorem

Let W be a subspace of $\mathbb{R}^n$, let $\mathbf{y}$ be any vector in $\mathbb{R}^n$, and let $\hat{\mathbf{y}}$ be the orthogonal projection of $\mathbf{y}$ onto W. Then $\hat{\mathbf{y}}$ is the closest point in W to $\mathbf{y}$, in the sense that

$$\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\| \tag{3}$$

for all $\mathbf{v}$ in W distinct from $\hat{\mathbf{y}}$.

The vector $\hat{\mathbf{y}}$ in Theorem 9 is called **the best approximation to y by elements of** W. Later sections in the text will examine problems where a given $\mathbf{y}$ must be replaced, or *approximated*, by a vector $\mathbf{v}$ in some fixed subspace W. The distance from $\mathbf{y}$ to $\mathbf{v}$, given by $\|\mathbf{y} - \mathbf{v}\|$, can be regarded as the "error" of using $\mathbf{v}$ in place of $\mathbf{y}$. Theorem 9 says that this error is minimized when $\mathbf{v} = \hat{\mathbf{y}}$.

Inequality (3) leads to a new proof that $\hat{\mathbf{y}}$ does not depend on the particular orthogonal basis used to compute it. If a different orthogonal basis for W was used to construct an orthogonal projection of $\mathbf{y}$, then this projection would also be the closest point in W to $\mathbf{y}$, namely $\hat{\mathbf{y}}$.

PROOF Take $\mathbf{v}$ in W distinct from $\hat{\mathbf{y}}$. See Figure 4. Then $\hat{\mathbf{y}} - \mathbf{v}$ is in W. By the Orthogonal Decomposition Theorem, $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to W. In particular, $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to $\hat{\mathbf{y}} - \mathbf{v}$ (which is in W). Since

$$\mathbf{y} - \mathbf{v} = (\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \mathbf{v})$$

the Pythagorean Theorem gives

$$\|\mathbf{y} - \mathbf{v}\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}} - \mathbf{v}\|^2$$

(See the right triangle outlined in teal in Figure 4. The length of each side is labeled.) Now $\|\hat{\mathbf{y}} - \mathbf{v}\|^2 > 0$ because $\hat{\mathbf{y}} - \mathbf{v} \neq \mathbf{0}$, and so inequality (3) follows immediately. ■

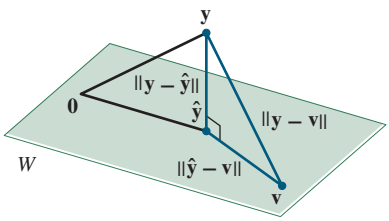


FIGURE 4 The orthogonal projection of $\mathbf{y}$ onto W is the closest point in W to $\mathbf{y}$.

EXAMPLE 3 If $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, and $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, as in Example 2, then the closest point in W to $\mathbf{y}$ is

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix}$$

EXAMPLE 4 The distance from a point $\mathbf{y}$ in $\mathbb{R}^n$ to a subspace W is defined as the distance from $\mathbf{y}$ to the nearest point in W. Find the distance from $\mathbf{y}$ to $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, where

$$\mathbf{y} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}, \quad \mathbf{u}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$$

SOLUTION By the Best Approximation Theorem, the distance from $\mathbf{y}$ to W is $\|\mathbf{y} - \hat{\mathbf{y}}\|$, where $\hat{\mathbf{y}} = \text{proj}_W \mathbf{y}$. Since $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for W,

$$\hat{\mathbf{y}} = \frac{15}{30}\mathbf{u}_1 + \frac{-21}{6}\mathbf{u}_2 = \frac{1}{2}\begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix} - \frac{7}{2}\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix}$$

$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix} - \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 6 \end{bmatrix}$$

$$\|\mathbf{y} - \hat{\mathbf{y}}\|^2 = 3^2 + 6^2 = 45$$

The distance from $\mathbf{y}$ to W is $\sqrt{45} = 3\sqrt{5}$.

The final theorem in this section shows how formula (2) for $\text{proj}_W \mathbf{y}$ is simplified when the basis for W is an orthonormal set.

THEOREM 10

If $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then

$$\text{proj}_W \mathbf{y} = (\mathbf{y}\cdot\mathbf{u}_1)\mathbf{u}_1 + (\mathbf{y}\cdot\mathbf{u}_2)\mathbf{u}_2 + \cdots + (\mathbf{y}\cdot\mathbf{u}_p)\mathbf{u}_p \tag{4}$$

If $U = [\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_p\,]$, then

$$\text{proj}_W \mathbf{y} = UU^T\mathbf{y} \quad \text{for all } \mathbf{y} \text{ in } \mathbb{R}^n \tag{5}$$

PROOF Formula (4) follows immediately from (2) in Theorem 8. Also, (4) shows that $\text{proj}_W \mathbf{y}$ is a linear combination of the columns of U using the weights $\mathbf{y}\cdot\mathbf{u}_1$, $\mathbf{y}\cdot\mathbf{u}_2, \ldots, \mathbf{y}\cdot\mathbf{u}_p$. The weights can be written as $\mathbf{u}_1^T\mathbf{y}, \mathbf{u}_2^T\mathbf{y}, \ldots, \mathbf{u}_p^T\mathbf{y}$, showing that they are the entries in $U^T\mathbf{y}$ and justifying (5).

Suppose U is an $n \times p$ matrix with orthonormal columns, and let W be the column space of U. Then

$$U^TU\mathbf{x} = I_p\mathbf{x} = \mathbf{x} \quad \text{for all } \mathbf{x} \text{ in } \mathbb{R}^p \qquad \text{Theorem 6}$$

$$UU^T\mathbf{y} = \text{proj}_W \mathbf{y} \quad \text{for all } \mathbf{y} \text{ in } \mathbb{R}^n \qquad \text{Theorem 10}$$

If U is an $n \times n$ (square) matrix with orthonormal columns, then U is an *orthogonal* matrix, the column space W is all of $\mathbb{R}^n$, and $UU^T\mathbf{y} = I\mathbf{y} = \mathbf{y}$ for all $\mathbf{y}$ in $\mathbb{R}^n$.

Although formula (4) is important for theoretical purposes, in practice it usually involves calculations with square roots of numbers (in the entries of the $\mathbf{u}_i$). Formula (2) is recommended for hand calculations.

Example 9 of Section 2.1 illustrates how matrix multiplication and transposition are used to detect a specified pattern illustrated using blue and white squares. Now that we have more experience working with bases for W and $W^{\perp}$, we are ready to discuss how to set up the matrix M in Figure 6. Let $\mathbf{w}$ be the vector generated from a pattern of blue and white squares by turning each blue square into a 1 and each white square into a 0, and then lining up each column below the column before it. See Figure 5.

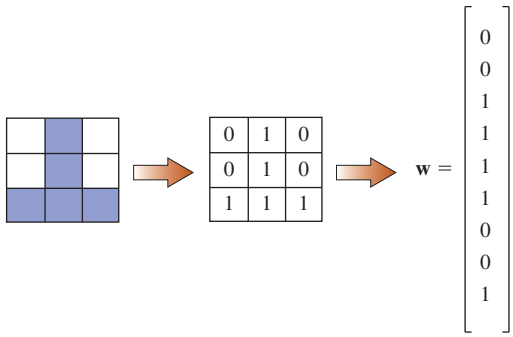


FIGURE 5 Creating a vector from colored squares.

Let $W = \text{span}\{\mathbf{w}\}$. Choose a basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{n-1}\}$ for $W^{\perp}$. Create the matrix $B = \begin{bmatrix} \mathbf{v}_1{}^T \\ \mathbf{v}_2{}^T \\ \vdots \\ \mathbf{v}_{n-1}{}^T \end{bmatrix}$. Notice $B\mathbf{u} = \mathbf{0}$ if and only if $\mathbf{u}$ is orthogonal to a set of basis vectors for $W^{\perp}$, which happens if and only if $\mathbf{u}$ is in W. Set $M = B^T B$. Then $\mathbf{u}^T M\mathbf{u} = \mathbf{u}^T B^T B\mathbf{u} = (B\mathbf{u})^T B\mathbf{u}$. By Theorem 1, $(B\mathbf{u})^T B\mathbf{u} = 0$ if and only if $B\mathbf{u} = 0$, and hence $\mathbf{u}^T M\mathbf{u} = 0$ if and only if $\mathbf{u} \in W$. But there are only two vectors in W consisting of zeros and ones: $1\mathbf{w} = \mathbf{w}$ and $0\mathbf{w} = \mathbf{0}$. Thus we can conclude that if $\mathbf{u}^T M\mathbf{u} = 0$, but $\mathbf{u}^T\mathbf{u} \neq 0$, then $\mathbf{u} = \mathbf{w}$. See Figure 6.

EXAMPLE 5 Find a matrix M that can be used in Figure 6 to identify the perp symbol.

SOLUTION First change the symbol into a vector. Set $\mathbf{w} = [0\ 0\ 1\ 1\ 1\ 1\ 0\ 0\ 1]^T$. Next set $W = \text{span}\{\mathbf{w}\}$ and find a basis for $W^{\perp}$: solving $\mathbf{x}^T\mathbf{w} = 0$ creates the homogeneous system of equations:

$$x_3 + x_4 + x_5 + x_6 + x_9 = 0$$

Treating x_3 as the basic variable and the remaining variables as free variables we get a basis for $W^{\perp}$. Transposing each vector in the basis and inserting it as a row of B we get

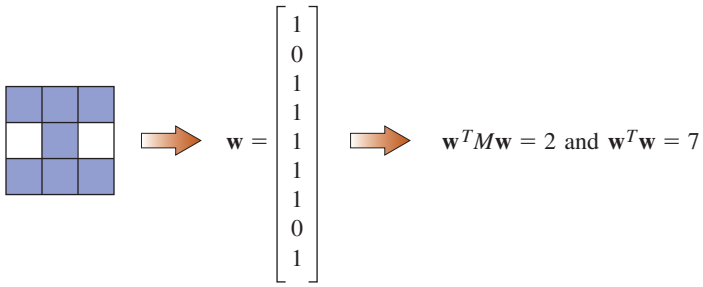


FIGURE 6 How AI detects the perp symbol.

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ and } M = B^T B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & -1 & -1 & -1 & 0 & 0 & -1 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Notice $\mathbf{w}^T M\mathbf{w} = 0$, but $\mathbf{w}^T\mathbf{w} \neq 0$. ■

Practice Problems

1. Let $\mathbf{u}_1 = \begin{bmatrix} -7 \\ 1 \\ 4 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -1 \\ 1 \\ -2 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} -9 \\ 1 \\ 6 \end{bmatrix}$, and $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Use the fact that $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal to compute $\text{proj}_W \mathbf{y}$.

2. Let W be a subspace of $\mathbb{R}^n$. Let $\mathbf{x}$ and $\mathbf{y}$ be vectors in $\mathbb{R}^n$ and let $\mathbf{z} = \mathbf{x} + \mathbf{y}$. If $\mathbf{u}$ is the projection of $\mathbf{x}$ onto W and $\mathbf{v}$ is the projection of $\mathbf{y}$ onto W, show that $\mathbf{u} + \mathbf{v}$ is the projection of $\mathbf{z}$ onto W.

6.3 Exercises

In Exercises 1 and 2, you may assume that $\{\mathbf{u}_1, \ldots, \mathbf{u}_4\}$ is an orthogonal basis for $\mathbb{R}^4$.

1. $\mathbf{u}_1 = \begin{bmatrix} 0 \\ 1 \\ -4 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} 3 \\ 5 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ -4 \end{bmatrix}$, $\mathbf{u}_4 = \begin{bmatrix} 5 \\ -3 \\ -1 \\ 1 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 10 \\ -8 \\ 2 \\ 0 \end{bmatrix}$. Write $\mathbf{x}$ as the sum of two vectors, one in Span $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ and the other in Span $\{\mathbf{u}_4\}$.

2. $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ -1 \\ 1 \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} 1 \\ 1 \\ -2 \\ -1 \end{bmatrix}$, $\mathbf{u}_4 = \begin{bmatrix} -1 \\ 1 \\ 1 \\ -2 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} 4 \\ 5 \\ -2 \\ 2 \end{bmatrix}$. Write $\mathbf{v}$ as the sum of two vectors, one in Span $\{\mathbf{u}_1\}$ and the other in Span $\{\mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$.

In Exercises 3–6, verify that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal set, and then find the orthogonal projection of $\mathbf{y}$ onto Span $\{\mathbf{u}_1, \mathbf{u}_2\}$.

3. $\mathbf{y} = \begin{bmatrix} -1 \\ 4 \\ 3 \end{bmatrix}, \mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$

4. $\mathbf{y} = \begin{bmatrix} 4 \\ 3 \\ -2 \end{bmatrix}, \mathbf{u}_1 = \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -4 \\ 3 \\ 0 \end{bmatrix}$

5. $\mathbf{y} = \begin{bmatrix} -1 \\ 2 \\ 6 \end{bmatrix}, \mathbf{u}_1 = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}$

6. $\mathbf{y} = \begin{bmatrix} -1 \\ 5 \\ 3 \end{bmatrix}, \mathbf{u}_1 = \begin{bmatrix} 4 \\ -1 \\ 1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 1 \\ -1 \\ -5 \end{bmatrix}$

In Exercises 7–10, let W be the subspace spanned by the $\mathbf{u}$'s, and write $\mathbf{y}$ as the sum of a vector in W and a vector orthogonal to W.

7. $\mathbf{y} = \begin{bmatrix} 1 \\ 3 \\ 5 \end{bmatrix}, \mathbf{u}_1 = \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 5 \\ 1 \\ 4 \end{bmatrix}$

8. $\mathbf{y} = \begin{bmatrix} -1 \\ 6 \\ 4 \end{bmatrix}, \mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -1 \\ 4 \\ -3 \end{bmatrix}$

9. $\mathbf{y} = \begin{bmatrix} 4 \\ 3 \\ 3 \\ -1 \end{bmatrix}, \mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -1 \\ 3 \\ 1 \\ -2 \end{bmatrix}, \mathbf{u}_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 1 \end{bmatrix}$

10. $\mathbf{y} = \begin{bmatrix} 3 \\ 4 \\ 5 \\ 4 \end{bmatrix}, \mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \mathbf{u}_3 = \begin{bmatrix} 0 \\ -1 \\ 1 \\ -1 \end{bmatrix}$

In Exercises 11 and 12, find the closest point to $\mathbf{y}$ in the subspace W spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$.

11. $\mathbf{y} = \begin{bmatrix} 3 \\ 1 \\ 5 \\ 1 \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} 3 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}$

12. $\mathbf{y} = \begin{bmatrix} 4 \\ 3 \\ 4 \\ 7 \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \\ -2 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \end{bmatrix}$

In Exercises 13 and 14, find the best approximation to $\mathbf{z}$ by vectors of the form $c_1\mathbf{v}_1 + c_2\mathbf{v}_2$.

13. $\mathbf{z} = \begin{bmatrix} 3 \\ -7 \\ 2 \\ 3 \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} 2 \\ -1 \\ -3 \\ 1 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix}$

14. $\mathbf{z} = \begin{bmatrix} 2 \\ 4 \\ 0 \\ -1 \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} 2 \\ 0 \\ -1 \\ -3 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 5 \\ -2 \\ 4 \\ 2 \end{bmatrix}$

15. Let $\mathbf{y} = \begin{bmatrix} 5 \\ -9 \\ 5 \end{bmatrix}, \mathbf{u}_1 = \begin{bmatrix} -3 \\ -5 \\ 1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}$. Find the distance from $\mathbf{y}$ to the plane in $\mathbb{R}^3$ spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$.

16. Let $\mathbf{y}$, $\mathbf{v}_1$, and $\mathbf{v}_2$ be as in Exercise 12. Find the distance from $\mathbf{y}$ to the subspace of $\mathbb{R}^4$ spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$.

17. Let $\mathbf{y} = \begin{bmatrix} 4 \\ 8 \\ 1 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 2/3 \\ 1/3 \\ 2/3 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2/3 \\ 2/3 \\ 1/3 \end{bmatrix}$, and $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$.

a. Let $U = [\,\mathbf{u}_1 \quad \mathbf{u}_2\,]$. Compute U^TU and UU^T.

b. Compute $\text{proj}_W \mathbf{y}$ and $(UU^T)\mathbf{y}$.

18. Let $\mathbf{y} = \begin{bmatrix} 7 \\ 9 \end{bmatrix}, \mathbf{u}_1 = \begin{bmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \end{bmatrix}$, and $W = \text{Span}\{\mathbf{u}_1\}$.

a. Let U be the 2×1 matrix whose only column is $\mathbf{u}_1$. Compute U^TU and UU^T.

b. Compute $\text{proj}_W \mathbf{y}$ and $(UU^T)\mathbf{y}$.

19. Let $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$, and $\mathbf{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Note that $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal but that $\mathbf{u}_3$ is not orthogonal to $\mathbf{u}_1$ or $\mathbf{u}_2$. It can be shown that $\mathbf{u}_3$ is not in the subspace W spanned

by $\mathbf{u}_1$ and $\mathbf{u}_2$. Use this fact to construct a nonzero vector $\mathbf{v}$ in $\mathbb{R}^3$ that is orthogonal to $\mathbf{u}_1$ and $\mathbf{u}_2$.

20. Let $\mathbf{u}_1$ and $\mathbf{u}_2$ be as in Exercise 19, and let $\mathbf{u}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. It can be shown that $\mathbf{u}_4$ is not in the subspace W spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$. Use this fact to construct a nonzero vector $\mathbf{v}$ in $\mathbb{R}^3$ that is orthogonal to $\mathbf{u}_1$ and $\mathbf{u}_2$.

In Exercises 21–30, all vectors and subspaces are in $\mathbb{R}^n$. Mark each statement True or False **(T/F)**. Justify each answer.

21. (T/F) If $\mathbf{z}$ is orthogonal to $\mathbf{u}_1$ and to $\mathbf{u}_2$ and if $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, then $\mathbf{z}$ must be in $W^\perp$.

22. (T/F) For each $\mathbf{y}$ and each subspace W, the vector $\mathbf{y} - \text{proj}_W \mathbf{y}$ is orthogonal to W.

23. (T/F) The orthogonal projection $\hat{\mathbf{y}}$ of $\mathbf{y}$ onto a subspace W can sometimes depend on the orthogonal basis for W used to compute $\hat{\mathbf{y}}$.

24. (T/F) If $\mathbf{y}$ is in a subspace W, then the orthogonal projection of $\mathbf{y}$ onto W is $\mathbf{y}$ itself.

25. (T/F) The best approximation to $\mathbf{y}$ by elements of a subspace W is given by the vector $\mathbf{y} - \text{proj}_W \mathbf{y}$.

26. (T/F) If W is a subspace of $\mathbb{R}^n$ and if $\mathbf{v}$ is in both W and $W^\perp$, then $\mathbf{v}$ must be the zero vector.

27. (T/F) In the Orthogonal Decomposition Theorem, each term in formula (2) for $\hat{\mathbf{y}}$ is itself an orthogonal projection of $\mathbf{y}$ onto a subspace of W.

28. (T/F) If $\mathbf{y} = \mathbf{z}_1 + \mathbf{z}_2$, where $\mathbf{z}_1$ is in a subspace W and $\mathbf{z}_2$ is in $W^\perp$, then $\mathbf{z}_1$ must be the orthogonal projection of $\mathbf{y}$ onto W.

29. (T/F) If the columns of an $n \times p$ matrix U are orthonormal, then $UU^T\mathbf{y}$ is the orthogonal projection of $\mathbf{y}$ onto the column space of U.

30. (T/F) If an $n \times p$ matrix U has orthonormal columns, then $UU^T\mathbf{x} = \mathbf{x}$ for all $\mathbf{x}$ in $\mathbb{R}^n$.

31. Let A be an $m \times n$ matrix. Prove that every vector $\mathbf{x}$ in $\mathbb{R}^n$ can be written in the form $\mathbf{x} = \mathbf{p} + \mathbf{u}$, where $\mathbf{p}$ is in Row A and $\mathbf{u}$ is in Nul A. Also, show that if the equation $A\mathbf{x} = \mathbf{b}$ is consistent, then there is a unique $\mathbf{p}$ in Row A such that $A\mathbf{p} = \mathbf{b}$.

32. Let W be a subspace of $\mathbb{R}^n$ with an orthogonal basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_p\}$, and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_q\}$ be an orthogonal basis for $W^\perp$.

a. Explain why $\{\mathbf{w}_1, \ldots, \mathbf{w}_p, \mathbf{v}_1, \ldots, \mathbf{v}_q\}$ is an orthogonal set.

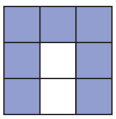
b. Explain why the set in part (a) spans $\mathbb{R}^n$.

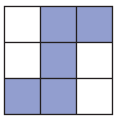
c. Show that $\dim W + \dim W^\perp = n$.

In Exercises 33–36, first change the given pattern into a vector $\mathbf{w}$ of zeros and ones and then use the method illustrated in Example 5 to find a matrix M so that $\mathbf{w}^T M \mathbf{w} = 0$, but $\mathbf{u}^T M \mathbf{u} \neq 0$ for all other nonzero vectors $\mathbf{u}$ of zeros and ones.

33. 

34. 

35. 

36. 

T 37. Let U be the 8×4 matrix in Exercise 43 in Section 6.2. Find the closest point to $\mathbf{y} = (1, 1, 1, 1, 1, 1, 1, 1)$ in Col U. Write the keystrokes or commands you use to solve this problem.

T 38. Let U be the matrix in Exercise 37. Find the distance from $\mathbf{b} = (1, 1, 1, 1, -1, -1, -1, -1)$ to Col U.

Solution to Practice Problems

1. Compute

$$\text{proj}_W \mathbf{y} = \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 = \frac{88}{66}\mathbf{u}_1 + \frac{-2}{6}\mathbf{u}_2$$

$$= \frac{4}{3}\begin{bmatrix} -7 \\ 1 \\ 4 \end{bmatrix} - \frac{1}{3}\begin{bmatrix} -1 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} -9 \\ 1 \\ 6 \end{bmatrix} = \mathbf{y}$$

In this case, $\mathbf{y}$ happens to be a linear combination of $\mathbf{u}_1$ and $\mathbf{u}_2$, so $\mathbf{y}$ is in W. The closest point in W to $\mathbf{y}$ is $\mathbf{y}$ itself.

2. Using Theorem 10, let U be a matrix whose columns consist of an orthonormal basis for W. Then $\text{proj}_W \mathbf{z} = UU^T\mathbf{z} = UU^T(\mathbf{x} + \mathbf{y}) = UU^T\mathbf{x} + UU^T\mathbf{y} = \text{proj}_W \mathbf{x} + \text{proj}_W \mathbf{y} = \mathbf{u} + \mathbf{v}$.

6.6 Machine Learning and Linear Models

Machine Learning

Machine learning uses linear models in situations where the machine is being *trained* to predict the outcome (dependent variables) based on the values of the inputs (independent variables). The machine is given a set of training data where the values of the independent and dependent variables are known. The machine then *learns* the relationship between the independent variables and the dependent variables. One type of learning is to fit a curve, such as a least-squares line or parabola, to the data. Once the machine has learned the pattern from the training data, it can then estimate the value of the output based on a given value for the input.

Least-Squares Lines

A common task in science and engineering is to analyze and understand relationships among several quantities that vary. This section describes a variety of situations in which data are used to build or verify a formula that predicts the value of one variable as a function of other variables. In each case, the problem will amount to solving a least-squares problem.

For easy application of the discussion to real problems that you may encounter later in your career, we choose notation that is commonly used in the statistical analysis of scientific and engineering data. Instead of $A\mathbf{x} = \mathbf{b}$, we write $X\boldsymbol{\beta} = \mathbf{y}$ and refer to X as the **design matrix**, $\boldsymbol{\beta}$ as the **parameter vector**, and $\mathbf{y}$ as the **observation vector**.

The simplest relation between two variables x and y is the linear equation $y = \beta_0 + \beta_1 x$.[1] Experimental data often produce points $(x_1, y_1), \ldots, (x_n, y_n)$ that, when graphed, seem to lie close to a line. We want to determine the parameters β_0 and β_1 that make the line as "close" to the points as possible.

Suppose β_0 and β_1 are fixed, and consider the line $y = \beta_0 + \beta_1 x$ in Figure 1. Corresponding to each data point (x_j, y_j) there is a point $(x_j, \beta_0 + \beta_1 x_j)$ on the line with the same x-coordinate. We call y_j the *observed* value of y and $\beta_0 + \beta_1 x_j$ the *predicted* y-value (determined by the line). The difference between an observed y-value and a predicted y-value is called a *residual*.

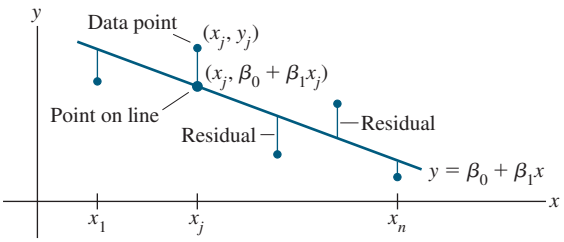


FIGURE 1 Fitting a line to experimental data.

There are several ways to measure how "close" the line is to the data. The usual choice (primarily because the mathematical calculations are simple) is to add the squares of the residuals. The **least-squares line** is the line $y = \beta_0 + \beta_1 x$ that minimizes the sum of the squares of the residuals. This line is also called a **line of regression of y on x**, because any errors in the data are assumed to be only in the y-coordinates. The coefficients β_0, β_1 of the line are called (linear) **regression coefficients**.[2]

If the data points were on the line, the parameters β_0 and β_1 would satisfy the equations

Predicted y-value		**Observed y-value**
$\beta_0 + \beta_1 x_1$	=	y_1
$\beta_0 + \beta_1 x_2$	=	y_2
$\vdots$		$\vdots$
$\beta_0 + \beta_1 x_n$	=	y_n

[1] This notation is commonly used for least-squares lines instead of $y = mx + b$.

[2] If the measurement errors are in x instead of y, simply interchange the coordinates of the data (x_j, y_j) before plotting the points and computing the regression line. If both coordinates are subject to possible error, then you might choose the line that minimizes the sum of the squares of the *orthogonal* (perpendicular) distances from the points to the line.

We can write this system as

$$X\boldsymbol{\beta} = \mathbf{y}, \quad \text{where } X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{1}$$

Of course, if the data points don't lie on a line, then there are no parameters β_0, β_1 for which the predicted y-values in $X\boldsymbol{\beta}$ equal the observed y-values in $\mathbf{y}$, and $X\boldsymbol{\beta} = \mathbf{y}$ has no solution. This is a least-squares problem, $A\mathbf{x} = \mathbf{b}$, with different notation!

The square of the distance between the vectors $X\boldsymbol{\beta}$ and $\mathbf{y}$ is precisely the sum of the squares of the residuals. The $\boldsymbol{\beta}$ that minimizes this sum also minimizes the distance between $X\boldsymbol{\beta}$ and $\mathbf{y}$. *Computing the least-squares solution of $X\boldsymbol{\beta} = \mathbf{y}$ is equivalent to finding the $\boldsymbol{\beta}$ that determines the least-squares line in Figure 1.*

EXAMPLE 1 Find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the data points $(2, 1)$, $(5, 2)$, $(7, 3)$, and $(8, 3)$.

SOLUTION Use the x-coordinates of the data to build the design matrix X in (1) and the y-coordinates to build the observation vector $\mathbf{y}$:

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$$

For the least-squares solution of $X\boldsymbol{\beta} = \mathbf{y}$, obtain the normal equations (with the new notation):

$$X^T X\boldsymbol{\beta} = X^T\mathbf{y}$$

That is, compute

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

$$X^T\mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

The normal equations are

$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

Hence

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}^{-1} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 142 & -22 \\ -22 & 4 \end{bmatrix} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 24 \\ 30 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}$$

Thus the least-squares line has the equation

$$y = \frac{2}{7} + \frac{5}{14}x$$

See Figure 2. ■

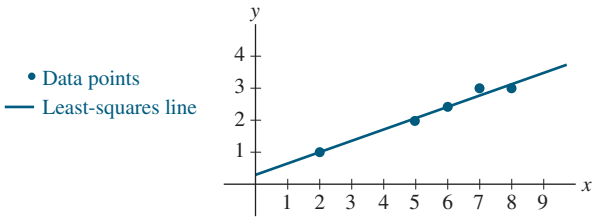


FIGURE 2 The least-squares line $y = \frac{2}{7} + \frac{5}{14}x$.

EXAMPLE 2 If a machine learns the data from Example 1 by creating a least-squares line, what outcome will it predict for the inputs 4 and 6?

SOLUTION The machine would perform the same calculations as in Example 1 to arrive at the least-squares line

$$y = \frac{2}{7} + \frac{5}{14}x$$

as a reasonable pattern to use to predict the outcomes.

For the value $x = 4$, the machine will predict an output of $y = \frac{2}{7} + \frac{5}{14}(4) = \frac{12}{7}$.

For the value $x = 6$, the machine will predict an output of $y = \frac{2}{7} + \frac{5}{14}(6) = \frac{17}{7}$.

See Figure 3.

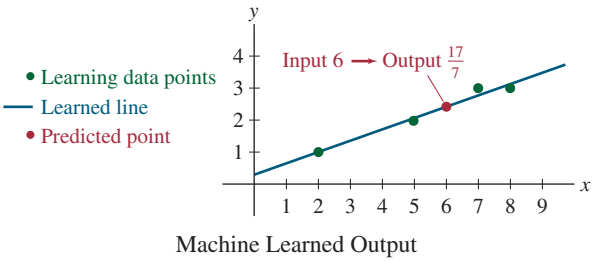


FIGURE 3 Machine-learned output.

A common practice before computing a least-squares line is to compute the average $\overline{x}$ of the original x-values and form a new variable $x^* = x - \overline{x}$. The new x-data are said to be in **mean-deviation form**. In this case, the two columns of the design matrix will be orthogonal. Solution of the normal equations is simplified, just as in Example 4 in Section 6.5. See Exercises 23 and 24.

The General Linear Model

In some applications, it is necessary to fit data points with something other than a straight line. In the examples that follow, the matrix equation is still $X\boldsymbol{\beta} = \mathbf{y}$, but the specific form of X changes from one problem to the next. Statisticians usually introduce a **residual vector** $\boldsymbol{\epsilon}$, defined by $\boldsymbol{\epsilon} = \mathbf{y} - X\boldsymbol{\beta}$, and write

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Any equation of this form is referred to as a **linear model**. Once X and $\mathbf{y}$ are determined, the goal is to minimize the length of $\boldsymbol{\epsilon}$, which amounts to finding a least-squares solution

of $X\boldsymbol{\beta} = \mathbf{y}$. In each case, the least-squares solution $\hat{\boldsymbol{\beta}}$ is a solution of the normal equations

$$X^T X\boldsymbol{\beta} = X^T\mathbf{y}$$

Least-Squares Fitting of Other Curves

When data points $(x_1, y_1), \ldots, (x_n, y_n)$ on a scatter plot do not lie close to any line, it may be appropriate to postulate some other functional relationship between x and y.

The next two examples show how to fit data by curves that have the general form

$$y = \beta_0 f_0(x) + \beta_1 f_1(x) + \cdots + \beta_k f_k(x) \tag{2}$$

where $f_0, \ldots, f_k$ are known functions and $\beta_0, \ldots, \beta_k$ are parameters that must be determined. As we will see, equation (2) describes a linear model because it is linear in the unknown parameters.

For a particular value of x, (2) gives a predicted, or "fitted," value of y. The difference between the observed value and the predicted value is the residual. The parameters $\beta_0, \ldots, \beta_k$ must be determined so as to minimize the sum of the squares of the residuals.

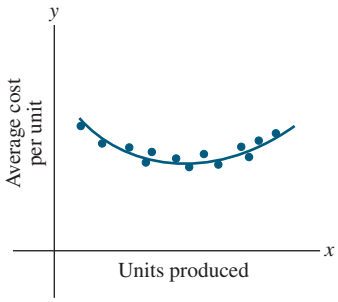


FIGURE 4
Average cost curve.

EXAMPLE 3 Suppose data points $(x_1, y_1), \ldots, (x_n, y_n)$ appear to lie along some sort of parabola instead of a straight line. For instance, if the x-coordinate denotes the production level for a company, and y denotes the average cost per unit of operating at a level of x units per day, then a typical average cost curve looks like a parabola that opens upward (Figure 4). In ecology, a parabolic curve that opens downward is used to model the net primary production of nutrients in a plant, as a function of the surface area of the foliage (Figure 5). Suppose we wish to approximate the data by an equation of the form

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 \tag{3}$$

Describe the linear model that produces a "least-squares fit" of the data by equation (3).

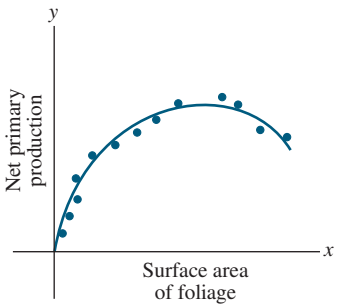


FIGURE 5
Production of nutrients.

SOLUTION Equation (3) describes the ideal relationship. Suppose the actual values of the parameters are $\beta_0, \beta_1, \beta_2$. Then the coordinates of the first data point (x_1, y_1) satisfy an equation of the form

$$y_1 = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \epsilon_1$$

where ϵ_1 is the residual error between the observed value y_1 and the predicted y-value $\beta_0 + \beta_1 x_1 + \beta_2 x_1^2$. Each data point determines a similar equation:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \epsilon_1 \\ y_2 &= \beta_0 + \beta_1 x_2 + \beta_2 x_2^2 + \epsilon_2 \\ &\vdots \qquad\qquad\quad \vdots \\ y_n &= \beta_0 + \beta_1 x_n + \beta_2 x_n^2 + \epsilon_n \end{aligned}$$

It is a simple matter to write this system of equations in the form $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$. To find X, inspect the first few rows of the system and look for the pattern.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$
$$\mathbf{y} \quad = \quad X \quad \boldsymbol{\beta} \quad + \quad \boldsymbol{\epsilon}$$

■

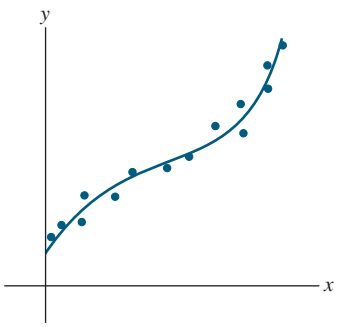


FIGURE 6
Data points along a cubic curve.

EXAMPLE 4 If data points tend to follow a pattern such as in Figure 6, then an appropriate model might be an equation of the form

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

Such data, for instance, could come from a company's total costs, as a function of the level of production. Describe the linear model that gives a least-squares fit of this type to data $(x_1, y_1), \ldots, (x_n, y_n)$.

SOLUTION By an analysis similar to that in Example 2, we obtain

Observation vector: $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, Design matrix: $X = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix}$, Parameter vector: $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$, Residual vector: $\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$

■

Multiple Regression

Suppose an experiment involves two independent variables—say, u and v—and one dependent variable, y. A simple equation for predicting y from u and v has the form

$$y = \beta_0 + \beta_1 u + \beta_2 v \tag{4}$$

A more general prediction equation might have the form

$$y = \beta_0 + \beta_1 u + \beta_2 v + \beta_3 u^2 + \beta_4 uv + \beta_5 v^2 \tag{5}$$

This equation is used in geology, for instance, to model erosion surfaces, glacial cirques, soil pH, and other quantities. In such cases, the least-squares fit is called a *trend surface*.

Equations (4) and (5) both lead to a linear model because they are linear in the unknown parameters (even though u and v are multiplied). In general, a linear model will arise whenever y is to be predicted by an equation of the form

$$y = \beta_0 f_0(u, v) + \beta_1 f_1(u, v) + \cdots + \beta_k f_k(u, v)$$

with $f_0, \ldots, f_k$ any sort of known functions and $\beta_0, \ldots, \beta_k$ unknown weights.

EXAMPLE 5 In geography, local models of terrain are constructed from data $(u_1, v_1, y_1), \ldots, (u_n, v_n, y_n)$, where u_j, v_j, and y_j are latitude, longitude, and altitude, respectively. Describe the linear model based on (4) that gives a least-squares fit to such data. The solution is called the *least-squares plane*. See Figure 7.

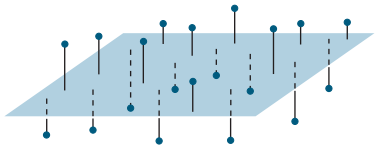


FIGURE 7 A least-squares plane.

SOLUTION We expect the data to satisfy the following equations:

$$\begin{aligned} y_1 &= \beta_0 + \beta_1 u_1 + \beta_2 v_1 + \epsilon_1 \\ y_2 &= \beta_0 + \beta_1 u_2 + \beta_2 v_2 + \epsilon_2 \\ &\vdots \qquad\qquad\quad \vdots \\ y_n &= \beta_0 + \beta_1 u_n + \beta_2 v_n + \epsilon_n \end{aligned}$$

This system has the matrix form $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where

$$\underset{\text{Observation vector}}{\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}, \quad \underset{\text{Design matrix}}{X = \begin{bmatrix} 1 & u_1 & v_1 \\ 1 & u_2 & v_2 \\ \vdots & \vdots & \vdots \\ 1 & u_n & v_n \end{bmatrix}}, \quad \underset{\text{Parameter vector}}{\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}}, \quad \underset{\text{Residual vector}}{\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}}$$

Example 5 shows that the linear model for multiple regression has the same abstract form as the model for the simple regression in the earlier examples. Linear algebra gives us the power to understand the general principle behind all the linear models. Once X is defined properly, the normal equations for $\boldsymbol{\beta}$ have the same matrix form, no matter how many variables are involved. Thus, for any linear model where X^TX is invertible, the least-squares $\hat{\boldsymbol{\beta}}$ is given by $(X^TX)^{-1}X^T\mathbf{y}$.

STUDY GUIDE offers additional resources for understanding the geometry of a linear model.

Practice Problem

When the monthly sales of a product are subject to seasonal fluctuations, a curve that approximates the sales data might have the form

$$y = \beta_0 + \beta_1 x + \beta_2 \sin(2\pi x/12)$$

where x is the time in months. The term $\beta_0 + \beta_1 x$ gives the basic sales trend, and the sine term reflects the seasonal changes in sales. Give the design matrix and the parameter vector for the linear model that leads to a least-squares fit of the equation above. Assume the data are $(x_1, y_1), \ldots, (x_n, y_n)$.

6.6 Exercises

In Exercises 1–4, find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the given data points.

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

2. $(1, 0), (2, 2), (3, 7), (4, 9)$

3. $(-1, 0), (0, 1), (1, 2), (2, 4)$

4. $(2, 3), (3, 2), (5, 1), (6, 0)$

5. If a machine learns the least-squares line that best fits the data in Exercise 1, what will the machine pick for the value of y when $x = 4$?

6. If a machine learns the least-squares line that best fits the data in Exercise 2, what will the machine pick for the value of y when $x = 3$?

7. If a machine learns the least-squares line that best fits the data in Exercise 1, what will the machine pick for the value of y when $x = 3$? How closely does this match the data point at $x = 3$ fed into the machine?

8. If a machine learns the least-squares line that best fits the data in Exercise 2, what will the machine pick for the value of y when $x = 4$? How closely does this match the data point at $x = 4$ fed into the machine?

9. If you enter the data from Exercise 1 into a machine and it returns a y value of 20 when $x = 2.5$, should you trust the machine? Justify your answer.

10. If you enter the data from Exercise 2 into a machine and it returns a y value of -1 when $x = 1.5$, should you trust the machine? Justify your answer.

11. Let X be the design matrix used to find the least-squares line to fit data $(x_1, y_1), \ldots, (x_n, y_n)$. Use a theorem in Section 6.5 to show that the normal equations have a unique solution if and only if the data include at least two data points with different x-coordinates.

12. Let X be the design matrix in Example 2 corresponding to a least-squares fit of a parabola to data $(x_1, y_1), \ldots, (x_n, y_n)$. Suppose x_1, x_2, and x_3 are distinct. Explain why there is only one parabola that fits the data best, in a least-squares sense. (See Exercise 11.)

13. A certain experiment produces the data $(1, 2.5)$, $(2, 4.3)$, $(3, 5.5)$, $(4, 6.1)$, $(5, 6.1)$. Describe the model that produces a least-squares fit of these points by a function of the form

$$y = \beta_1 x + \beta_2 x^2$$

Such a function might arise, for example, as the revenue from the sale of x units of a product, when the amount offered for sale affects the price to be set for the product.

a. Give the design matrix, the observation vector, and the unknown parameter vector.

T b. Find the associated least-squares curve for the data.

c. If a machine learned the curve you found in (b), what output would it provide for an input of $x = 6$?

14. A simple curve that often makes a good model for the variable costs of a company, as a function of the sales level x, has the form $y = \beta_1 x + \beta_2 x^2 + \beta_3 x^3$. There is no constant term because fixed costs are not included.

a. Give the design matrix and the parameter vector for the linear model that leads to a least-squares fit of the equation above, with data $(x_1, y_1), \ldots, (x_n, y_n)$.

T b. Find the least-squares curve of the form above to fit the data $(4, 1.58)$, $(6, 2.08)$, $(8, 2.5)$, $(10, 2.8)$, $(12, 3.1)$, $(14, 3.4)$, $(16, 3.8)$, and $(18, 4.32)$, with values in thousands. If possible, produce a graph that shows the data points and the graph of the cubic approximation.

c. If a machine learned the curve you found in (b), what output would it provide for an input of $x = 9$?

15. A certain experiment produces the data $(1, 7.9)$, $(2, 5.4)$, and $(3, -.9)$. Describe the model that produces a least-squares fit of these points by a function of the form

$$y = A \cos x + B \sin x$$

16. Suppose radioactive substances A and B have decay constants of .02 and .07, respectively. If a mixture of these two substances at time $t = 0$ contains M_A grams of A and M_B grams of B, then a model for the total amount y of the mixture present at time t is

$$y = M_A e^{-.02t} + M_B e^{-.07t} \tag{6}$$

Suppose the initial amounts M_A and M_B are unknown, but a scientist is able to measure the total amounts present at several times and records the following points (t_i, y_i): $(10, 21.34)$, $(11, 20.68)$, $(12, 20.05)$, $(14, 18.87)$, and $(15, 18.30)$.

a. Describe a linear model that can be used to estimate M_A and M_B.

T b. Find the least-squares curve based on (6).



Halley's Comet last appeared in 1986 and will reappear in 2061.

T 17. According to Kepler's first law, a comet should have an elliptic, parabolic, or hyperbolic orbit (with gravitational attractions from the planets ignored). In suitable polar coordinates, the position (r, ϑ) of a comet satisfies an equation of the form

$$r = \beta + e(r \cdot \cos \vartheta)$$

where β is a constant and e is the *eccentricity* of the orbit, with $0 \le e < 1$ for an ellipse, $e = 1$ for a parabola, and $e > 1$ for a hyperbola. Suppose observations of a newly discovered comet provide the data below. Determine the type of orbit, and predict where the comet will be when $\vartheta = 4.6$ (radians).[3]

ϑ	.88	1.10	1.42	1.77	2.14
r	3.00	2.30	1.65	1.25	1.01

T 18. A healthy child's systolic blood pressure p (in millimeters of mercury) and weight w (in pounds) are approximately related by the equation

$$\beta_0 + \beta_1 \ln w = p$$

Use the following experimental data to estimate the systolic blood pressure of a healthy child weighing 100 pounds.

w	44	61	81	113	131
$\ln w$	3.78	4.11	4.39	4.73	4.88
p	91	98	103	110	112

T 19. To measure the takeoff performance of an airplane, the horizontal position of the plane was measured every second, from $t = 0$ to $t = 12$. The positions (in feet) were: 0, 8.8, 29.9, 62.0, 104.7, 159.1, 222.0, 294.5, 380.4, 471.1, 571.7, 686.8, and 809.2.

a. Find the least-squares cubic curve $y = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$ for these data.

b. If a machine learned the curve given in part (a), what would it estimate the velocity of the plane to be when $t = 4.5$ seconds?

20. Let $\overline{x} = \dfrac{1}{n}(x_1 + \cdots + x_n)$ and $\overline{y} = \dfrac{1}{n}(y_1 + \cdots + y_n)$. Show that the least-squares line for the data $(x_1, y_1), \ldots, (x_n, y_n)$ must pass through $(\overline{x}, \overline{y})$. That is, show that $\overline{x}$ and $\overline{y}$ satisfy the linear equation $\overline{y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{x}$. [*Hint:* Derive this equation from the vector equation $\mathbf{y} = X\hat{\boldsymbol{\beta}} + \boldsymbol{\epsilon}$. Denote the first column of X by $\mathbf{1}$. Use the fact that the residual vector $\boldsymbol{\epsilon}$ is orthogonal to the column space of X and hence is orthogonal to $\mathbf{1}$.]

[3] The basic idea of least-squares fitting of data is due to K. F. Gauss (and, independently, to A. Legendre), whose initial rise to fame occurred in 1801 when he used the method to determine the path of the asteroid *Ceres*. Forty days after the asteroid was discovered, it disappeared behind the sun. Gauss predicted it would appear ten months later and gave its location. The accuracy of the prediction astonished the European scientific community.

Given data for a least-squares problem, $(x_1, y_1), \ldots, (x_n, y_n)$, the following abbreviations are helpful:

$$\sum x = \sum_{i=1}^n x_i, \quad \sum x^2 = \sum_{i=1}^n x_i^2,$$
$$\sum y = \sum_{i=1}^n y_i, \quad \sum xy = \sum_{i=1}^n x_i y_i$$

The normal equations for a least-squares line $y = \hat{\beta}_0 + \hat{\beta}_1 x$ may be written in the form

$$\begin{aligned} n\hat{\beta}_0 + \hat{\beta}_1 \sum x &= \sum y \\ \hat{\beta}_0 \sum x + \hat{\beta}_1 \sum x^2 &= \sum xy \end{aligned} \tag{7}$$

21. Derive the normal equations (7) from the matrix form given in this section.

22. Use a matrix inverse to solve the system of equations in (7) and thereby obtain formulas for $\hat{\beta}_0$ and $\hat{\beta}_1$ that appear in many statistics texts.

23. a. Rewrite the data in Example 1 with new x-coordinates in mean deviation form. Let X be the associated design matrix. Why are the columns of X orthogonal?

b. Write the normal equations for the data in part (a), and solve them to find the least-squares line, $y = \beta_0 + \beta_1 x^*$, where $x^* = x - 5.5$.

24. Suppose the x-coordinates of the data $(x_1, y_1), \ldots, (x_n, y_n)$ are in mean deviation form, so that $\sum x_i = 0$. Show that if X is the design matrix for the least-squares line in this case, then $X^T X$ is a diagonal matrix.

Exercises 25 and 26 involve a design matrix X with two or more columns and a least-squares solution $\hat{\boldsymbol{\beta}}$ of $\mathbf{y} = X\boldsymbol{\beta}$. Consider the following numbers.

(i) $\|X\hat{\boldsymbol{\beta}}\|^2$—the sum of the squares of the "regression term." Denote this number by SS(R).

(ii) $\|\mathbf{y} - X\hat{\boldsymbol{\beta}}\|^2$—the sum of the squares for the error term. Denote this number by SS(E).

(iii) $\|\mathbf{y}\|^2$—the "total" sum of the squares of the y-values. Denote this number by SS(T).

Every statistics text that discusses regression and the linear model $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$ introduces these numbers, though terminology and notation vary somewhat. To simplify matters, assume that the mean of the y-values is zero. In this case, SS(T) is proportional to what is called the *variance* of the set of y-values.

25. Justify the equation SS(T) = SS(R) + SS(E). [*Hint:* Use a theorem, and explain why the hypotheses of the theorem are satisfied.] This equation is extremely important in statistics, both in regression theory and in the analysis of variance.

26. Show that $\|X\hat{\boldsymbol{\beta}}\|^2 = \hat{\boldsymbol{\beta}}^T X^T \mathbf{y}$. [*Hint:* Rewrite the left side and use the fact that $\hat{\boldsymbol{\beta}}$ satisfies the normal equations.] This formula for SS(R) is used in statistics. From this and from Exercise 25, obtain the standard formula for SS(E):

$$\text{SS(E)} = \mathbf{y}^T \mathbf{y} - \hat{\boldsymbol{\beta}}^T X^T \mathbf{y}$$

This page is intentionally left blank

9.2 Linear Programming—Geometric Method

Since the 1950s, the variety and size of industrial linear programming problems have grown along with the dramatic increase in computing power. Still, at their core, linear programming problems have a concise mathematical description, discussed in this section. The final example in the section presents a geometric view of linear programming that is important for visualizing the algebraic approach needed for larger problems.

Generally speaking, a linear programming problem involves a system of linear inequalities in variables $x_1, \ldots, x_n$ and a linear functional f from $\mathbb{R}^n$ into $\mathbb{R}$. The system typically has many free variables, and the problem is to find a solution $\mathbf{x}$ that maximizes or minimizes $f(\mathbf{x})$.

EXAMPLE 1 The Shady-Lane grass seed company blends two types of seed mixtures, EverGreen and QuickGreen. Each bag of EverGreen contains 3 pounds of fescue seed, 1 pound of rye seed, and 1 pound of bluegrass seed. Each bag of QuickGreen contains 2 pounds of fescue seed, 2 pounds of rye seed, and 1 pound of bluegrass seed. The company has 1,200 pounds of fescue seed, 800 pounds of rye seed, and 450 pounds of bluegrass seed available to put into its mixtures. The company makes a profit of $2 on each bag of EverGreen and $3 on each bag of QuickGreen that it produces. Set up the mathematical problem that determines the number of bags of each mixture that Shady-Lane should make in order to maximize its profit.

SOLUTION The phrase "maximize ... profit" identifies the goal or objective of the problem. The first step, then, is to create a formula for the profit. Begin by naming the quantities that can vary. Let x_1 be the number of bags of EverGreen and x_2 the number

of bags of QuickGreen that are produced. Since the profit on each bag of EverGreen is \$2 and the profit on each bag of QuickGreen is \$3, the total profit (in dollars) is

$$2x_1 + 3x_2 \qquad \text{(profit function)}$$

The next step is to write inequalities or equalities that x_1 and x_2 must satisfy, one for each of the ingredients that are in limited supply. Notice that each bag of EverGreen requires 3 pounds of fescue seed and each bag of QuickGreen requires 2 pounds of fescue seed. So the total amount of fescue seed required is $3x_1 + 2x_2$ pounds. Since only 1,200 pounds are available, x_1 and x_2 must satisfy

$$3x_1 + 2x_2 \leq 1{,}200 \qquad \text{(fescue)}$$

Similarly, EverGreen needs 1 pound of rye seed per bag, QuickGreen needs 2 pounds per bag, and only 800 pounds of rye seed are available. Thus, the total amount of rye seed required is $x_1 + 2x_2$, and x_1 and x_2 must satisfy

$$x_1 + 2x_2 \leq 800 \qquad \text{(rye)}$$

As for the bluegrass seed, EverGreen requires 1 pound per bag and QuickGreen requires 1 pound per bag. Since 450 pounds are available,

$$x_1 + x_2 \leq 450 \qquad \text{(bluegrass)}$$

Of course, x_1 and x_2 cannot be negative, so x_1 and x_2 must also satisfy

$$x_1 \geq 0 \quad \text{and} \quad x_2 \geq 0$$

The problem is summarized mathematically as

$$\begin{array}{lll} \text{Maximize} & 2x_1 + 3x_2 & \text{(profit function)} \\ \text{subject to} & 3x_1 + 2x_2 \leq 1{,}200 & \text{(fescue)} \\ & x_1 + 2x_2 \leq 800 & \text{(rye)} \\ & x_1 + x_2 \leq 450 & \text{(bluegrass)} \end{array}$$

and $x_1 \geq 0, x_2 \geq 0$. ■

EXAMPLE 2 An oil refining company has two refineries that produce three grades of unleaded gasoline. Each day refinery A produces 12,000 gallons of regular, 4,000 gallons of premium, and 1,000 gallons of super, at a cost of \$3,500. Each day refinery B produces 4,000 gallons of regular, 4,000 gallons of premium, and 5,000 gallons of super, at a cost of \$3,000. An order is received for 48,000 gallons of regular, 32,000 gallons of premium, and 20,000 gallons of super. Set up a mathematical problem that determines the number of days each refinery should operate in order to fill the order at the least cost.

SOLUTION Suppose that refinery A operates x_1 days and refinery B operates x_2 days. The cost of doing this is $3{,}500x_1 + 3{,}000x_2$ dollars. The problem is to find a production schedule (x_1, x_2) that minimizes this cost and also ensures that the required gasoline is produced.

Since refinery A produces 12,000 gallons of regular each day and refinery B produces 4,000 gallons of regular each day, the total produced is $12{,}000x_1 + 4{,}000x_2$. The total should be at least 48,000 gallons. That is,

$$12{,}000x_1 + 4{,}000x_2 \geq 48{,}000$$

Similarly, for the premium,

$$4{,}000x_1 + 4{,}000x_2 \geq 32{,}000$$

and, for the super,

$$1{,}000x_1 + 5{,}000x_2 \geq 20{,}000$$

As in Example 1, x_1 and x_2 cannot be negative, so $x_1 \geq 0$ and $x_2 \geq 0$.

The problem is summarized mathematically as

$$\begin{aligned}
&\text{Minimize} && 3{,}500x_1 + 3{,}000x_2 && \text{(cost function)}\\
&\text{subject to} && 12{,}000x_1 + 4{,}000x_2 \geq 48{,}000 && \text{(regular)}\\
&&& 4{,}000x_1 + 4{,}000x_2 \geq 32{,}000 && \text{(premium)}\\
&&& 1{,}000x_1 + 5{,}000x_2 \geq 20{,}000 && \text{(super)}
\end{aligned}$$

and $x_1 \geq 0, x_2 \geq 0$. ■

The examples show how a linear programming problem involves finding the maximum (or minimum) of a linear function, called the **objective function**, subject to certain linear constraints. In many situations, the constraints take the form of linear inequalities and the variables are restricted to nonnegative values. Here is a precise statement of the so-called canonical form of a linear programming problem.

DEFINITION

Given $\mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$ in $\mathbb{R}^m$, $\mathbf{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$ in $\mathbb{R}^n$, and an $m \times n$ matrix $A = [\,a_{ij}\,]$, the **canonical linear programming problem** is the following:

Find an n-tuple $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ in $\mathbb{R}^n$ to maximize

$$f(x_1, \ldots, x_n) = c_1x_1 + c_2x_2 + \cdots + c_nx_n$$

subject to the constraints

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

and

$$x_j \geq 0 \quad \text{for } j = 1, \ldots, n$$

This may be restated in vector-matrix notation[1] as follows:

$$\text{Maximize } f(\mathbf{x}) = \mathbf{c}^T\mathbf{x} \tag{1}$$

$$\text{subject to the constraints } A\mathbf{x} \leq \mathbf{b} \tag{2}$$

$$\text{and } \mathbf{x} \geq \mathbf{0} \tag{3}$$

where an inequality between two vectors applies to each of their coordinates.

Any vector $\mathbf{x}$ that satisfies (2) and (3) is called a **feasible solution**, and the set of all feasible solutions, denoted by $\mathcal{F}$, is called the **feasible set**. A vector $\bar{\mathbf{x}}$ in $\mathcal{F}$ is an **optimal solution** if $f(\bar{\mathbf{x}}) = \max_{\mathbf{x}\in\mathcal{F}} f(\mathbf{x})$.

[1] Some authors write $\mathbf{c}^T\mathbf{x}$ as $\mathbf{c} \cdot \mathbf{x}$, using the dot product.

The canonical statement of the problem is really not as restrictive as it might seem. To minimize a function $h(\mathbf{x})$, replace it with the problem of maximizing the function $-h(\mathbf{x})$. A constraint inequality of the sort

$$a_{i1}x_1 + \cdots + a_{in}x_n \geq b_i$$

can be replaced by

$$-a_{i1}x_1 - \cdots - a_{in}x_n \leq -b_i$$

An equality constraint

$$a_{i1}x_1 + \cdots + a_{in}x_n = b_i$$

can be replaced by two inequalities

$$a_{i1}x_1 + \cdots + a_{in}x_n \leq b_i$$
$$-a_{i1}x_1 - \cdots - a_{in}x_n \leq -b_i$$

With an arbitrary canonical linear programming problem, two things can go wrong. If the constraint inequalities are inconsistent, then $\mathcal{F}$ is the empty set. If the objective function takes on arbitrarily large values in $\mathcal{F}$, then the desired maximum does not exist. In the former case, the problem is said to be **infeasible**; in the latter case, the problem is called **unbounded**.

EXAMPLE 3 The problem

$$\begin{array}{ll} \text{Maximize} & 5x \\ \text{subject to} & x \leq 3 \\ & -x \leq -4 \\ & x \geq 0 \end{array}$$

is infeasible, since there is no x such that $x \leq 3$ and $x \geq 4$. ■

EXAMPLE 4 The problem

$$\begin{array}{ll} \text{Maximize} & 5x \\ \text{subject to} & -x \leq 3 \\ & x \geq 0 \end{array}$$

is unbounded. The values of $5x$ may be arbitrarily large, as x is only required to satisfy $x \geq 0$ (and $x \geq -3$). ■

Fortunately, these are the only two things that can go wrong.

THEOREM 6

If the feasible set $\mathcal{F}$ is nonempty and if the objective function is bounded above on $\mathcal{F}$, then the canonical linear programming problem has at least one optimal solution. Furthermore, at least one of the optimal solutions is an extreme point of $\mathcal{F}$.[2]

[2] The feasible set is the solution of a system of linear inequalities. Geometrically, this corresponds to the intersection of a finite number of (closed) half-spaces, sometimes called a polyhedral set. Intuitively, the extreme points correspond to the "corner points," or vertices, of this polyhedral set. The notion of an extreme point is discussed more fully in Section 8.5.

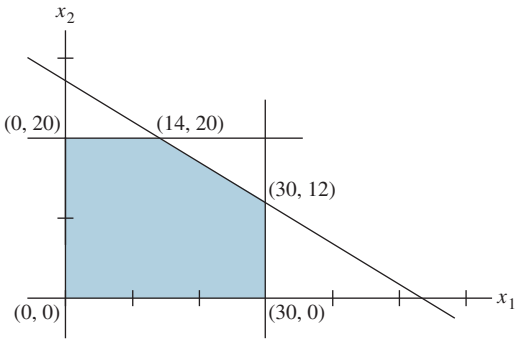
A proof of Theorem 6 is in Steven R. Lay, *Convex Sets and Their Applications* (New York: John Wiley & Sons, 1982; Mineola, NY: Dover Publications, 2007), p. 171.

Theorem 6 describes when an optimal solution exists, and it suggests a possible technique for finding one. That is, evaluate the objective function at each of the extreme points of $\mathcal{F}$ and select the point that gives the largest value. This works well in simple cases such as the next two examples. The geometric approach is limited to two or three dimensions, but it provides an important visualization of the nature of the solution set and how the objective function interacts with the feasible set to identify extreme points.

EXAMPLE 5 Maximize $f(x_1, x_2) = 2x_1 + 3x_2$

$$\begin{aligned} \text{subject to} \quad & x_1 \leq 30 \\ & x_2 \leq 20 \\ & x_1 + 2x_2 \leq 54 \\ \text{and } x_1 \geq 0, & \; x_2 \geq 0. \end{aligned}$$

SOLUTION Figure 1 shows the shaded pentagonal feasible set, obtained by graphing each of the constraint inequalities. (For simplicity, points in this section are displayed as ordered pairs or triples.) There are five extreme points, corresponding to the five vertices of the feasible set. They are found by solving the appropriate pairs of linear equations. For example, the extreme point $(14, 20)$ is found by solving the linear system $x_1 + 2x_2 = 54$ and $x_2 = 20$. The table below shows the value of the objective function at each extreme point. Evidently, the maximum is 96 at $x_1 = 30$ and $x_2 = 12$.



(x_1, x_2)	$2x_1+3x_2$
(0, 0)	0
(30, 0)	60
(30, 12)	96 ←
(14, 20)	88
(0, 20)	60

FIGURE 1

Another geometric technique that can be used when the problem involves two variables is to graph several *level lines* for the objective function. These are parallel lines, and the objective function has a constant value on each line. (See Figure 2.) The

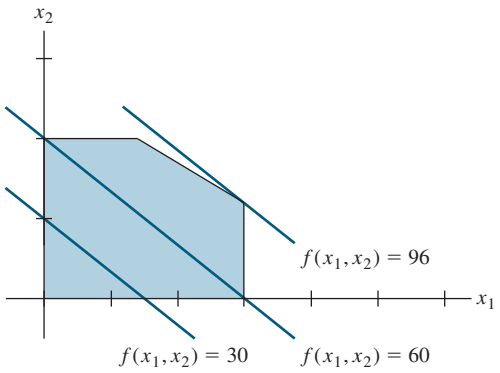


FIGURE 2

values of the objective function $f(x_1, x_2)$ increase as (x_1, x_2) moves from left to right. The level line farthest to the right that still intersects the feasible set is the line through the vertex (30, 12). Thus, the point (30, 12) yields the maximum value of $f(x_1, x_2)$ over the feasible set. ■

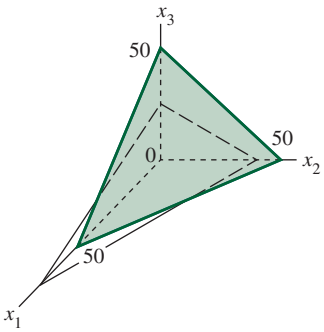


FIGURE 3

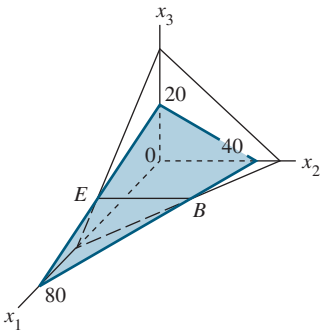


FIGURE 4

EXAMPLE 6 Maximize $f(x_1, x_2, x_3) = 2x_1 + 3x_2 + 4x_3$

$$\begin{aligned} \text{subject to} \quad & x_1 + x_2 + x_3 \le 50 \\ & x_1 + 2x_2 + 4x_3 \le 80 \\ \text{and } & x_1 \ge 0, x_2 \ge 0, x_3 \ge 0. \end{aligned}$$

SOLUTION Each of the five inequalities above determines a "half-space" in $\mathbb{R}^3$—a plane together with all points on one side of the plane. The feasible set of this linear programming problem is the intersection of these half-spaces, which is a convex set in the first octant of $\mathbb{R}^3$.

When the first inequality is changed to an equality, the graph is a plane that intercepts each coordinate axis 50 units from the origin and determines the equilateral triangular region shown in Figure 3. Since (0, 0, 0) satisfies the inequality, so do all the other points "below" the plane. In a similar fashion, the second (in)equality determines a triangular region on a plane (shown in Figure 4) that passes somewhat closer to the origin. The two planes intersect in a line that contains segment EB.

The quadrilateral surface $BCDE$ forms a boundary of the feasible set, because it is below the equilateral triangular region. Beyond EB, however, the two planes change position relative to the origin, so the planar region ABE forms another bounding surface for the feasible set. The vertices of the feasible set are the points A, B, C, D, E, and 0 (the origin). See Figure 5, which has all sides of the feasible set shaded except the large "top" piece. To find the coordinates of B, solve the system

$$\begin{cases} x_1 + x_2 + x_3 = 50 \\ x_1 + 2x_2 + 4x_3 = 80 \\ x_3 = 0 \end{cases} \Rightarrow \begin{cases} x_1 + x_2 = 50 \\ x_1 + 2x_2 = 80 \end{cases}$$

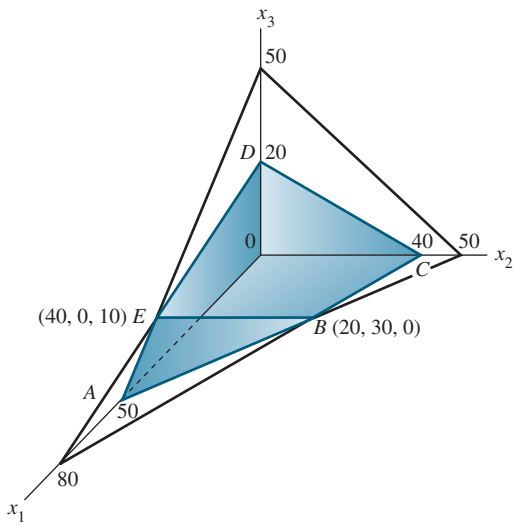


FIGURE 5

Obtain $x_2 = 30$, and find that B is $(20, 30, 0)$. For E, solve

$$\begin{cases} x_1 + x_2 + x_3 = 50 \\ x_1 + 2x_2 + 4x_3 = 80 \\ x_2 = 0 \end{cases} \Rightarrow \begin{cases} x_1 + x_3 = 50 \\ x_1 + 4x_3 = 80 \end{cases}$$

Obtain $x_3 = 10$, and find that $E = (40, 0, 10)$.

Now that the feasible set and its extreme points are clearly seen, the next step is to examine the objective function $f(x_1, x_2, x_3) = 2x_1 + 3x_2 + 4x_3$. The sets on which f is constant are planes, rather than lines, all having $(2, 3, 4)$ as a normal vector to the plane. This normal vector has a direction different from the normal vectors $(1, 1, 1)$ and $(1, 2, 4)$ to the two faces $BCDE$ and ABE. So the level sets of f are not parallel to any of the bounding surfaces of the feasible set. Figure 6 shows just the feasible set and a level set on which f has the value 120. This plane passes through C, E, and the point $(30, 20, 0)$ on the edge of the feasible set between A and B, which shows that the vertex B is "above" this level plane. In fact, $f(20, 30, 0) = 130$. Thus the unique solution of the linear programming problem is at $B = (20, 30, 0)$. ■

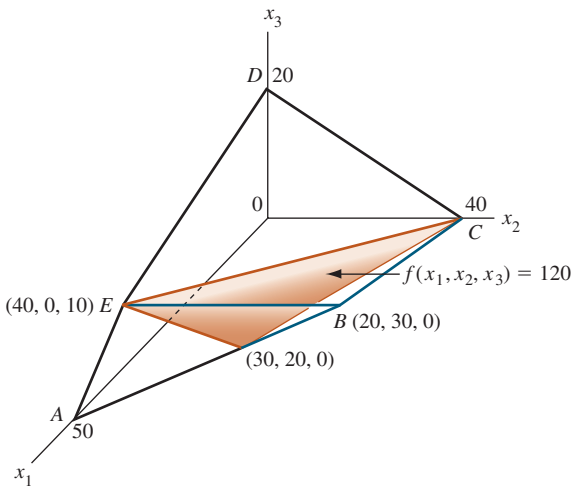


FIGURE 6

Practice Problems

1. Consider the following problem:

$$\begin{aligned} &\text{Maximize} && 2x_1 + x_2 \\ &\text{subject to} && x_1 - 2x_2 \geq -8 \\ & && 3x_1 + 2x_2 \leq 24 \\ &\text{and } x_1 \geq 0, x_2 \geq 0. \end{aligned}$$

Write this problem in the form of a canonical linear programming problem: Maximize $\mathbf{c}^T\mathbf{x}$ subject to $A\mathbf{x} \leq \mathbf{b}$ and $\mathbf{x} \geq 0$. Specify A, $\mathbf{b}$, and $\mathbf{c}$.

2. Graph the feasible set for Practice Problem 1.

3. Find the extreme points of the feasible set in Practice Problem 2.

4. Use the answer to Practice Problem 3 to find the solution to the linear programming problem in Practice Problem 1.

9.2 Exercises

1. Betty plans to invest a total of $12,000 in mutual funds, certificates of deposit (CDs), and a high-yield savings account. Because of the risk involved in mutual funds, she wants to invest no more in mutual funds than the sum of her CDs and savings. She also wants the amount in savings to be at least half the amount in CDs. Her expected returns are 11% on the mutual funds, 8% on the CDs and 6% on savings. How much money should Betty invest in each area in order to have the largest return on her investments? Set this up as a linear programming problem in the following form: Maximize $\mathbf{c}^T\mathbf{x}$ subject to $A\mathbf{x} \le \mathbf{b}$ and $\mathbf{x} \ge \mathbf{0}$. Do *not* find the solution.

2. A dog breeder decides to feed his dogs a combination of two dog foods: Pixie Power and Misty Might. He wants the dogs to receive four nutritional factors each month. The amounts of these factors (a, b, c, and d) contained in 1 bag of each dog food are shown in the following chart, together with the total amounts needed.

	a	**b**	**c**	**d**
Pixie Power	3	2	1	2
Misty Might	2	4	3	1
Needed	28	30	20	25

The costs per bag are $50 for Pixie Power and $40 for Misty Might. How many bags of each dog food should be blended to meet the nutritional requirements at the lowest cost? Set this up as a linear programming problem in the following form: Minimize $\mathbf{c}^T\mathbf{x}$ subject to $A\mathbf{x} \ge \mathbf{b}$ and $\mathbf{x} \ge \mathbf{0}$. Do *not* find the solution.

In Exercises 3–6, find vectors $\mathbf{b}$ and $\mathbf{c}$ and matrix A so that each problem is set up as a canonical linear programming problem: Maximize $\mathbf{c}^T\mathbf{x}$ subject to $A\mathbf{x} \le \mathbf{b}$ and $\mathbf{x} \ge \mathbf{0}$. Do *not* find the solution.

3. Maximize $3x_1 + 4x_2 - 2x_3$
subject to $x_1 + 2x_2 \le 20$
$3x_2 + 5x_3 \ge 10$
and $x_1 \ge 0, x_2 \ge 0, x_3 \ge 0$.

4. Maximize $3x_1 + x_2 + 5x_3$
subject to $5x_1 + 7x_2 + x_3 \le 25$
$2x_1 + 3x_2 + 4x_3 = 40$
and $x_1 \ge 0, x_2 \ge 0, x_3 \ge 0$.

5. Minimize $7x_1 - 3x_2 + x_3$
subject to $x_1 - 4x_2 \ge 35$
$x_2 - 2x_3 = 20$
and $x_1 \ge 0, x_2 \ge 0, x_3 \ge 0$.

6. Minimize $x_1 + 5x_2 - 2x_3$
subject to $2x_1 + x_2 + 4x_3 \le 27$
$x_1 - 6x_2 + 3x_3 \ge 40$
and $x_1 \ge 0, x_2 \ge 0, x_3 \ge 0$.

In Exercises 7–10, solve the linear programming problems.

7. Maximize $80x_1 + 65x_2$
subject to $2x_1 + x_2 \le 32$
$x_1 + x_2 \le 18$
$x_1 + 3x_2 \le 24$
and $x_1 \ge 0, x_2 \ge 0$.

8. Minimize $5x_1 + 3x_2$
subject to $2x_1 + 5x_2 \ge 10$
$3x_1 + x_2 \ge 6$
$x_1 + 7x_2 \ge 7$
and $x_1 \ge 0, x_2 \ge 0$.

9. Maximize $2x_1 + 7x_2$
subject to $-2x_1 + x_2 \le -4$
$x_1 - 2x_2 \le -4$
and $x_1 \ge 0, x_2 \ge 0$.

10. Maximize $5x_1 + 12x_2$
subject to $x_1 - x_2 \le 3$
$-x_1 + 2x_2 \le -4$
and $x_1 \ge 0, x_2 \ge 0$.

In Exercises 11–14, mark each statement True or False **(T/F)**. Justify each answer.

11. **(T/F)** In a canonical linear programming problem, a nonnegative vector $\mathbf{x}$ is a feasible solution if it satisfies $A\mathbf{x} \le \mathbf{b}$.

12. **(T/F)** If a canonical linear programming problem does not have an optimal solution, then either the objective function is not bounded on the feasible set $\mathcal{F}$ or $\mathcal{F}$ is the empty set.

13. **(T/F)** A vector $\overline{\mathbf{x}}$ is an optimal solution of a canonical linear programming problem if $f(\overline{\mathbf{x}})$ is equal to the maximum value of the linear functional f on the feasible set $\mathcal{F}$.

14. **(T/F)** If $\overline{\mathbf{x}}$ is an optimal solution of a canonical linear programming problem, then $\overline{\mathbf{x}}$ is an extreme point of the feasible set.

15. Solve the linear programming problem in Example 1.

16. Solve the linear programming problem in Example 2.

17. The Benri Company manufactures two kinds of kitchen gadgets: invertible widgets and collapsible whammies. The production process is divided into three departments: fabricating, packing, and shipping. The hours of labor required for each operation and the hours available in each department each day are shown below.

	Widgets	**Whammies**	**Time available**
Fabricating	5.0	2.0	200
Packing	0.2	0.4	16
Shipping	0.2	0.2	10

Suppose that the profit on each widget is \$20 and the profit on each whammy is \$26. How many widgets and how many whammies should be made each day to maximize the company's profit?

Exercises 18–21 use the notion of a convex set, studied in Section 8.3. A set S in $\mathbb{R}^n$ is convex if, for each $\mathbf{p}$ and $\mathbf{q}$ in S, the line segment between $\mathbf{p}$ and $\mathbf{q}$ lies in S. [This line segment is the set of points of the form $(1-t)\mathbf{p}+t\mathbf{q}$ for $0 \le t \le 1$.]

18. Let $\mathcal{F}$ be the feasible set of all solutions $\mathbf{x}$ of a linear programming problem $A\mathbf{x} \le \mathbf{b}$ with $\mathbf{x} \ge \mathbf{0}$. Assume that $\mathcal{F}$ is nonempty. Show that $\mathcal{F}$ is a convex set in $\mathbb{R}^n$. [*Hint:* Consider points $\mathbf{p}$ and $\mathbf{q}$ in $\mathcal{F}$ and t such that $0 \le t \le 1$. Show that $(1-t)\mathbf{p}+t\mathbf{q}$ is in $\mathcal{F}$.]

19. Let $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. The inequality $ax_1 + bx_2 \le c$ for some real number c may be written as $\mathbf{v}^T\mathbf{x} \le c$. The set S of all $\mathbf{x}$ that satisfy this inequality is called a **closed half-space** of $\mathbb{R}^2$. Show that S is convex. [See the Hint for Exercise 18.]

20. The feasible set in Example 5 is the intersection of five closed half-spaces. By Exercise 19, these half-spaces are convex sets. Show that the intersection of any five convex sets $S_1, \ldots, S_5$ in $\mathbb{R}^n$ is a convex set.

21. If c is in $\mathbb{R}^n$ and if f is defined on $\mathbb{R}^n$ by $f(\mathbf{x}) = \mathbf{c}^T\mathbf{x}$, then f is called a linear functional, and for any real number d, $\{\mathbf{x}: f(\mathbf{x}) = d\}$ is called a level set of f. (See level sets in Figure 2 of Example 5.) Show that any such level set is convex.

Solutions to Practice Problems

1. The first inequality has the wrong direction, so multiply by -1. This gives the following problem:

$$\begin{aligned} &\text{Maximize} && 2x_1 + x_2 \\ &\text{subject to} && -x_1 + 2x_2 \le 8 \\ &&& 3x_1 + 2x_2 \le 24 \\ &\text{and } x_1 \ge 0, x_2 \ge 0. \end{aligned}$$

This corresponds to the canonical form

$$\text{Maximize } \mathbf{c}^T\mathbf{x} \text{ subject to } A\mathbf{x} \le \mathbf{b} \text{ and } \mathbf{x} \ge \mathbf{0}$$

when

$$\mathbf{b} = \begin{bmatrix} 8 \\ 24 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad \text{and} \quad A = \begin{bmatrix} -1 & 2 \\ 3 & 2 \end{bmatrix}$$

2. To graph the inequality $-x_1 + 2x_2 \le 8$, first graph the corresponding equality $-x_1 + 2x_2 = 8$. The intercepts are easy to find: $(0, 4)$ and $(-8, 0)$. Figure 7 shows the straight line through these two points.

The graph of the inequality consists of this line together with all points on one side of the line. To determine which side, pick a point not on the line to see if its coordinates satisfy the inequality. For example, try the origin, $(0, 0)$. The inequality

$$-(0) + 2(0) \le 8$$

is a true statement. Thus the origin and all other points below the line satisfy the inequality. As another example, substituting the coordinates of the point $(0, 8)$ into the inequality produces a false statement:

$$-(0) + 2(8) \le 8$$

Thus $(0, 8)$ and all other points above the line do not satisfy the inequality. Figure 7 shows small arrows beneath the graph of $-x_1 + 2x_2 = 8$, to indicate which side is to be included.

For the inequality

$$3x_1 + 2x_2 \le 24$$

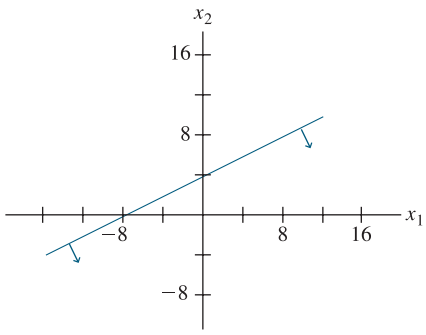


FIGURE 7 Graph of $-x_1 + 2x_2 \le 8$.

draw the graph of $3x_1 + 2x_2 = 24$, using the intercepts $(0, 12)$ and $(8, 0)$ or two other convenient points. Since $(0, 0)$ satisfies the inequality, the feasible set is on the side of the line containing the origin. The inequality $x_1 \ge 0$ gives the right half-plane, and the inequality $x_2 \ge 0$ gives the upper half-plane. All of these are graphed in Figure 8, and their common solution is the shaded feasible set.

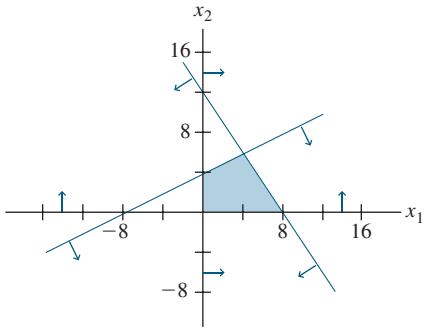


FIGURE 8 Graph of the feasible set.

3. There are four extreme points in the feasible set:

1. The origin: $(0, 0)$

2. The x_2-intercept of the first equality: $(0, 4)$

3. The x_1-intercept of the second equality: $(8, 0)$

4. The intersection of the two equalities.

For the fourth extreme point, solve the system of equations $-x_1 + 2x_2 = 8$ and $3x_1 + 2x_2 = 24$ to obtain $x_1 = 4$ and $x_2 = 6$.

4. To find the maximum value of the objective function $2x_1 + x_2$, evaluate it at each of the four extreme points of the feasible set.

	$2x_1 + x_2$	
$(0, 0)$	$2(0) + 1(0) = 0$	
$(0, 4)$	$2(0) + 1(4) = 4$	
$(8, 0)$	$2(8) + 1(0) = 16$	← max
$(4, 6)$	$2(4) + 1(6) = 14$	

The maximum value is 16, attained when $x_1 = 8$ and $x_2 = 0$.