Guidelines

B.Sc. (H) Computer Science

DSC-03 (Mathematics for Computing)

S.No	Topic	Reference: Table of Content	Reference: Book
1	**Unit 1-** Introduction to Matrix Algebra: Echelon form of a Matrix, Rank of a Matrix, Determinant and Inverse of a matrix, Solution of System of Homogeneous & Non-Homogeneous Equations: Gauss elimination and Gauss Jordan Method.	7.1	[2]
		7.3	
		7.4 Pg.282-285	
		7.5	
		7.7 Pg 293-295	
		7.8 Pg 301-304	
2	**Unit 2 -** Vector Space and Linear Transformation: Vector Space, Sub-spaces, Linear Combinations, Linear Span, Convex Sets (Follow any Book) , Linear Independence/Dependence, Basis & Dimension, Linear transformation on finite dimensional vector spaces, Inner Product Space, Schwarz Inequality, Orthonormal Basis, Gram-Schmidt Orthogonalization Process	4.1 – 4.5 (Except Page no. 208 -212)	[3]
		5.1 – 5.4	
		7.5	
3	**Unit 3 -** EigenValue and EigenVector: Characteristic Polynomial, Cayley Hamilton Theorem (Only in numericals), Eigen Value And eigen vector of a matrix, eigenspaces, Diagonalization, Positive Definite Matrices, Applications to Markov Matrices	6.1 Introduction to eigen value (*Refer 4.2 for applications)	[1]
		6.2 Diagonalization	
		6.4 Symmetric Matrices	
		Cayley Hamilton Theorem Page no. 384	[3]
		6.5 Positive Definite Matrices	[1]
		8.3 Applications of Markov Matrix	[1]
4	**Unit 4 -** Vector Calculus: Vector Algebra,Laws of Vector Algebra, Dot Product, Cross Product,Vector and Scalar Fields, Ordinary Derivative of Vectors, Space Curves, Partial Derivatives, Del Operator, Gradient of a Scalar Field, Directional Derivative, Gradient of Matrices, Divergence of a Vector Field, Laplacian Operator, Curl of a Vector Field.	9.1 Vectors in 2-Space and 3-Space	[2]
		9.2 Inner Product (Dot Product)	
		9.3 Vector Product (Cross Product)	
		9.4 Vector and Scalar Functions and Their Fields. Vector Calculus: Derivatives	
		9.7 Gradient of a Scalar Field. Directional Derivative	
		9.8 Divergence of a Vector Field	
		9.9 Curl of a Vector Field	

Reference:

1. Strang Gilbert, “Introduction to Linear Algebra”, 5th Edition, Wellesley-Cambridge Press, 2021.
2. Kreyszig Erwin, “Advanced Engineering Mathematics”, 10th Edition, Wiley, 2015.
3. Stephen Andrilli and David Hecker, “Elementary Linear Algebra”, Fourth Edition, Academic Press, 2010, ISBN: 978-0-12-374751-8

* Deisenroth, Marc Peter, Faisal A. Aldo and Ong Chengsoonm “Mathematics for Machine Learning, 1st Edition, Cambridge University Press, 2020.

7.1 Matrices, Vectors: Addition and Scalar Multiplication

The basic concepts and rules of matrix and vector algebra are introduced in Secs. 7.1 and 7.2 and are followed by **linear systems** (systems of linear equations), a main application, in Sec. 7.3.

Let us first take a leisurely look at matrices before we formalize our discussion. A **matrix** is a rectangular array of numbers or functions which we will enclose in brackets. For example,

$$
\begin{bmatrix} 0.3 & 1 & -5 \\ 0 & -0.2 & 16 \end{bmatrix}, \quad \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \tag{1}
$$

$$
\begin{bmatrix} e^{-x} & 2x^2 \\ e^{6x} & 4x \end{bmatrix}, \quad [a_1 \quad a_2 \quad a_3], \quad \begin{bmatrix} 4 \\ \frac{1}{2} \end{bmatrix}
$$

are matrices. The numbers (or functions) are called **entries** or, less commonly, *elements* of the matrix. The first matrix in (1) has two **rows**, which are the horizontal lines of entries. Furthermore, it has three **columns**, which are the vertical lines of entries. The second and third matrices are **square matrices**, which means that each has as many rows as columns—3 and 2, respectively. The entries of the second matrix have two indices, signifying their location within the matrix. The first index is the number of the row and the second is the number of the column, so that together the entry's position is uniquely identified. For example, a_{23} (read *a two three*) is in Row 2 and Column 3, etc. The notation is standard and applies to all matrices, including those that are not square.

Matrices having just a single row or column are called **vectors**. Thus, the fourth matrix in (1) has just one row and is called a **row vector**. The last matrix in (1) has just one column and is called a **column vector**. Because the goal of the indexing of entries was to uniquely identify the position of an element within a matrix, one index suffices for vectors, whether they are row or column vectors. Thus, the third entry of the row vector in (1) is denoted by a_3.

Matrices are handy for storing and processing data in applications. Consider the following two common examples.

EXAMPLE 1 Linear Systems, a Major Application of Matrices

We are given a system of linear equations, briefly a **linear system**, such as

$$
\begin{aligned}
4x_1 + 6x_2 + 9x_3 &= 6 \\
6x_1 \qquad\quad - 2x_3 &= 20 \\
5x_1 - 8x_2 + x_3 &= 10
\end{aligned}
$$

where x_1, x_2, x_3 are the **unknowns**. We form the **coefficient matrix**, call it **A**, by listing the coefficients of the unknowns in the position in which they appear in the linear equations. In the second equation, there is no unknown x_2, which means that the coefficient of x_2 is 0 and hence in matrix **A**, $a_{22} = 0$, Thus,

$$\mathbf{A} = \begin{bmatrix} 4 & 6 & 9 \\ 6 & 0 & -2 \\ 5 & -8 & 1 \end{bmatrix}. \quad \text{We form another matrix} \quad \tilde{\mathbf{A}} = \begin{bmatrix} 4 & 6 & 9 & 6 \\ 6 & 0 & -2 & 20 \\ 5 & -8 & 1 & 10 \end{bmatrix}$$

by augmenting **A** with the right sides of the linear system and call it the augmented matrix of the system.

Since we can go back and recapture the system of linear equations directly from the augmented matrix $\tilde{\mathbf{A}}$, $\tilde{\mathbf{A}}$ contains all the information of the system and can thus be used to solve the linear system. This means that we can just use the augmented matrix to do the calculations needed to solve the system. We shall explain this in detail in Sec. 7.3. Meanwhile you may verify by substitution that the solution is $x_1 = 3, x_2 = \frac{1}{2}, x_3 = -1$.

The notation x_1, x_2, x_3 for the unknowns is practical but not essential; we could choose x, y, z or some other letters. ■

EXAMPLE 2 Sales Figures in Matrix Form

Sales figures for three products I, II, III in a store on Monday (Mon), Tuesday (Tues), $\cdots$ may for each week be arranged in a matrix

$$\mathbf{A} = \begin{array}{c} \begin{array}{ccccccc} \text{Mon} & \text{Tues} & \text{Wed} & \text{Thur} & \text{Fri} & \text{Sat} & \text{Sun} \end{array} \\ \begin{bmatrix} 40 & 33 & 81 & 0 & 21 & 47 & 33 \\ 0 & 12 & 78 & 50 & 50 & 96 & 90 \\ 10 & 0 & 0 & 27 & 43 & 78 & 56 \end{bmatrix} \end{array} \cdot \begin{array}{c} \text{I} \\ \text{II} \\ \text{III} \end{array}$$

If the company has 10 stores, we can set up 10 such matrices, one for each store. Then, by adding corresponding entries of these matrices, we can get a matrix showing the total sales of each product on each day. Can you think of other data which can be stored in matrix form? For instance, in transportation or storage problems? Or in listing distances in a network of roads? ■

General Concepts and Notations

Let us formalize what we just have discussed. We shall denote matrices by capital boldface letters **A**, **B**, **C**, $\cdots$, or by writing the general entry in brackets; thus $\mathbf{A} = [a_{jk}]$, and so on. By an ***m* × *n* matrix** (read *m by n matrix*) we mean a matrix with m rows and n columns—rows always come first! $m \times n$ is called the **size** of the matrix. Thus an $m \times n$ matrix is of the form

$$\mathbf{A} = [a_{jk}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}. \tag{2}$$

The matrices in (1) are of sizes 2×3, 3×3, 2×2, 1×3, and 2×1, respectively.

Each entry in (2) has two subscripts. The first is the *row number* and the second is the *column number*. Thus a_{21} is the entry in Row 2 and Column 1.

If $m = n$, we call **A** an $n \times n$ **square matrix**. Then its diagonal containing the entries $a_{11}, a_{22}, \cdots, a_{nn}$ is called the **main diagonal** of **A**. Thus the main diagonals of the two square matrices in (1) are a_{11}, a_{22}, a_{33} and e^{-x}, $4x$, respectively.

Square matrices are particularly important, as we shall see. A matrix of any size $m \times n$ is called a **rectangular matrix**; this includes square matrices as a special case.

Vectors

A **vector** is a matrix with only one row or column. Its entries are called the **components** of the vector. We shall denote vectors by ***lowercase*** boldface letters $\mathbf{a}, \mathbf{b}, \cdots$ or by its general component in brackets, $\mathbf{a} = [a_j]$, and so on. Our special vectors in (1) suggest that a (general) **row vector** is of the form

$$\mathbf{a} = [a_1 \quad a_2 \quad \cdots \quad a_n]. \qquad \text{For instance,} \qquad \mathbf{a} = [-2 \quad 5 \quad 0.8 \quad 0 \quad 1].$$

A **column vector** is of the form

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}. \qquad \text{For instance,} \qquad \mathbf{b} = \begin{bmatrix} 4 \\ 0 \\ -7 \end{bmatrix}.$$

Addition and Scalar Multiplication of Matrices and Vectors

What makes matrices and vectors really useful and particularly suitable for computers is the fact that we can calculate with them almost as easily as with numbers. Indeed, we now introduce rules for addition and for scalar multiplication (multiplication by numbers) that were suggested by practical applications. (Multiplication of matrices by matrices follows in the next section.) We first need the concept of equality.

DEFINITION

Equality of Matrices

Two matrices $\mathbf{A} = [a_{jk}]$ and $\mathbf{B} = [b_{jk}]$ are **equal**, written $\mathbf{A} = \mathbf{B}$, if and only if they have the same size and the corresponding entries are equal, that is, $a_{11} = b_{11}$, $a_{12} = b_{12}$, and so on. Matrices that are not equal are called **different**. Thus, matrices of different sizes are always different.

EXAMPLE 3 **Equality of Matrices**

Let

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \qquad \text{and} \qquad \mathbf{B} = \begin{bmatrix} 4 & 0 \\ 3 & -1 \end{bmatrix}.$$

Then

$$\mathbf{A} = \mathbf{B} \qquad \text{if and only if} \qquad \begin{matrix} a_{11} = 4, & a_{12} = 0, \\ a_{21} = 3, & a_{22} = -1. \end{matrix}$$

The following matrices are all different. Explain!

$$\begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix} \qquad \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix} \qquad \begin{bmatrix} 4 & 1 \\ 2 & 3 \end{bmatrix} \qquad \begin{bmatrix} 1 & 3 & 0 \\ 4 & 2 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 1 & 3 \\ 0 & 4 & 2 \end{bmatrix}$$ ■

DEFINITION

Addition of Matrices

The **sum** of two matrices $\mathbf{A} = [a_{jk}]$ and $\mathbf{B} = [b_{jk}]$ ***of the same size*** is written $\mathbf{A} + \mathbf{B}$ and has the entries $a_{jk} + b_{jk}$ obtained by adding the corresponding entries of $\mathbf{A}$ and $\mathbf{B}$. Matrices of different sizes cannot be added.

As a special case, the **sum** $\mathbf{a} + \mathbf{b}$ of two row vectors or two column vectors, which must have the same number of components, is obtained by adding the corresponding components.

EXAMPLE 4 **Addition of Matrices and Vectors**

If $\quad \mathbf{A} = \begin{bmatrix} -4 & 6 & 3 \\ 0 & 1 & 2 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 5 & -1 & 0 \\ 3 & 1 & 0 \end{bmatrix}$, then $\mathbf{A} + \mathbf{B} = \begin{bmatrix} 1 & 5 & 3 \\ 3 & 2 & 2 \end{bmatrix}$.

$\mathbf{A}$ in Example 3 and our present $\mathbf{A}$ cannot be added. If $\mathbf{a} = [5 \quad 7 \quad 2]$ and $\mathbf{b} = [-6 \quad 2 \quad 0]$, then $\mathbf{a} + \mathbf{b} = [-1 \quad 9 \quad 2]$.

An application of matrix addition was suggested in Example 2. Many others will follow. ■

DEFINITION

Scalar Multiplication (Multiplication by a Number)

The **product** of any $m \times n$ matrix $\mathbf{A} = [a_{jk}]$ and any **scalar** c (number c) is written $c\mathbf{A}$ and is the $m \times n$ matrix $c\mathbf{A} = [ca_{jk}]$ obtained by multiplying each entry of $\mathbf{A}$ by c.

Here $(-1)\mathbf{A}$ is simply written $-\mathbf{A}$ and is called the **negative** of $\mathbf{A}$. Similarly, $(-k)\mathbf{A}$ is written $-k\mathbf{A}$. Also, $\mathbf{A} + (-\mathbf{B})$ is written $\mathbf{A} - \mathbf{B}$ and is called the **difference** of $\mathbf{A}$ and $\mathbf{B}$ (which must have the same size!).

EXAMPLE 5 **Scalar Multiplication**

If $\quad \mathbf{A} = \begin{bmatrix} 2.7 & -1.8 \\ 0 & 0.9 \\ 9.0 & -4.5 \end{bmatrix}$, then $-\mathbf{A} = \begin{bmatrix} -2.7 & 1.8 \\ 0 & -0.9 \\ -9.0 & 4.5 \end{bmatrix}$, $\frac{10}{9}\mathbf{A} = \begin{bmatrix} 3 & -2 \\ 0 & 1 \\ 10 & -5 \end{bmatrix}$, $0\mathbf{A} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$.

If a matrix $\mathbf{B}$ shows the distances between some cities in miles, $1.609\mathbf{B}$ gives these distances in kilometers. ■

Rules for Matrix Addition and Scalar Multiplication. From the familiar laws for the addition of numbers we obtain similar laws for the addition of matrices of the same size $m \times n$, namely,

$$\textbf{(3)} \qquad \begin{aligned} &\text{(a)} \quad && \mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A} \\ &\text{(b)} && (\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}) \qquad \text{(written } \mathbf{A} + \mathbf{B} + \mathbf{C}\text{)} \\ &\text{(c)} && \mathbf{A} + \mathbf{0} = \mathbf{A} \\ &\text{(d)} && \mathbf{A} + (-\mathbf{A}) = \mathbf{0}. \end{aligned}$$

Here $\mathbf{0}$ denotes the **zero matrix** (of size $m \times n$), that is, the $m \times n$ matrix with all entries zero. If $m = 1$ or $n = 1$, this is a vector, called a **zero vector**.

Hence matrix addition is *commutative* and *associative* [by (3a) and (3b)].
Similarly, for scalar multiplication we obtain the rules

$$\textbf{(4)}\qquad \begin{aligned} &\text{(a)} & c(\mathbf{A}+\mathbf{B}) &= c\mathbf{A}+c\mathbf{B} \\ &\text{(b)} & (c+k)\mathbf{A} &= c\mathbf{A}+k\mathbf{A} \\ &\text{(c)} & c(k\mathbf{A}) &= (ck)\mathbf{A} \qquad \text{(written } ck\mathbf{A}\text{)} \\ &\text{(d)} & 1\mathbf{A} &= \mathbf{A}. \end{aligned}$$

PROBLEM SET 7.1

1–7 GENERAL QUESTIONS

1. **Equality.** Give reasons why the five matrices in Example 3 are all different.
2. **Double subscript notation.** If you write the matrix in Example 2 in the form $\mathbf{A} = [a_{jk}]$, what is a_{31}? a_{13}? a_{26}? a_{33}?
3. **Sizes.** What sizes do the matrices in Examples 1, 2, 3, and 5 have?
4. **Main diagonal.** What is the main diagonal of **A** in Example 1? Of **A** and **B** in Example 3?
5. **Scalar multiplication.** If **A** in Example 2 shows the number of items sold, what is the matrix **B** of units sold if a unit consists of **(a)** 5 items and **(b)** 10 items?
6. If a 12×12 matrix **A** shows the distances between 12 cities in kilometers, how can you obtain from **A** the matrix **B** showing these distances in miles?
7. **Addition of vectors.** Can you add: A row and a column vector with different numbers of components? With the same number of components? Two row vectors with the same number of components but different numbers of zeros? A vector and a scalar? A vector with four components and a 2×2 matrix?

8–16 ADDITION AND SCALAR MULTIPLICATION OF MATRICES AND VECTORS

Let

$$\mathbf{A} = \begin{bmatrix} 0 & 2 & 4 \\ 6 & 5 & 5 \\ 1 & 0 & -3 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 0 & 5 & 2 \\ 5 & 3 & 4 \\ -2 & 4 & -2 \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} 5 & 2 \\ -2 & 4 \\ 1 & 0 \end{bmatrix}, \qquad \mathbf{D} = \begin{bmatrix} -4 & 1 \\ 5 & 0 \\ 2 & -1 \end{bmatrix},$$

$$\mathbf{E} = \begin{bmatrix} 0 & 2 \\ 3 & 4 \\ 3 & -1 \end{bmatrix}$$

$$\mathbf{u} = \begin{bmatrix} 1.5 \\ 0 \\ -3.0 \end{bmatrix}, \qquad \mathbf{v} = \begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix}, \qquad \mathbf{w} = \begin{bmatrix} -5 \\ -30 \\ 10 \end{bmatrix}.$$

Find the following expressions, indicating which of the rules in (3) or (4) they illustrate, or give reasons why they are not defined.

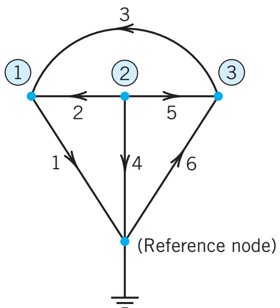
8. $2\mathbf{A} + 4\mathbf{B}$, $4\mathbf{B} + 2\mathbf{A}$, $0\mathbf{A} + \mathbf{B}$, $0.4\mathbf{B} - 4.2\mathbf{A}$
9. $3\mathbf{A}$, $0.5\mathbf{B}$, $3\mathbf{A} + 0.5\mathbf{B}$, $3\mathbf{A} + 0.5\mathbf{B} + \mathbf{C}$
10. $(4 \cdot 3)\mathbf{A}$, $4(3\mathbf{A})$, $14\mathbf{B} - 3\mathbf{B}$, $11\mathbf{B}$
11. $8\mathbf{C} + 10\mathbf{D}$, $2(5\mathbf{D} + 4\mathbf{C})$, $0.6\mathbf{C} - 0.6\mathbf{D}$, $0.6(\mathbf{C} - \mathbf{D})$
12. $(\mathbf{C} + \mathbf{D}) + \mathbf{E}$, $(\mathbf{D} + \mathbf{E}) + \mathbf{C}$, $0(\mathbf{C} - \mathbf{E}) + 4\mathbf{D}$, $\mathbf{A} - 0\mathbf{C}$
13. $(2 \cdot 7)\mathbf{C}$, $2(7\mathbf{C})$, $-\mathbf{D} + 0\mathbf{E}$, $\mathbf{E} - \mathbf{D} + \mathbf{C} + \mathbf{u}$
14. $(5\mathbf{u} + 5\mathbf{v}) - \frac{1}{2}\mathbf{w}$, $-20(\mathbf{u} + \mathbf{v}) + 2\mathbf{w}$, $\mathbf{E} - (\mathbf{u} + \mathbf{v})$, $10(\mathbf{u} + \mathbf{v}) + \mathbf{w}$
15. $(\mathbf{u} + \mathbf{v}) - \mathbf{w}$, $\mathbf{u} + (\mathbf{v} - \mathbf{w})$, $\mathbf{C} + 0\mathbf{w}$, $0\mathbf{E} + \mathbf{u} - \mathbf{v}$
16. $15\mathbf{v} - 3\mathbf{w} - 0\mathbf{u}$, $-3\mathbf{w} + 15\mathbf{v}$, $\mathbf{D} - \mathbf{u} + 3\mathbf{C}$, $8.5\mathbf{w} - 11.1\mathbf{u} + 0.4\mathbf{v}$
17. **Resultant of forces.** If the above vectors **u**, **v**, **w** represent forces in space, their sum is called their *resultant*. Calculate it.
18. **Equilibrium.** By definition, forces are *in equilibrium* if their resultant is the zero vector. Find a force **p** such that the above **u**, **v**, **w**, and **p** are in equilibrium.
19. **General rules.** Prove (3) and (4) for general 2×3 matrices and scalars c and k.

20. TEAM PROJECT. Matrices for Networks. Matrices have various engineering applications, as we shall see. For instance, they can be used to characterize connections in electrical networks, in nets of roads, in production processes, etc., as follows.

(a) Nodal Incidence Matrix. The network in Fig. 155 consists of six *branches* (connections) and four *nodes* (points where two or more branches come together). One node is the *reference node* (grounded node, whose voltage is zero). We number the other nodes and number and direct the branches. This we do arbitrarily. The network can now be described by a matrix $\mathbf{A} = [a_{jk}]$, where

$$a_{jk} = \begin{cases} +1 \text{ if branch } k \text{ leaves node } (j) \\ -1 \text{ if branch } k \text{ enters node } (j) \\ 0 \text{ if branch } k \text{ does not touch node } (j). \end{cases}$$

$\mathbf{A}$ is called the *nodal incidence matrix* of the network. Show that for the network in Fig. 155 the matrix $\mathbf{A}$ has the given form.



$$\begin{array}{l} \text{Branch} \\ \text{Node } (1) \\ \text{Node } (2) \\ \text{Node } (3) \end{array} \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{bmatrix} 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & -1 & -1 \end{bmatrix} \end{array}$$

Fig. 155. Network and nodal incidence matrix in Team Project 20(a)

(b) Find the nodal incidence matrices of the networks in Fig. 156.

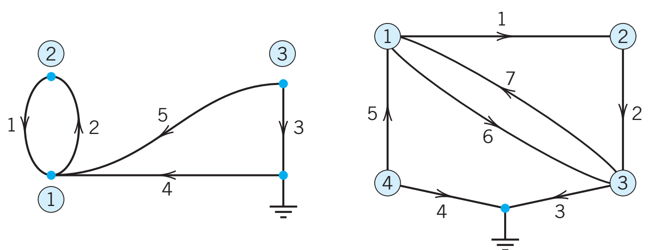


Fig. 156. Electrical networks in Team Project 20(b)

(c) Sketch the three networks corresponding to the nodal incidence matrices

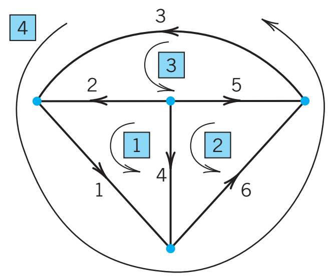
$$\begin{bmatrix} 1 & 0 & 0 & 1 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & -1 & 0 & 0 & 1 \\ -1 & 1 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 \end{bmatrix},$$

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ -1 & 1 & 0 & 1 & 0 \\ 0 & -1 & -1 & 0 & 1 \end{bmatrix}.$$

(d) Mesh Incidence Matrix. A network can also be characterized by the *mesh incidence matrix* $\mathbf{M} = [m_{jk}]$, where

$$m_{jk} = \begin{cases} +1 \text{ if branch } k \text{ is in mesh } \boxed{j} \\ \qquad \text{and has the same orientation} \\ -1 \text{ if branch } k \text{ is in mesh } \boxed{j} \\ \qquad \text{and has the opposite orientation} \\ 0 \text{ if branch } k \text{ is not in mesh } \boxed{j} \end{cases}$$

and a mesh is a loop with no branch in its interior (or in its exterior). Here, the meshes are numbered and directed (oriented) in an arbitrary fashion. Show that for the network in Fig. 157, the matrix $\mathbf{M}$ has the given form, where Row 1 corresponds to mesh 1, etc.



$$\mathbf{M} = \begin{bmatrix} 1 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 1 \\ 0 & -1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

Fig. 157. Network and matrix **M** in Team Project 20(d)

(d) Computer graphics. To visualize a three-dimensional object with plane faces (e.g., a cube), we may store the position vectors of the vertices with respect to a suitable $x_1x_2x_3$-coordinate system (and a list of the connecting edges) and then obtain a two-dimensional image on a video screen by projecting the object onto a coordinate plane, for instance, onto the x_1x_2-plane by setting $x_3 = 0$. To change the appearance of the image, we can impose a linear transformation on the position vectors stored. Show that a diagonal matrix $\mathbf{D}$ with main diagonal entries 3, 1, $\frac{1}{2}$ gives from an $\mathbf{x} = [x_j]$ the new position vector $\mathbf{y} = \mathbf{D}\mathbf{x}$, where $y_1 = 3x_1$ (stretch in the x_1-direction by a factor 3), $y_2 = x_2$ (unchanged), $y_3 = \frac{1}{2}x_3$ (contraction in the x_3-direction). What effect would a scalar matrix have?

(e) Rotations in space. Explain $\mathbf{y} = \mathbf{A}\mathbf{x}$ geometrically when $\mathbf{A}$ is one of the three matrices

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

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

What effect would these transformations have in situations such as that described in (d)?

7.3 Linear Systems of Equations. Gauss Elimination

We now come to one of the most important use of matrices, that is, using matrices to solve systems of linear equations. We showed informally, in Example 1 of Sec. 7.1, how to represent the information contained in a system of linear equations by a matrix, called the augmented matrix. This matrix will then be used in solving the linear system of equations. Our approach to solving linear systems is called the Gauss elimination method. Since this method is so fundamental to linear algebra, the student should be alert.

A shorter term for systems of linear equations is just **linear systems**. Linear systems model many applications in engineering, economics, statistics, and many other areas. Electrical networks, traffic flow, and commodity markets may serve as specific examples of applications.

Linear System, Coefficient Matrix, Augmented Matrix

A **linear system of *m* equations in *n* unknowns** $x_1, \cdots, x_n$ is a set of equations of the form

$$\textbf{(1)} \qquad \begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= b_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots&\cdots\cdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m. \end{aligned}$$

The system is called *linear* because each variable x_j appears in the first power only, just as in the equation of a straight line. $a_{11}, \cdots, a_{mn}$ are given numbers, called the **coefficients** of the system. $b_1, \cdots, b_m$ on the right are also given numbers. If all the b_j are zero, then (1) is called a **homogeneous system**. If at least one b_j is not zero, then (1) is called a **nonhomogeneous system**.

A **solution** of (1) is a set of numbers $x_1, \cdots, x_n$ that satisfies all the m equations. A **solution vector** of (1) is a vector $\mathbf{x}$ whose components form a solution of (1). If the system (1) is homogeneous, it always has at least the **trivial solution** $x_1 = 0, \cdots, x_n = 0$.

Matrix Form of the Linear System (1). From the definition of matrix multiplication we see that the m equations of (1) may be written as a single vector equation

$$\mathbf{Ax} = \mathbf{b} \tag{2}$$

where the **coefficient matrix** $\mathbf{A} = [a_{jk}]$ is the $m \times n$ matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

are column vectors. We assume that the coefficients a_{jk} are not all zero, so that $\mathbf{A}$ is not a zero matrix. Note that $\mathbf{x}$ has n components, whereas $\mathbf{b}$ has m components. The matrix

$$\tilde{\mathbf{A}} = \left[\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdots & \cdot & \cdot \\ a_{m1} & \cdots & a_{mn} & b_m \end{array}\right]$$

is called the **augmented matrix** of the system (1). The dashed vertical line could be omitted, as we shall do later. It is merely a reminder that the last column of $\tilde{\mathbf{A}}$ did not come from matrix $\mathbf{A}$ but came from vector $\mathbf{b}$. Thus, we *augmented* the matrix $\mathbf{A}$.

Note that ***the augmented matrix*** $\tilde{\mathbf{A}}$ ***determines the system* (1) *completely*** because it contains all the given numbers appearing in (1).

EXAMPLE 1 Geometric Interpretation. Existence and Uniqueness of Solutions

If $m = n = 2$, we have two equations in two unknowns x_1, x_2

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1 \\ a_{21}x_1 + a_{22}x_2 &= b_2. \end{aligned}$$

If we interpret x_1, x_2 as coordinates in the x_1x_2-plane, then each of the two equations represents a straight line, and (x_1, x_2) is a solution if and only if the point P with coordinates x_1, x_2 lies on both lines. Hence there are three possible cases (see Fig. 158 on next page):

(*a*) Precisely one solution if the lines intersect

(*b*) Infinitely many solutions if the lines coincide

(*c*) No solution if the lines are parallel

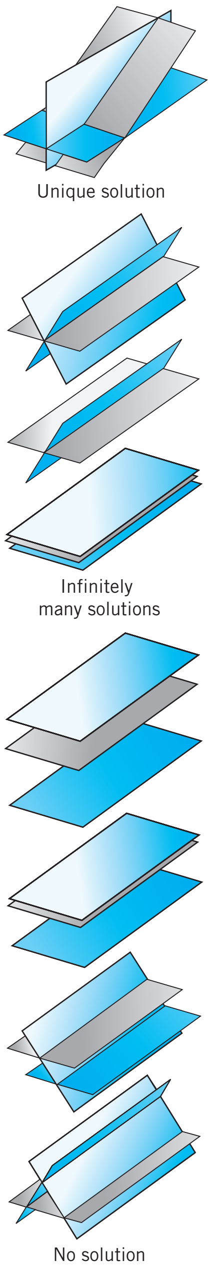
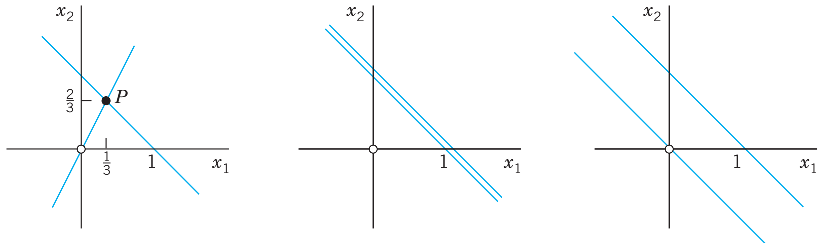


Fig. 158. Three equations in three unknowns interpreted as planes in space

For instance,

Case (a)	Case (b)	Case (c)
$x_1 + x_2 = 1$ $2x_1 - x_2 = 0$	$x_1 + x_2 = 1$ $2x_1 + 2x_2 = 2$	$x_1 + x_2 = 1$ $x_1 + x_2 = 0$



If the system is homogenous, Case (c) cannot happen, because then those two straight lines pass through the origin, whose coordinates (0, 0) constitute the trivial solution. Similarly, our present discussion can be extended from two equations in two unknowns to three equations in three unknowns. We give the geometric interpretation of three possible cases concerning solutions in Fig. 158. Instead of straight lines we have planes and the solution depends on the positioning of these planes in space relative to each other. The student may wish to come up with some specific examples. ■

Our simple example illustrated that a system (1) may have no solution. This leads to such questions as: Does a given system (1) have a solution? Under what conditions does it have precisely one solution? If it has more than one solution, how can we characterize the set of all solutions? We shall consider such questions in Sec. 7.5.

First, however, let us discuss an important systematic method for solving linear systems.

Gauss Elimination and Back Substitution

The Gauss elimination method can be motivated as follows. Consider a linear system that is in ***triangular form*** (in full, *upper* triangular form) such as

$$\begin{aligned} 2x_1 + 5x_2 &= 2 \\ 13x_2 &= -26 \end{aligned}$$

(*Triangular* means that all the nonzero entries of the corresponding coefficient matrix lie above the diagonal and form an upside-down 90° triangle.) Then we can solve the system by **back substitution**, that is, we solve the last equation for the variable, $x_2 = -26/13 = -2$, and then work backward, substituting $x_2 = -2$ into the first equation and solving it for x_1, obtaining $x_1 = \frac{1}{2}(2 - 5x_2) = \frac{1}{2}(2 - 5 \cdot (-2)) = 6$. This gives us the idea of first reducing a general system to triangular form. For instance, let the given system be

$$\begin{aligned} 2x_1 + 5x_2 &= 2 \\ -4x_1 + 3x_2 &= -30. \end{aligned} \qquad \text{Its augmented matrix is} \qquad \begin{bmatrix} 2 & 5 & 2 \\ -4 & 3 & -30 \end{bmatrix}.$$

We leave the first equation as it is. We eliminate x_1 from the second equation, to get a triangular system. For this we add twice the first equation to the second, and we do the same operation on the ***rows*** of the augmented matrix. This gives $-4x_1 + 4x_1 + 3x_2 + 10x_2 = -30 + 2 \cdot 2$, that is,

$$\begin{aligned} 2x_1 + 5x_2 &= 2 \\ 13x_2 &= -26 \end{aligned} \qquad \text{Row 2 + 2 Row 1} \qquad \begin{bmatrix} 2 & 5 & 2 \\ 0 & 13 & -26 \end{bmatrix}$$

where Row 2 + 2 Row 1 means "Add twice Row 1 to Row 2" in the original matrix. This is the **Gauss elimination** (for 2 equations in 2 unknowns) giving the triangular form, from which back substitution now yields $x_2 = -2$ and $x_1 = 6$, as before.

Since a linear system is completely determined by its augmented matrix, ***Gauss elimination can be done by merely considering the matrices***, as we have just indicated. We do this again in the next example, emphasizing the matrices by writing them first and the equations behind them, just as a help in order not to lose track.

EXAMPLE 2 Gauss Elimination. Electrical Network

Solve the linear system

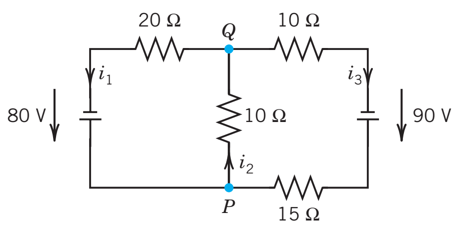
$$\begin{aligned} x_1 - x_2 + x_3 &= 0 \\ -x_1 + x_2 - x_3 &= 0 \\ 10x_2 + 25x_3 &= 90 \\ 20x_1 + 10x_2 &= 80. \end{aligned}$$

***Derivation from the circuit in Fig.* 159 (*Optional*).** This is the system for the unknown currents $x_1 = i_1$, $x_2 = i_2$, $x_3 = i_3$ in the electrical network in Fig. 159. To obtain it, we label the currents as shown, choosing directions arbitrarily; if a current will come out negative, this will simply mean that the current flows against the direction of our arrow. The current entering each battery will be the same as the current leaving it. The equations for the currents result from Kirchhoff's laws:

Kirchhoff's Current Law (KCL). *At any point of a circuit, the sum of the inflowing currents equals the sum of the outflowing currents.*

Kirchhoff's Voltage Law (KVL). *In any closed loop, the sum of all voltage drops equals the impressed electromotive force.*

Node P gives the first equation, node Q the second, the right loop the third, and the left loop the fourth, as indicated in the figure.



Node P:	$i_1 - i_2 + i_3 = 0$
Node Q:	$-i_1 + i_2 - i_3 = 0$
Right loop:	$10i_2 + 25i_3 = 90$
Left loop:	$20i_1 + 10i_2 = 80$

Fig. 159. Network in Example 2 and equations relating the currents

Solution by Gauss Elimination. This system could be solved rather quickly by noticing its particular form. But this is not the point. The point is that the Gauss elimination is systematic and will work in general, also for large systems. We apply it to our system and then do back substitution. As indicated, let us write the augmented matrix of the system first and then the system itself:

$$\begin{array}{c} \text{Augmented Matrix } \tilde{\mathbf{A}} \\ \begin{array}{l} \text{Pivot 1} \longrightarrow \\ \\ \text{Eliminate} \longrightarrow \\ \\ \end{array} \left[\begin{array}{rrr|r} 1 & -1 & 1 & 0 \\ -1 & 1 & -1 & 0 \\ 0 & 10 & 25 & 90 \\ 20 & 10 & 0 & 80 \end{array}\right] \end{array} \qquad \begin{array}{c} \text{Equations} \\ \begin{array}{l} \text{Pivot 1} \longrightarrow \\ \\ \text{Eliminate} \longrightarrow \\ \\ \end{array} \begin{array}{rcrcrcr} x_1 & - & x_2 & + & x_3 & = & 0 \\ -x_1 & + & x_2 & - & x_3 & = & 0 \\ & & 10x_2 & + & 25x_3 & = & 90 \\ 20x_1 & + & 10x_2 & & & = & 80. \end{array} \end{array}$$

Step 1. Elimination of x_1

Call the first row of $\mathbf{A}$ the **pivot row** and the first equation the **pivot equation**. Call the coefficient 1 of its x_1-term the **pivot** in this step. Use this equation to eliminate x_1 (get rid of x_1) in the other equations. For this, do:

Add 1 times the pivot equation to the second equation.

Add -20 times the pivot equation to the fourth equation.

This corresponds to **row operations** on the augmented matrix as indicated in BLUE behind the ***new matrix*** in (3). So the operations are performed on the ***preceding matrix***. The result is

$$\left[\begin{array}{rrr|r} 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 10 & 25 & 90 \\ 0 & 30 & -20 & 80 \end{array}\right] \begin{array}{l} \\ \text{Row 2 + Row 1} \\ \\ \text{Row 4} - \text{20 Row 1} \end{array} \qquad \begin{array}{rcrcrcr} x_1 & - & x_2 & + & x_3 & = & 0 \\ & & & & 0 & = & 0 \\ & & 10x_2 & + & 25x_3 & = & 90 \\ & & 30x_2 & - & 20x_3 & = & 80. \end{array} \tag{3}$$

Step 2. Elimination of x_2

The first equation remains as it is. We want the new second equation to serve as the next pivot equation. But since it has no x_2-term (in fact, it is $0 = 0$), we must first change the order of the equations and the corresponding rows of the new matrix. We put $0 = 0$ at the end and move the third equation and the fourth equation one place up. This is called **partial pivoting** (as opposed to the rarely used *total pivoting*, in which the order of the unknowns is also changed). It gives

$$\begin{array}{l} \\ \text{Pivot 10} \longrightarrow \\ \text{Eliminate 30} \longrightarrow \\ \\ \end{array} \left[\begin{array}{rrr|r} 1 & -1 & 1 & 0 \\ 0 & 10 & 25 & 90 \\ 0 & 30 & -20 & 80 \\ 0 & 0 & 0 & 0 \end{array}\right] \qquad \begin{array}{l} \\ \text{Pivot 10} \longrightarrow \\ \text{Eliminate } 30x_2 \longrightarrow \\ \\ \end{array} \begin{array}{rcrcrcr} x_1 & - & x_2 & + & x_3 & = & 0 \\ & & 10x_2 & + & 25x_3 & = & 90 \\ & & 30x_2 & - & 20x_3 & = & 80 \\ & & & & 0 & = & 0. \end{array}$$

To eliminate x_2, do:

Add -3 times the pivot equation to the third equation.

The result is

$$\left[\begin{array}{rrr|r} 1 & -1 & 1 & 0 \\ 0 & 10 & 25 & 90 \\ 0 & 0 & -95 & -190 \\ 0 & 0 & 0 & 0 \end{array}\right] \begin{array}{l} \\ \\ \text{Row 3} - \text{3 Row 2} \\ \\ \end{array} \qquad \begin{array}{rcrcrcr} x_1 & - & x_2 & + & x_3 & = & 0 \\ & & 10x_2 & + & 25x_3 & = & 90 \\ & & & - & 95x_3 & = & -190 \\ & & & & 0 & = & 0. \end{array} \tag{4}$$

Back Substitution. ***Determination of x_3, x_2, x_1 (in this order)***

Working backward from the last to the first equation of this "triangular" system (4), we can now readily find x_3, then x_2, and then x_1:

$$\begin{array}{rcrcrcr} & & & - & 95x_3 & = & -190 \\ & & 10x_2 & + & 25x_3 & = & 90 \\ x_1 & - & x_2 & + & x_3 & = & 0 \end{array} \qquad \begin{aligned} x_3 &= i_3 = 2\,[\mathrm{A}] \\ x_2 &= \tfrac{1}{10}(90 - 25x_3) = i_2 = 4\,[\mathrm{A}] \\ x_1 &= x_2 - x_3 = i_1 = 2\,[\mathrm{A}] \end{aligned}$$

where A stands for "amperes." This is the answer to our problem. The solution is unique. ■

Elementary Row Operations. Row-Equivalent Systems

Example 2 illustrates the operations of the Gauss elimination. These are the first two of three operations, which are called

Elementary Row Operations for Matrices:

Interchange of two rows

Addition of a constant multiple of one row to another row

Multiplication of a row by a ***nonzero*** *constant c*

CAUTION! These operations are for rows, ***not for columns***! They correspond to the following

Elementary Operations for Equations:

Interchange of two equations

Addition of a constant multiple of one equation to another equation

Multiplication of an equation by a ***nonzero*** *constant c*

Clearly, the interchange of two equations does not alter the solution set. Neither does their addition because we can undo it by a corresponding subtraction. Similarly for their multiplication, which we can undo by multiplying the new equation by $1/c$ (since $c \neq 0$), producing the original equation.

We now call a linear system S_1 **row-equivalent** to a linear system S_2 if S_1 can be obtained from S_2 by (finitely many!) row operations. This justifies Gauss elimination and establishes the following result.

THEOREM 1

Row-Equivalent Systems

Row-equivalent linear systems have the same set of solutions.

Because of this theorem, systems having the same solution sets are often called *equivalent systems*. But note well that we are dealing with ***row operations***. No column operations on the augmented matrix are permitted in this context because they would generally alter the solution set.

A linear system (1) is called **overdetermined** if it has more equations than unknowns, as in Example 2, **determined** if $m = n$, as in Example 1, and **underdetermined** if it has fewer equations than unknowns.

Furthermore, a system (1) is called **consistent** if it has at least one solution (thus, one solution or infinitely many solutions), but **inconsistent** if it has no solutions at all, as $x_1 + x_2 = 1, x_1 + x_2 = 0$ in Example 1, Case (c).

Gauss Elimination: The Three Possible Cases of Systems

We have seen, in Example 2, that Gauss elimination can solve linear systems that have a unique solution. This leaves us to apply Gauss elimination to a system with infinitely many solutions (in Example 3) and one with no solution (in Example 4).

EXAMPLE 3 Gauss Elimination if Infinitely Many Solutions Exist

Solve the following linear system of three equations in four unknowns whose augmented matrix is

$$(5)\qquad \left[\begin{array}{cccc|c} 3.0 & 2.0 & 2.0 & -5.0 & 8.0 \\ 0.6 & 1.5 & 1.5 & -5.4 & 2.7 \\ 1.2 & -0.3 & -0.3 & 2.4 & 2.1 \end{array}\right]. \qquad \text{Thus,} \qquad \begin{aligned} 3.0x_1 + 2.0x_2 + 2.0x_3 - 5.0x_4 &= 8.0 \\ \boxed{0.6x_1} + 1.5x_2 + 1.5x_3 - 5.4x_4 &= 2.7 \\ \boxed{1.2x_1} - 0.3x_2 - 0.3x_3 + 2.4x_4 &= 2.1. \end{aligned}$$

Solution. As in the previous example, we circle pivots and box terms of equations and corresponding entries to be eliminated. We indicate the operations in terms of equations and operate on both equations and matrices.

Step 1. Elimination of x_1 from the second and third equations by adding

$$-0.6/3.0 = -0.2 \text{ times the first equation to the second equation,}$$

$$-1.2/3.0 = -0.4 \text{ times the first equation to the third equation.}$$

This gives the following, in which the pivot of the next step is circled.

$$(6)\qquad \left[\begin{array}{cccc|c} 3.0 & 2.0 & 2.0 & -5.0 & 8.0 \\ 0 & 1.1 & 1.1 & -4.4 & 1.1 \\ 0 & -1.1 & -1.1 & 4.4 & -1.1 \end{array}\right] \begin{array}{l} \\ \text{Row 2} - 0.2\text{ Row 1} \\ \text{Row 3} - 0.4\text{ Row 1} \end{array} \qquad \begin{aligned} 3.0x_1 + 2.0x_2 + 2.0x_3 - 5.0x_4 &= 8.0 \\ 1.1x_2 + 1.1x_3 - 4.4x_4 &= 1.1 \\ \boxed{-1.1x_2} - 1.1x_3 + 4.4x_4 &= -1.1. \end{aligned}$$

Step 2. Elimination of x_2 from the third equation of (6) by adding

$$1.1/1.1 = 1 \text{ times the second equation to the third equation.}$$

This gives

$$(7)\qquad \left[\begin{array}{cccc|c} 3.0 & 2.0 & 2.0 & -5.0 & 8.0 \\ 0 & 1.1 & 1.1 & -4.4 & 1.1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \begin{array}{l} \\ \\ \text{Row 3} + \text{Row 2} \end{array} \qquad \begin{aligned} 3.0x_1 + 2.0x_2 + 2.0x_3 - 5.0x_4 &= 8.0 \\ 1.1x_2 + 1.1x_3 - 4.4x_4 &= 1.1 \\ 0 &= 0. \end{aligned}$$

Back Substitution. From the second equation, $x_2 = 1 - x_3 + 4x_4$. From this and the first equation, $x_1 = 2 - x_4$. Since x_3 and x_4 remain arbitrary, we have infinitely many solutions. If we choose a value of x_3 and a value of x_4, then the corresponding values of x_1 and x_2 are uniquely determined.

On Notation. If unknowns remain arbitrary, it is also customary to denote them by other letters $t_1, t_2, \cdots$. In this example we may thus write $x_1 = 2 - x_4 = 2 - t_2$, $x_2 = 1 - x_3 + 4x_4 = 1 - t_1 + 4t_2$, $x_3 = t_1$ (first arbitrary unknown), $x_4 = t_2$ (second arbitrary unknown). ■

EXAMPLE 4 Gauss Elimination if no Solution Exists

What will happen if we apply the Gauss elimination to a linear system that has no solution? The answer is that in this case the method will show this fact by producing a contradiction. For instance, consider

$$\left[\begin{array}{ccc|c} 3 & 2 & 1 & 3 \\ 2 & 1 & 1 & 0 \\ 6 & 2 & 4 & 6 \end{array}\right] \qquad \begin{aligned} 3x_1 + 2x_2 + x_3 &= 3 \\ \boxed{2x_1} + x_2 + x_3 &= 0 \\ \boxed{6x_1} + 2x_2 + 4x_3 &= 6. \end{aligned}$$

Step 1. Elimination of x_1 from the second and third equations by adding

$$-\tfrac{2}{3} \text{ times the first equation to the second equation,}$$

$$-\tfrac{6}{3} = -2 \text{ times the first equation to the third equation.}$$

This gives

$$\left[\begin{array}{ccc|c} 3 & 2 & 1 & 3 \\ 0 & -\frac{1}{3} & \frac{1}{3} & -2 \\ 0 & -2 & 2 & 0 \end{array}\right] \quad \begin{array}{l} \\ \text{Row 2} - \frac{2}{3}\text{ Row 1} \\ \text{Row 3} - 2\text{ Row 1} \end{array} \qquad \begin{aligned} 3x_1 + 2x_2 + \;x_3 &= \;\;3 \\ -\tfrac{1}{3}x_2 + \tfrac{1}{3}x_3 &= -2 \\ \boxed{-\,2x_2} + 2x_3 &= \;\;0. \end{aligned}$$

Step 2. Elimination of x_2 from the third equation gives

$$\left[\begin{array}{ccc|c} 3 & 2 & 1 & 3 \\ 0 & -\frac{1}{3} & \frac{1}{3} & -2 \\ 0 & 0 & 0 & 12 \end{array}\right] \quad \begin{array}{l} \\ \\ \text{Row 3} - 6\text{ Row 2} \end{array} \qquad \begin{aligned} 3x_1 + 2x_2 + \;x_3 &= \;\;3 \\ -\tfrac{1}{3}x_2 + \tfrac{1}{3}x_3 &= -\,2 \\ 0 &= \;\;12. \end{aligned}$$

The false statement $0 = 12$ shows that the system has no solution. ■

Row Echelon Form and Information From It

At the end of the Gauss elimination the form of the coefficient matrix, the augmented matrix, and the system itself are called the **row echelon form**. In it, rows of zeros, if present, are the last rows, and, in each nonzero row, the leftmost nonzero entry is farther to the right than in the previous row. For instance, in Example 4 the coefficient matrix and its augmented in row echelon form are

$$\text{(8)} \qquad \begin{bmatrix} 3 & 2 & 1 \\ 0 & -\frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \left[\begin{array}{ccc|c} 3 & 2 & 1 & 3 \\ 0 & -\frac{1}{3} & \frac{1}{3} & -2 \\ 0 & 0 & 0 & 12 \end{array}\right].$$

Note that we do not require that the leftmost nonzero entries be 1 since this would have no theoretic or numeric advantage. (The so-called *reduced echelon form*, in which those entries *are* 1, will be discussed in Sec. 7.8.)

The original system of m equations in n unknowns has augmented matrix $[\mathbf{A} \mid \mathbf{b}]$. This is to be row reduced to matrix $[\mathbf{R} \mid \mathbf{f}]$. The two systems $\mathbf{Ax} = \mathbf{b}$ and $\mathbf{Rx} = \mathbf{f}$ are equivalent: if either one has a solution, so does the other, and the solutions are identical.

At the end of the Gauss elimination (before the back substitution), the row echelon form of the augmented matrix will be

$$\text{(9)} \qquad \left[\begin{array}{ccccc|c} r_{11} & r_{12} & \cdots\cdots & & r_{1n} & f_1 \\ & r_{22} & \cdots\cdots & & r_{2n} & f_2 \\ & & \ddots & & \vdots & \vdots \\ & & & r_{rr} & \cdots \; r_{rn} & f_r \\ & & & & & f_{r+1} \\ & & & & & \vdots \\ & & & & & f_m \end{array}\right].$$

Here, $r \leqq m$, $r_{11} \neq 0$, and all entries in the blue triangle and blue rectangle are zero.

The number of nonzero rows, r, in the row-reduced coefficient matrix $\mathbf{R}$ is called the **rank of R** and also the **rank of A**. Here is the method for determining whether $\mathbf{Ax} = \mathbf{b}$ has solutions and what they are:

(a) No solution. If r is less than m (meaning that $\mathbf{R}$ actually has at least one row of all 0s) *and* at least one of the numbers $f_{r+1}, f_{r+2}, \cdots, f_m$ is not zero, then the system

$\mathbf{Rx} = \mathbf{f}$ is inconsistent: No solution is possible. Therefore the system $\mathbf{Ax} = \mathbf{b}$ is inconsistent as well. See Example 4, where $r = 2 < m = 3$ and $f_{r+1} = f_3 = 12$.

If the system is consistent (either $r = m$, or $r < m$ *and* all the numbers $f_{r+1}, f_{r+2}, \cdots, f_m$ are zero), then there are solutions.

(b) Unique solution. If the system is consistent and $r = n$, there is exactly one solution, which can be found by back substitution. See Example 2, where $r = n = 3$ and $m = 4$.

(c) Infinitely many solutions. To obtain any of these solutions, choose values of $x_{r+1}, \cdots, x_n$ arbitrarily. Then solve the rth equation for x_r (in terms of those arbitrary values), then the $(r-1)$st equation for x_{r-1}, and so on up the line. See Example 3.

Orientation. Gauss elimination is reasonable in computing time and storage demand. We shall consider those aspects in Sec. 20.1 in the chapter on numeric linear algebra. Section 7.4 develops fundamental concepts of linear algebra such as linear independence and rank of a matrix. These in turn will be used in Sec. 7.5 to fully characterize the behavior of linear systems in terms of existence and uniqueness of solutions.

PROBLEM SET 7.3

1–14 GAUSS ELIMINATION

Solve the linear system given explicitly or by its augmented matrix. Show details.

1. $$\begin{aligned} 4x - 6y &= -11 \\ -3x + 8y &= 10 \end{aligned}$$

2. $$\begin{bmatrix} 3.0 & -0.5 & 0.6 \\ 1.5 & 4.5 & 6.0 \end{bmatrix}$$

3. $$\begin{aligned} x + y - z &= 9 \\ 8y + 6z &= -6 \\ -2x + 4y - 6z &= 40 \end{aligned}$$

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

5. $$\begin{bmatrix} 13 & 12 & -6 \\ -4 & 7 & -73 \\ 11 & -13 & 157 \end{bmatrix}$$

6. $$\begin{bmatrix} 4 & -8 & 3 & 16 \\ -1 & 2 & -5 & -21 \\ 3 & -6 & 1 & 7 \end{bmatrix}$$

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

8. $$\begin{aligned} 4y + 3z &= 8 \\ 2x - z &= 2 \\ 3x + 2y &= 5 \end{aligned}$$

9. $$\begin{aligned} -2y - 2z &= -8 \\ 3x + 4y - 5z &= 13 \end{aligned}$$

10. $$\begin{bmatrix} 5 & -7 & 3 & 17 \\ -15 & 21 & -9 & 50 \end{bmatrix}$$

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

12. $$\begin{bmatrix} 2 & -2 & 4 & 0 & 0 \\ -3 & 3 & -6 & 5 & 15 \\ 1 & -1 & 2 & 0 & 0 \end{bmatrix}$$

13. $$\begin{aligned} 10x + 4y - 2z &= -4 \\ -3w - 17x + y + 2z &= 2 \\ w + x + y &= 6 \\ 8w - 34x + 16y - 10z &= 4 \end{aligned}$$

14. $$\begin{bmatrix} 2 & 3 & 1 & -11 & 1 \\ 5 & -2 & 5 & -4 & 5 \\ 1 & -1 & 3 & -3 & 3 \\ 3 & 4 & -7 & 2 & -7 \end{bmatrix}$$

15. Equivalence relation. By definition, an *equivalence relation* on a set is a relation satisfying three conditions: (named as indicated)

(i) Each element A of the set is equivalent to itself (*Reflexivity*).

(ii) If A is equivalent to B, then B is equivalent to A (*Symmetry*).

(iii) If A is equivalent to B and B is equivalent to C, then A is equivalent to C (*Transitivity*).

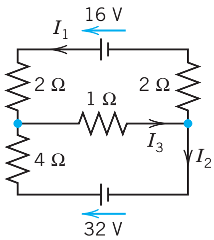
Show that row equivalence of matrices satisfies these three conditions. *Hint.* Show that for each of the three elementary row operations these conditions hold.

16. CAS PROJECT. Gauss Elimination and Back Substitution. Write a program for Gauss elimination and back substitution **(a)** that does not include pivoting and **(b)** that does include pivoting. Apply the programs to Probs. 11–14 and to some larger systems of your choice.

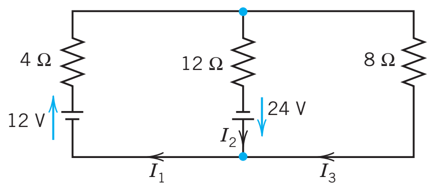
17–21 MODELS OF NETWORKS

In Probs. 17–19, using Kirchhoff's laws (see Example 2) and showing the details, find the currents:

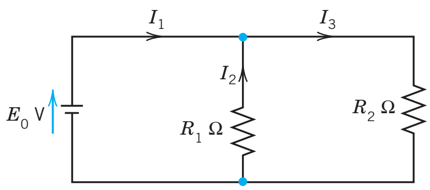
17.



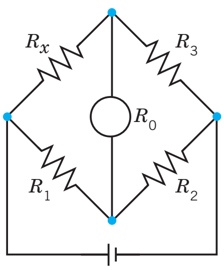
18.



19.

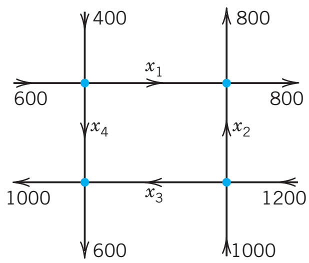


20. Wheatstone bridge. Show that if $R_x/R_3 = R_1/R_2$ in the figure, then $I = 0$. (R_0 is the resistance of the instrument by which I is measured.) This bridge is a method for determining R_x. R_1, R_2, R_3 are known. R_3 is variable. To get R_x, make $I = 0$ by varying R_3. Then calculate $R_x = R_3R_1/R_2$.



Wheatstone bridge

Problem 20



Net of one-way streets

Problem 21

21. Traffic flow. Methods of electrical circuit analysis have applications to other fields. For instance, applying the analog of Kirchhoff's Current Law, find the traffic flow (cars per hour) in the net of one-way streets (in the directions indicated by the arrows) shown in the figure. Is the solution unique?

22. Models of markets. Determine the equilibrium solution ($D_1 = S_1, D_2 = S_2$) of the two-commodity market with linear model (D, S, P = demand, supply, price; index 1 = first commodity, index 2 = second commodity)

$$D_1 = 40 - 2P_1 - P_2, \qquad S_1 = 4P_1 - P_2 + 4,$$

$$D_2 = 5P_1 - 2P_2 + 16, \qquad S_2 = 3P_2 - 4.$$

23. Balancing a chemical equation $x_1C_3H_8 + x_2O_2 \rightarrow x_3CO_2 + x_4H_2O$ means finding integer x_1, x_2, x_3, x_4 such that the numbers of atoms of carbon (C), hydrogen (H), and oxygen (O) are the same on both sides of this reaction, in which propane C_3H_8 and O_2 give carbon dioxide and water. Find the smallest positive integers $x_1, \cdots, x_4$.

24. PROJECT. Elementary Matrices. The idea is that elementary operations can be accomplished by matrix multiplication. If **A** is an $m \times n$ matrix on which we want to do an elementary operation, then there is a matrix **E** such that **EA** is the new matrix after the operation. Such an **E** is called an **elementary matrix**. This idea can be helpful, for instance, in the design of algorithms. (*Computationally*, it is generally preferable to do row operations *directly*, rather than by multiplication by **E**.)

(a) Show that the following are elementary matrices, for interchanging Rows 2 and 3, for adding -5 times the first row to the third, and for multiplying the fourth row by 8.

$$\mathbf{E}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{E}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -5 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$\mathbf{E}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}.$$

Apply $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$ to a vector and to a 4×3 matrix of your choice. Find $\mathbf{B} = \mathbf{E}_3\mathbf{E}_2\mathbf{E}_1\mathbf{A}$, where $\mathbf{A} = [a_{jk}]$ is the general 4×2 matrix. Is $\mathbf{B}$ equal to $\mathbf{C} = \mathbf{E}_1\mathbf{E}_2\mathbf{E}_3\mathbf{A}$?

(b) Conclude that $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$ are obtained by doing the corresponding elementary operations on the 4×4 unit matrix. Prove that *if* $\mathbf{M}$ *is obtained from* $\mathbf{A}$ *by an elementary row operation, then*

$$\mathbf{M} = \mathbf{EA},$$

where $\mathbf{E}$ *is obtained from the* $n \times n$ *unit matrix* $\mathbf{I}_n$ *by the same row operation.*

7.4 Linear Independence. Rank of a Matrix. Vector Space

Since our next goal is to fully characterize the behavior of linear systems in terms of existence and uniqueness of solutions (Sec. 7.5), we have to introduce new fundamental linear algebraic concepts that will aid us in doing so. Foremost among these are ***linear independence*** and the ***rank of a matrix***. Keep in mind that these concepts are intimately linked with the important Gauss elimination method and how it works.

Linear Independence and Dependence of Vectors

Given any set of m vectors $\mathbf{a}_{(1)}, \cdots, \mathbf{a}_{(m)}$ (with the same number of components), a **linear combination** of these vectors is an expression of the form

$$c_1\mathbf{a}_{(1)} + c_2\mathbf{a}_{(2)} + \cdots + c_m\mathbf{a}_{(m)}$$

where $c_1, c_2, \cdots, c_m$ are any scalars. Now consider the equation

$$c_1\mathbf{a}_{(1)} + c_2\mathbf{a}_{(2)} + \cdots + c_m\mathbf{a}_{(m)} = \mathbf{0}. \tag{1}$$

Clearly, this vector equation (1) holds if we choose all c_j's zero, because then it becomes $\mathbf{0} = \mathbf{0}$. If this is the only m-tuple of scalars for which (1) holds, then our vectors $\mathbf{a}_{(1)}, \cdots, \mathbf{a}_{(m)}$ are said to form a *linearly independent set* or, more briefly, we call them **linearly independent**. Otherwise, if (1) also holds with scalars not all zero, we call these vectors **linearly dependent**. This means that we can express at least one of the vectors as a linear combination of the other vectors. For instance, if (1) holds with, say, $c_1 \neq 0$, we can solve (1) for $\mathbf{a}_{(1)}$:

$$\mathbf{a}_{(1)} = k_2\mathbf{a}_{(2)} + \cdots + k_m\mathbf{a}_{(m)} \qquad \text{where } k_j = -c_j/c_1.$$

(Some k_j's may be zero. Or even all of them, namely, if $\mathbf{a}_{(1)} = \mathbf{0}$.)

Why is linear independence important? Well, if a set of vectors is linearly dependent, then we can get rid of at least one or perhaps more of the vectors until we get a linearly independent set. This set is then the smallest "truly essential" set with which we can work. Thus, we cannot express any of the vectors, of this set, linearly in terms of the others.

7.4 pg.282-285---:)https://github.com/AvinashShrivastav/MATHEMATICS-FOR-COMPUTING

Apply $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$ to a vector and to a 4×3 matrix of your choice. Find $\mathbf{B} = \mathbf{E}_3\mathbf{E}_2\mathbf{E}_1\mathbf{A}$, where $\mathbf{A} = [a_{jk}]$ is the general 4×2 matrix. Is $\mathbf{B}$ equal to $\mathbf{C} = \mathbf{E}_1\mathbf{E}_2\mathbf{E}_3\mathbf{A}$?

(b) Conclude that $\mathbf{E}_1$, $\mathbf{E}_2$, $\mathbf{E}_3$ are obtained by doing the corresponding elementary operations on the 4×4 unit matrix. Prove that *if* $\mathbf{M}$ *is obtained from* $\mathbf{A}$ *by an elementary row operation, then*

$$\mathbf{M} = \mathbf{EA},$$

where $\mathbf{E}$ *is obtained from the* $n \times n$ *unit matrix* $\mathbf{I}_n$ *by the same row operation.*

7.4 Linear Independence. Rank of a Matrix. Vector Space

Since our next goal is to fully characterize the behavior of linear systems in terms of existence and uniqueness of solutions (Sec. 7.5), we have to introduce new fundamental linear algebraic concepts that will aid us in doing so. Foremost among these are ***linear independence*** and the ***rank of a matrix***. Keep in mind that these concepts are intimately linked with the important Gauss elimination method and how it works.

Linear Independence and Dependence of Vectors

Given any set of m vectors $\mathbf{a}_{(1)}, \cdots, \mathbf{a}_{(m)}$ (with the same number of components), a **linear combination** of these vectors is an expression of the form

$$c_1\mathbf{a}_{(1)} + c_2\mathbf{a}_{(2)} + \cdots + c_m\mathbf{a}_{(m)}$$

where $c_1, c_2, \cdots, c_m$ are any scalars. Now consider the equation

$$c_1\mathbf{a}_{(1)} + c_2\mathbf{a}_{(2)} + \cdots + c_m\mathbf{a}_{(m)} = \mathbf{0}. \tag{1}$$

Clearly, this vector equation (1) holds if we choose all c_j's zero, because then it becomes $\mathbf{0} = \mathbf{0}$. If this is the only m-tuple of scalars for which (1) holds, then our vectors $\mathbf{a}_{(1)}, \cdots, \mathbf{a}_{(m)}$ are said to form a *linearly independent set* or, more briefly, we call them **linearly independent**. Otherwise, if (1) also holds with scalars not all zero, we call these vectors **linearly dependent**. This means that we can express at least one of the vectors as a linear combination of the other vectors. For instance, if (1) holds with, say, $c_1 \neq 0$, we can solve (1) for $\mathbf{a}_{(1)}$:

$$\mathbf{a}_{(1)} = k_2\mathbf{a}_{(2)} + \cdots + k_m\mathbf{a}_{(m)} \qquad \text{where } k_j = -c_j/c_1.$$

(Some k_j's may be zero. Or even all of them, namely, if $\mathbf{a}_{(1)} = \mathbf{0}$.)

Why is linear independence important? Well, if a set of vectors is linearly dependent, then we can get rid of at least one or perhaps more of the vectors until we get a linearly independent set. This set is then the smallest "truly essential" set with which we can work. Thus, we cannot express any of the vectors, of this set, linearly in terms of the others.

EXAMPLE 1 **Linear Independence and Dependence**

The three vectors

$$\begin{aligned} \mathbf{a}_{(1)} &= [\ \ 3 \quad\ \ 0 \quad 2 \quad\ \ 2] \\ \mathbf{a}_{(2)} &= [-6 \quad 42 \quad 24 \quad 54] \\ \mathbf{a}_{(3)} &= [\ 21 \quad -21 \quad 0 \quad -15] \end{aligned}$$

are linearly dependent because

$$6\mathbf{a}_{(1)} - \tfrac{1}{2}\mathbf{a}_{(2)} - \mathbf{a}_{(3)} = \mathbf{0}.$$

Although this is easily checked by vector arithmetic (do it!), it is not so easy to discover. However, a systematic method for finding out about linear independence and dependence follows below.

The first two of the three vectors are linearly independent because $c_1\mathbf{a}_{(1)} + c_2\mathbf{a}_{(2)} = \mathbf{0}$ implies $c_2 = 0$ (from the second components) and then $c_1 = 0$ (from any other component of $\mathbf{a}_{(1)}$. ■

Rank of a Matrix

DEFINITION

The **rank** of a matrix **A** is the maximum number of linearly independent row vectors of **A**. It is denoted by rank **A**.

Our further discussion will show that the rank of a matrix is an important key concept for understanding general properties of matrices and linear systems of equations.

EXAMPLE 2 **Rank**

The matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 0 & 2 & 2 \\ -6 & 42 & 24 & 54 \\ 21 & -21 & 0 & -15 \end{bmatrix} \tag{2}$$

has rank 2, because Example 1 shows that the first two row vectors are linearly independent, whereas all three row vectors are linearly dependent.

Note further that rank $\mathbf{A} = 0$ if and only if $\mathbf{A} = \mathbf{0}$. This follows directly from the definition. ■

We call a matrix $\mathbf{A}_1$ **row-equivalent** to a matrix $\mathbf{A}_2$ if $\mathbf{A}_1$ can be obtained from $\mathbf{A}_2$ by (finitely many!) elementary row operations.

Now the maximum number of linearly independent row vectors of a matrix does not change if we change the order of rows or multiply a row by a nonzero c or take a linear combination by adding a multiple of a row to another row. This shows that rank is **invariant** under elementary row operations:

THEOREM 1 **Row-Equivalent Matrices**

Row-equivalent matrices have the same rank.

Hence we can determine the rank of a matrix by reducing the matrix to row-echelon form, as was done in Sec. 7.3. Once the matrix is in row-echelon form, we count the number of nonzero rows, which is precisely the rank of the matrix.

EXAMPLE 3 **Determination of Rank**

For the matrix in Example 2 we obtain successively

$$\mathbf{A} = \begin{bmatrix} 3 & 0 & 2 & 2 \\ -6 & 42 & 24 & 54 \\ 21 & -21 & 0 & -15 \end{bmatrix} \quad \text{(given)}$$

$$\begin{bmatrix} 3 & 0 & 2 & 2 \\ 0 & 42 & 28 & 58 \\ 0 & -21 & -14 & -29 \end{bmatrix} \quad \begin{matrix} \\ \text{Row 2} + 2 \text{ Row 1} \\ \text{Row 3} - 7 \text{ Row 1} \end{matrix}$$

$$\begin{bmatrix} 3 & 0 & 2 & 2 \\ 0 & 42 & 28 & 58 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \begin{matrix} \\ \\ \text{Row 3} + \frac{1}{2}\text{ Row 2.} \end{matrix}$$

The last matrix is in row-echelon form and has two nonzero rows. Hence rank $\mathbf{A} = 2$, as before. ■

Examples 1–3 illustrate the following useful theorem (with $p = 3$, $n = 3$, and the rank of the matrix $= 2$).

THEOREM 2 **Linear Independence and Dependence of Vectors**

Consider p vectors that each have n components. Then these vectors are linearly independent if the matrix formed, with these vectors as row vectors, has rank p. However, these vectors are linearly dependent if that matrix has rank less than p.

Further important properties will result from the basic

THEOREM 3 **Rank in Terms of Column Vectors**

The rank r of a matrix $\mathbf{A}$ *equals the maximum number of linearly independent* ***column*** *vectors of* $\mathbf{A}$.

Hence $\mathbf{A}$ *and its transpose* $\mathbf{A}^\mathsf{T}$ *have the same rank.*

PROOF In this proof we write simply "rows" and "columns" for row and column vectors. Let $\mathbf{A}$ be an $m \times n$ matrix of rank $\mathbf{A} = r$. Then by definition of rank, $\mathbf{A}$ has r linearly independent rows which we denote by $\mathbf{v}_{(1)}, \cdots, \mathbf{v}_{(r)}$ (regardless of their position in $\mathbf{A}$), and all the rows $\mathbf{a}_{(1)}, \cdots, \mathbf{a}_{(m)}$ of $\mathbf{A}$ are linear combinations of those, say,

$$\begin{aligned} \mathbf{a}_{(1)} &= c_{11}\mathbf{v}_{(1)} + c_{12}\mathbf{v}_{(2)} + \cdots + c_{1r}\mathbf{v}_{(r)} \\ \mathbf{a}_{(2)} &= c_{21}\mathbf{v}_{(1)} + c_{22}\mathbf{v}_{(2)} + \cdots + c_{2r}\mathbf{v}_{(r)} \\ &\;\;\vdots \\ \mathbf{a}_{(m)} &= c_{m1}\mathbf{v}_{(1)} + c_{m2}\mathbf{v}_{(2)} + \cdots + c_{mr}\mathbf{v}_{(r)}. \end{aligned} \tag{3}$$

These are vector equations for rows. To switch to columns, we write (3) in terms of components as n such systems, with $k = 1, \cdots, n$,

$$
(4) \qquad \begin{aligned}
a_{1k} &= c_{11}v_{1k} + c_{12}v_{2k} + \cdots + c_{1r}v_{rk} \\
a_{2k} &= c_{21}v_{1k} + c_{22}v_{2k} + \cdots + c_{2r}v_{rk} \\
&\;\;\vdots \\
a_{mk} &= c_{m1}v_{1k} + c_{m2}v_{2k} + \cdots + c_{mr}v_{rk}
\end{aligned}
$$

and collect components in columns. Indeed, we can write (4) as

$$
(5) \qquad \begin{bmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{bmatrix} = v_{1k} \begin{bmatrix} c_{11} \\ c_{21} \\ \vdots \\ c_{m1} \end{bmatrix} + v_{2k} \begin{bmatrix} c_{12} \\ c_{22} \\ \vdots \\ c_{m2} \end{bmatrix} + \cdots + v_{rk} \begin{bmatrix} c_{1r} \\ c_{2r} \\ \vdots \\ c_{mr} \end{bmatrix}
$$

where $k = 1, \cdots, n$. Now the vector on the left is the kth column vector of $\mathbf{A}$. We see that each of these n columns is a linear combination of the same r columns on the right. Hence $\mathbf{A}$ cannot have more linearly independent columns than rows, whose number is rank $\mathbf{A} = r$. Now rows of $\mathbf{A}$ are columns of the transpose $\mathbf{A}^\mathsf{T}$. For $\mathbf{A}^\mathsf{T}$ our conclusion is that $\mathbf{A}^\mathsf{T}$ cannot have more linearly independent columns than rows, so that $\mathbf{A}$ cannot have more linearly independent rows than columns. Together, the number of linearly independent columns of $\mathbf{A}$ must be r, the rank of $\mathbf{A}$. This completes the proof. ■

EXAMPLE 4 **Illustration of Theorem 3**

The matrix in (2) has rank 2. From Example 3 we see that the first two row vectors are linearly independent and by "working backward" we can verify that Row 3 $= 6$ Row 1 $- \frac{1}{2}$ Row 2. Similarly, the first two columns are linearly independent, and by reducing the last matrix in Example 3 by columns we find that

$$
\text{Column } 3 = \tfrac{2}{3}\,\text{Column } 1 + \tfrac{2}{3}\,\text{Column } 2 \qquad \text{and} \qquad \text{Column } 4 = \tfrac{2}{3}\,\text{Column } 1 + \tfrac{29}{21}\,\text{Column } 2.
$$
■

Combining Theorems 2 and 3 we obtain

THEOREM 4 **Linear Dependence of Vectors**

Consider p vectors each having n components. If $n < p$, then these vectors are linearly dependent.

PROOF The matrix $\mathbf{A}$ with those p vectors as row vectors has p rows and $n < p$ columns; hence by Theorem 3 it has rank $\mathbf{A} \leqq n < p$, which implies linear dependence by Theorem 2. ■

Vector Space

The following related concepts are of general interest in linear algebra. In the present context they provide a clarification of essential properties of matrices and their role in connection with linear systems.

27–35 VECTOR SPACE

Is the given set of vectors a vector space? Give reasons. If your answer is yes, determine the dimension and find a basis. ($v_1, v_2, \cdots$ denote components.)

27. All vectors in R^3 with $v_1 - v_2 + 2v_3 = 0$

28. All vectors in R^3 with $3v_2 + v_3 = k$

29. All vectors in R^2 with $v_1 \geqq v_2$

30. All vectors in R^n with the first $n - 2$ components zero

31. All vectors in R^5 with positive components

32. All vectors in R^3 with $3v_1 - 2v_2 + v_3 = 0$, $4v_1 + 5v_2 = 0$

33. All vectors in R^3 with $3v_1 - v_3 = 0$, $2v_1 + 3v_2 - 4v_3 = 0$

34. All vectors in R^n with $|v_j| = 1$ for $j = 1, \cdots, n$

35. All vectors in R^4 with $v_1 = 2v_2 = 3v_3 = 4v_4$

7.5 Solutions of Linear Systems: Existence, Uniqueness

Rank, as just defined, gives complete information about existence, uniqueness, and general structure of the solution set of linear systems as follows.

A linear system of equations in n unknowns has a unique solution if the coefficient matrix and the augmented matrix have the same rank n, and infinitely many solutions if that common rank is less than n. The system has no solution if those two matrices have different rank.

To state this precisely and prove it, we shall use the generally important concept of a **submatrix** of $\mathbf{A}$. By this we mean any matrix obtained from $\mathbf{A}$ by omitting some rows or columns (or both). By definition this includes $\mathbf{A}$ itself (as the matrix obtained by omitting no rows or columns); this is practical.

THEOREM 1

Fundamental Theorem for Linear Systems

(a) Existence. *A linear system of m equations in n unknowns* $x_1, \cdots, x_n$

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots &\cdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m
\end{aligned}
\tag{1}
$$

is **consistent**, *that is, has solutions, if and only if the coefficient matrix* $\mathbf{A}$ *and the augmented matrix* $\tilde{\mathbf{A}}$ *have the same rank. Here,*

$$
\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \cdot & \cdots & \cdot \\ \cdot & \cdots & \cdot \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \quad \textit{and} \quad \tilde{\mathbf{A}} = \left[\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdots & \cdot & \cdot \\ a_{m1} & \cdots & a_{mn} & b_m \end{array}\right]
$$

(b) Uniqueness. *The system* (1) *has precisely one solution if and only if this common rank r of* $\mathbf{A}$ *and* $\tilde{\mathbf{A}}$ *equals n.*

(c) Infinitely many solutions. *If this common rank r is less than n, the system* (1) *has infinitely many solutions. All of these solutions are obtained by determining r suitable unknowns* (*whose submatrix of coefficients must have rank r*) *in terms of the remaining $n - r$ unknowns, to which arbitrary values can be assigned.* (See Example 3 in Sec. 7.3.)

(d) Gauss elimination (Sec. 7.3). *If solutions exist, they can all be obtained by the Gauss elimination.* (This method will automatically reveal whether or not solutions exist; see Sec. 7.3.)

PROOF **(a)** We can write the system (1) in vector form $\mathbf{Ax} = \mathbf{b}$ or in terms of column vectors $\mathbf{c}_{(1)}, \cdots, \mathbf{c}_{(n)}$ of $\mathbf{A}$:

$$\mathbf{c}_{(1)}x_1 + \mathbf{c}_{(2)}x_2 + \cdots + \mathbf{c}_{(n)}x_n = \mathbf{b}. \tag{2}$$

$\widetilde{\mathbf{A}}$ is obtained by augmenting $\mathbf{A}$ by a single column $\mathbf{b}$. Hence, by Theorem 3 in Sec. 7.4, rank $\widetilde{\mathbf{A}}$ equals rank $\mathbf{A}$ or rank $\mathbf{A} + 1$. Now if (1) has a solution $\mathbf{x}$, then (2) shows that $\mathbf{b}$ must be a linear combination of those column vectors, so that $\widetilde{\mathbf{A}}$ and $\mathbf{A}$ have the same maximum number of linearly independent column vectors and thus the same rank.

Conversely, if rank $\widetilde{\mathbf{A}} =$ rank $\mathbf{A}$, then $\mathbf{b}$ must be a linear combination of the column vectors of $\mathbf{A}$, say,

$$\mathbf{b} = \alpha_1\mathbf{c}_{(1)} + \cdots + \alpha_n\mathbf{c}_{(n)} \tag{2*}$$

since otherwise rank $\widetilde{\mathbf{A}} =$ rank $\mathbf{A} + 1$. But (2*) means that (1) has a solution, namely, $x_1 = \alpha_1, \cdots, x_n = \alpha_n$, as can be seen by comparing (2*) and (2).

(b) If rank $\mathbf{A} = n$, the n column vectors in (2) are linearly independent by Theorem 3 in Sec. 7.4. We claim that then the representation (2) of $\mathbf{b}$ is unique because otherwise

$$\mathbf{c}_{(1)}x_1 + \cdots + \mathbf{c}_{(n)}x_n = \mathbf{c}_{(1)}\tilde{x}_1 + \cdots + \mathbf{c}_{(n)}\tilde{x}_n.$$

This would imply (take all terms to the left, with a minus sign)

$$(x_1 - \tilde{x}_1)\mathbf{c}_{(1)} + \cdots + (x_n - \tilde{x}_n)\mathbf{c}_{(n)} = \mathbf{0}$$

and $x_1 - \tilde{x}_1 = 0, \cdots, x_n - \tilde{x}_n = 0$ by linear independence. But this means that the scalars $x_1, \cdots, x_n$ in (2) are uniquely determined, that is, the solution of (1) is unique.

(c) If rank $\mathbf{A} =$ rank $\widetilde{\mathbf{A}} = r < n$, then by Theorem 3 in Sec. 7.4 there is a linearly independent set K of r column vectors of $\mathbf{A}$ such that the other $n - r$ column vectors of $\mathbf{A}$ are linear combinations of those vectors. We renumber the columns and unknowns, denoting the renumbered quantities by $\hat{}$, so that $\{\hat{\mathbf{c}}_{(1)}, \cdots, \hat{\mathbf{c}}_{(r)}\}$ is that linearly independent set K. Then (2) becomes

$$\hat{\mathbf{c}}_{(1)}\hat{x}_1 + \cdots + \hat{\mathbf{c}}_{(r)}\hat{x}_r + \hat{\mathbf{c}}_{(r+1)}\hat{x}_{r+1} + \cdots + \hat{\mathbf{c}}_{(n)}\hat{x}_n = \mathbf{b},$$

$\hat{\mathbf{c}}_{(r+1)}, \cdots, \hat{\mathbf{c}}_{(n)}$ are linear combinations of the vectors of K, and so are the vectors $\hat{x}_{r+1}\hat{\mathbf{c}}_{(r+1)}, \cdots, \hat{x}_n\hat{\mathbf{c}}_{(n)}$. Expressing these vectors in terms of the vectors of K and collecting terms, we can thus write the system in the form

$$\hat{\mathbf{c}}_{(1)}y_1 + \cdots + \hat{\mathbf{c}}_{(r)}y_r = \mathbf{b} \tag{3}$$

with $y_j = \hat{x}_j + \beta_j$, where β_j results from the $n - r$ terms $\hat{\mathbf{c}}_{(r+1)}\hat{x}_{r+1}, \cdots, \hat{\mathbf{c}}_{(n)}\hat{x}_n$; here, $j = 1, \cdots, r$. Since the system has a solution, there are $y_1, \cdots, y_r$ satisfying (3). These scalars are unique since K is linearly independent. Choosing $\hat{x}_{r+1}, \cdots, \hat{x}_n$ fixes the β_j and corresponding $\hat{x}_j = y_j - \beta_j$, where $j = 1, \cdots, r$.

(d) This was discussed in Sec. 7.3 and is restated here as a reminder. ■

The theorem is illustrated in Sec. 7.3. In Example 2 there is a unique solution since rank $\tilde{\mathbf{A}}$ = rank $\mathbf{A}$ $= n = 3$ (as can be seen from the last matrix in the example). In Example 3 we have rank $\tilde{\mathbf{A}}$ = rank $\mathbf{A}$ $= 2 < n = 4$ and can choose x_3 and x_4 arbitrarily. In Example 4 there is no solution because rank $\mathbf{A} = 2 <$ rank $\tilde{\mathbf{A}} = 3$.

Homogeneous Linear System

Recall from Sec. 7.3 that a linear system (1) is called **homogeneous** if all the b_j's are zero, and **nonhomogeneous** if one or several b_j's are not zero. For the homogeneous system we obtain from the Fundamental Theorem the following results.

THEOREM 2

Homogeneous Linear System

A homogeneous linear system

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0
\end{aligned}
\tag{4}
$$

always has the **trivial solution** $x_1 = 0, \cdots, x_n = 0$. *Nontrivial solutions exist if and only if rank* $\mathbf{A} < n$. *If* rank $\mathbf{A} = r < n$, *these solutions, together with* $\mathbf{x} = \mathbf{0}$, *form a vector space* (see Sec. 7.4) *of dimension* $n - r$ called the **solution space** of (4).

In particular, if $\mathbf{x}_{(1)}$ *and* $\mathbf{x}_{(2)}$ *are solution vectors of* (4), *then* $\mathbf{x} = c_1\mathbf{x}_{(1)} + c_2\mathbf{x}_{(2)}$ *with any scalars* c_1 *and* c_2 *is a solution vector of* (4). (This ***does not hold*** for nonhomogeneous systems. Also, the term *solution space is used for homogeneous* systems only.)

PROOF The first proposition can be seen directly from the system. It agrees with the fact that $\mathbf{b} = \mathbf{0}$ implies that rank $\tilde{\mathbf{A}}$ = rank $\mathbf{A}$, so that a homogeneous system is always ***consistent***. If rank $\mathbf{A} = n$, the trivial solution is the unique solution according to (b) in Theorem 1. If rank $\mathbf{A} < n$, there are nontrivial solutions according to (c) in Theorem 1. The solutions form a vector space because if $\mathbf{x}_{(1)}$ and $\mathbf{x}_{(2)}$ are any of them, then $\mathbf{A}\mathbf{x}_{(1)} = \mathbf{0}$, $\mathbf{A}\mathbf{x}_{(2)} = \mathbf{0}$, and this implies $\mathbf{A}(\mathbf{x}_{(1)} + \mathbf{x}_{(2)}) = \mathbf{A}\mathbf{x}_{(1)} + \mathbf{A}\mathbf{x}_{(2)} = \mathbf{0}$ as well as $\mathbf{A}(c\mathbf{x}_{(1)}) = c\mathbf{A}\mathbf{x}_{(1)} = \mathbf{0}$, where c is arbitrary. If rank $\mathbf{A} = r < n$, Theorem 1 (c) implies that we can choose $n - r$ suitable unknowns, call them $x_{r+1}, \cdots, x_n$, in an arbitrary fashion, and every solution is obtained in this way. Hence a basis for the solution space, briefly called a **basis of solutions** of (4), is $\mathbf{y}_{(1)}, \cdots, \mathbf{y}_{(n-r)}$, where the basis vector $\mathbf{y}_{(j)}$ is obtained by choosing $x_{r+j} = 1$ and the other $x_{r+1}, \cdots, x_n$ zero; the corresponding first r components of this solution vector are then determined. Thus the solution space of (4) has dimension $n - r$. This proves Theorem 2. ■

The solution space of (4) is also called the **null space** of $\mathbf{A}$ because $\mathbf{Ax} = \mathbf{0}$ for every $\mathbf{x}$ in the solution space of (4). Its dimension is called the **nullity** of $\mathbf{A}$. Hence Theorem 2 states that

$$\text{rank}\,\mathbf{A} + \text{nullity}\,\mathbf{A} = n \tag{5}$$

where n is the number of unknowns (number of columns of $\mathbf{A}$).

Furthermore, by the definition of rank we have rank $\mathbf{A} \leqq m$ in (4). Hence if $m < n$, then rank $\mathbf{A} < n$. By Theorem 2 this gives the practically important

THEOREM 3

Homogeneous Linear System with Fewer Equations Than Unknowns

A homogeneous linear system with fewer equations than unknowns always has nontrivial solutions.

Nonhomogeneous Linear Systems

The characterization of all solutions of the linear system (1) is now quite simple, as follows.

THEOREM 4

Nonhomogeneous Linear System

If a nonhomogeneous linear system (1) *is consistent, then all of its solutions are obtained as*

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{x}_h \tag{6}$$

where $\mathbf{x}_0$ *is any* (fixed) *solution of* (1) *and* $\mathbf{x}_h$ *runs through all the solutions of the corresponding homogeneous system* (4).

PROOF The difference $\mathbf{x}_h = \mathbf{x} - \mathbf{x}_0$ of any two solutions of (1) is a solution of (4) because $\mathbf{Ax}_h = \mathbf{A}(\mathbf{x} - \mathbf{x}_0) = \mathbf{Ax} - \mathbf{Ax}_0 = \mathbf{b} - \mathbf{b} = \mathbf{0}$. Since $\mathbf{x}$ is any solution of (1), we get all the solutions of (1) if in (6) we take any solution $\mathbf{x}_0$ of (1) and let $\mathbf{x}_h$ vary throughout the solution space of (4). ■

This covers a main part of our discussion of characterizing the solutions of systems of linear equations. Our next main topic is determinants and their role in linear equations.

7.6 For Reference: Second- and Third-Order Determinants

We created this section as a quick general reference section on second- and third-order determinants. It is completely independent of the theory in Sec. 7.7 and suffices as a reference for many of our examples and problems. Since this section is for reference, ***go on to the next section, consulting this material only when needed***.

A **determinant of second order** is denoted and defined by

$$D = \det \mathbf{A} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}. \tag{1}$$

So here we have ***bars*** (whereas a matrix has ***brackets***).

Note the following. The signs on the right are $+ - +$. Each of the three terms on the right is an entry in the first column of D times its **minor**, that is, the second-order determinant obtained from D by deleting the row and column of that entry; thus, for a_{11} delete the first row and first column, and so on.

If we write out the minors in (4), we obtain

$$D = a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{21}a_{13}a_{32} - a_{21}a_{12}a_{33} + a_{31}a_{12}a_{23} - a_{31}a_{13}a_{22}. \tag{4*}$$

Cramer's Rule for Linear Systems of Three Equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= b_3 \end{aligned} \tag{5}$$

is

$$x_1 = \frac{D_1}{D}, \quad x_2 = \frac{D_2}{D}, \quad x_3 = \frac{D_3}{D} \qquad (D \neq 0) \tag{6}$$

with the *determinant D of the system* given by (4) and

$$D_1 = \begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix}, \quad D_2 = \begin{vmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{vmatrix}, \quad D_3 = \begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{vmatrix}.$$

Note that D_1, D_2, D_3 are obtained by replacing Columns 1, 2, 3, respectively, by the column of the right sides of (5).

Cramer's rule (6) can be derived by eliminations similar to those for (3), but it also follows from the general case (Theorem 4) in the next section.

7.7 Determinants. Cramer's Rule

Determinants were originally introduced for solving linear systems. Although ***impractical in computations***, they have important engineering applications in eigenvalue problems (Sec. 8.1), differential equations, vector algebra (Sec. 9.3), and in other areas. They can be introduced in several equivalent ways. Our definition is particularly for dealing with linear systems.

A **determinant of order** n is a scalar associated with an $n \times n$ (hence ***square!***) matrix $\mathbf{A} = [a_{jk}]$, and is denoted by

$$D = \det \mathbf{A} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}. \tag{1}$$

For $n = 1$, this determinant is defined by

$$D = a_{11}. \tag{2}$$

For $n \geqq 2$ by

$$D = a_{j1}C_{j1} + a_{j2}C_{j2} + \cdots + a_{jn}C_{jn} \qquad (j = 1, 2, \cdots, \text{or } n) \tag{3a}$$

or

$$D = a_{1k}C_{1k} + a_{2k}C_{2k} + \cdots + a_{nk}C_{nk} \quad (k = 1, 2, \cdots, \text{or } n). \tag{3b}$$

Here,

$$C_{jk} = (-1)^{j+k}M_{jk}$$

and M_{jk} is a determinant of order $n - 1$, namely, the determinant of the submatrix of **A** obtained from **A** by omitting the row and column of the entry a_{jk}, that is, the jth row and the kth column.

In this way, D is defined in terms of n determinants of order $n - 1$, each of which is, in turn, defined in terms of $n - 1$ determinants of order $n - 2$, and so on—until we finally arrive at second-order determinants, in which those submatrices consist of single entries whose determinant is defined to be the entry itself.

From the definition it follows that *we may **expand** D by any row or column,* that is, choose in (3) the entries in any row or column, similarly when expanding the C_{jk}'s in (3), and so on.

This definition is unambiguous, that is, it yields the same value for D no matter which columns or rows we choose in expanding. A proof is given in App. 4.

Terms used in connection with determinants are taken from matrices. In D we have n^2 **entries** a_{jk}, also n **rows** and n **columns**, and a **main diagonal** on which $a_{11}, a_{22}, \cdots, a_{nn}$ stand. Two terms are new:

M_{jk} is called the **minor** *of* a_{jk} *in* D, and C_{jk} the **cofactor** *of* a_{jk} *in* D.

For later use we note that (3) may also be written in terms of minors

$$D = \sum_{k=1}^{n} (-1)^{j+k} a_{jk} M_{jk} \qquad (j = 1, 2, \cdots, \text{or } n) \tag{4a}$$

$$D = \sum_{j=1}^{n} (-1)^{j+k} a_{jk} M_{jk} \qquad (k = 1, 2, \cdots, \text{or } n). \tag{4b}$$

EXAMPLE 1 Minors and Cofactors of a Third-Order Determinant

In (4) of the previous section the minors and cofactors of the entries in the first column can be seen directly. For the entries in the second row the minors are

$$M_{21} = \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix}, \qquad M_{22} = \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}, \qquad M_{23} = \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}$$

and the cofactors are $C_{21} = -M_{21}$, $C_{22} = +M_{22}$, and $C_{23} = -M_{23}$. Similarly for the third row—write these down yourself. And verify that the signs in C_{jk} form a **checkerboard pattern**

$$\begin{matrix} + & - & + \\ - & + & - \\ + & - & + \end{matrix}$$

■

EXAMPLE 2 **Expansions of a Third-Order Determinant**

$$D = \begin{vmatrix} 1 & 3 & 0 \\ 2 & 6 & 4 \\ -1 & 0 & 2 \end{vmatrix} = 1\begin{vmatrix} 6 & 4 \\ 0 & 2 \end{vmatrix} - 3\begin{vmatrix} 2 & 4 \\ -1 & 2 \end{vmatrix} + 0\begin{vmatrix} 2 & 6 \\ -1 & 0 \end{vmatrix}$$

$$= 1(12 - 0) - 3(4 + 4) + 0(0 + 6) = -12.$$

This is the expansion by the first row. The expansion by the third column is

$$D = 0\begin{vmatrix} 2 & 6 \\ -1 & 0 \end{vmatrix} - 4\begin{vmatrix} 1 & 3 \\ -1 & 0 \end{vmatrix} + 2\begin{vmatrix} 1 & 3 \\ 2 & 6 \end{vmatrix} = 0 - 12 + 0 = -12.$$

Verify that the other four expansions also give the value -12. ■

EXAMPLE 3 **Determinant of a Triangular Matrix**

$$\begin{vmatrix} -3 & 0 & 0 \\ 6 & 4 & 0 \\ -1 & 2 & 5 \end{vmatrix} = -3\begin{vmatrix} 4 & 0 \\ 2 & 5 \end{vmatrix} = -3 \cdot 4 \cdot 5 = -60.$$

Inspired by this, can you formulate a little theorem on determinants of triangular matrices? Of diagonal matrices? ■

General Properties of Determinants

There is an attractive way of finding determinants (1) that consists of applying elementary row operations to (1). By doing so we obtain an "upper triangular" determinant (see Sec. 7.1, for definition with "matrix" replaced by "determinant") whose value is then very easy to compute, being just the product of its diagonal entries. This approach is ***similar*** (***but not the same!***) to what we did to matrices in Sec. 7.3. *In particular, be aware that interchanging two rows in a determinant introduces a multiplicative factor of* -1 *to the value of the determinant*! Details are as follows.

THEOREM 1 **Behavior of an *n*th-Order Determinant under Elementary Row Operations**

(a) *Interchange of two rows multiplies the value of the determinant by* -1.

(b) *Addition of a multiple of a row to another row does not alter the value of the determinant.*

(c) *Multiplication of a row by a nonzero constant c multiplies the value of the determinant by c.* (This holds also when $c = 0$, but no longer gives an elementary row operation.)

PROOF **(a)** By induction. The statement holds for $n = 2$ because

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc, \qquad \text{but} \qquad \begin{vmatrix} c & d \\ a & b \end{vmatrix} = bc - ad.$$

7.8 Pg 301-304---:)https://github.com/AvinashShrivastav/MATHEMATICS-FOR-COMPUTING

(a) Line through two points. Derive from $D = 0$ in (12) the familiar formula

$$\frac{x - x_1}{x_1 - x_2} = \frac{y - y_1}{y_1 - y_2}.$$

(b) Plane. Find the analog of (12) for a plane through three given points. Apply it when the points are (1, 1, 1), (3, 2, 6), (5, 0, 5).

(c) Circle. Find a similar formula for a circle in the plane through three given points. Find and sketch the circle through (2, 6), (6, 4), (7, 1).

(d) Sphere. Find the analog of the formula in (c) for a sphere through four given points. Find the sphere through (0, 0, 5), (4, 0, 1), (0, 4, 1), (0, 0, −3) by this formula or by inspection.

(e) General conic section. Find a formula for a general conic section (the vanishing of a determinant of 6th order). Try it out for a quadratic parabola and for a more general conic section of your own choice.

21–25 CRAMER'S RULE

Solve by Cramer's rule. Check by Gauss elimination and back substitution. Show details.

21. $\begin{aligned} 3x - 5y &= 15.5 \\ 6x + 16y &= 5.0 \end{aligned}$

22. $\begin{aligned} 2x - 4y &= -24 \\ 5x + 2y &= 0 \end{aligned}$

23. $\begin{aligned} 3y - 4z &= 16 \\ 2x - 5y + 7z &= -27 \\ -x \qquad - 9z &= 9 \end{aligned}$

24. $\begin{aligned} 3x - 2y + z &= 13 \\ -2x + y + 4z &= 11 \\ x + 4y - 5z &= -31 \end{aligned}$

25. $\begin{aligned} -4w + x + y \qquad &= -10 \\ w - 4x \qquad + z &= 1 \\ w \qquad - 4y + z &= -7 \\ x + y - 4z &= 10 \end{aligned}$

7.8 Inverse of a Matrix. Gauss–Jordan Elimination

***In this section we consider* square *matrices exclusively*.**

The **inverse** of an $n \times n$ matrix $\mathbf{A} = [a_{jk}]$ is denoted by $\mathbf{A}^{-1}$ and is an $n \times n$ matrix such that

$$\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \tag{1}$$

where $\mathbf{I}$ is the $n \times n$ unit matrix (see Sec. 7.2).

If $\mathbf{A}$ has an inverse, then $\mathbf{A}$ is called a **nonsingular matrix**. If $\mathbf{A}$ has no inverse, then $\mathbf{A}$ is called a **singular matrix**.

If $\mathbf{A}$ *has an inverse, the inverse is unique.*

Indeed, if both $\mathbf{B}$ and $\mathbf{C}$ are inverses of $\mathbf{A}$, then $\mathbf{AB} = \mathbf{I}$ and $\mathbf{CA} = \mathbf{I}$, so that we obtain the uniqueness from

$$\mathbf{B} = \mathbf{IB} = (\mathbf{CA})\mathbf{B} = \mathbf{C}(\mathbf{AB}) = \mathbf{CI} = \mathbf{C}.$$

We prove next that $\mathbf{A}$ has an inverse (is nonsingular) if and only if it has maximum possible rank n. The proof will also show that $\mathbf{Ax} = \mathbf{b}$ implies $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ provided $\mathbf{A}^{-1}$ exists, and will thus give a motivation for the inverse as well as a relation to linear systems. (But this will ***not*** give a good method of solving $\mathbf{Ax} = \mathbf{b}$ ***numerically*** because the Gauss elimination in Sec. 7.3 requires fewer computations.)

THEOREM 1

Existence of the Inverse

The inverse $\mathbf{A}^{-1}$ *of an* $n \times n$ *matrix* $\mathbf{A}$ *exists if and only if* rank $\mathbf{A} = n$, *thus* (by Theorem 3, Sec. 7.7) *if and only if* det $\mathbf{A} \neq 0$. *Hence* $\mathbf{A}$ *is nonsingular if* rank $\mathbf{A} = n$, *and is singular if* rank $\mathbf{A} < n$.

PROOF Let $\mathbf{A}$ be a given $n \times n$ matrix and consider the linear system

$$\mathbf{Ax} = \mathbf{b}. \tag{2}$$

If the inverse $\mathbf{A}^{-1}$ exists, then multiplication from the left on both sides and use of (1) gives

$$\mathbf{A}^{-1}\mathbf{Ax} = \mathbf{x} = \mathbf{A}^{-1}\mathbf{b}.$$

This shows that (2) has a solution $\mathbf{x}$, which is unique because, for another solution $\mathbf{u}$, we have $\mathbf{Au} = \mathbf{b}$, so that $\mathbf{u} = \mathbf{A}^{-1}\mathbf{b} = \mathbf{x}$. Hence $\mathbf{A}$ must have rank n by the Fundamental Theorem in Sec. 7.5.

Conversely, let rank $\mathbf{A} = n$. Then by the same theorem, the system (2) has a unique solution $\mathbf{x}$ for any $\mathbf{b}$. Now the back substitution following the Gauss elimination (Sec. 7.3) shows that the components x_j of $\mathbf{x}$ are linear combinations of those of $\mathbf{b}$. Hence we can write

$$\mathbf{x} = \mathbf{Bb} \tag{3}$$

with $\mathbf{B}$ to be determined. Substitution into (2) gives

$$\mathbf{Ax} = \mathbf{A}(\mathbf{Bb}) = (\mathbf{AB})\mathbf{b} = \mathbf{Cb} = \mathbf{b} \qquad (\mathbf{C} = \mathbf{AB})$$

for any $\mathbf{b}$. Hence $\mathbf{C} = \mathbf{AB} = \mathbf{I}$, the unit matrix. Similarly, if we substitute (2) into (3) we get

$$\mathbf{x} = \mathbf{Bb} = \mathbf{B}(\mathbf{Ax}) = (\mathbf{BA})\mathbf{x}$$

for any $\mathbf{x}$ (and $\mathbf{b} = \mathbf{Ax}$). Hence $\mathbf{BA} = \mathbf{I}$. Together, $\mathbf{B} = \mathbf{A}^{-1}$ exists. ■

Determination of the Inverse by the Gauss–Jordan Method

To actually determine the inverse $\mathbf{A}^{-1}$ of a nonsingular $n \times n$ matrix $\mathbf{A}$, we can use a variant of the Gauss elimination (Sec. 7.3), called the **Gauss–Jordan elimination**.[3] The idea of the method is as follows.

Using $\mathbf{A}$, we form n linear systems

$$\mathbf{Ax}_{(1)} = \mathbf{e}_{(1)}, \quad \cdots, \quad \mathbf{Ax}_{(n)} = \mathbf{e}_{(n)}$$

where the vectors $\mathbf{e}_{(1)}, \cdots, \mathbf{e}_{(n)}$ are the columns of the $n \times n$ unit matrix $\mathbf{I}$; thus, $\mathbf{e}_{(1)} = [1 \quad 0 \quad \cdots \quad 0]^\mathsf{T}$, $\mathbf{e}_{(2)} = [0 \quad 1 \quad 0 \quad \cdots \quad 0]^\mathsf{T}$, etc. These are n vector equations in the unknown vectors $\mathbf{x}_{(1)}, \cdots, \mathbf{x}_{(n)}$. We combine them into a single matrix equation

[3]WILHELM JORDAN (1842–1899), German geodesist and mathematician. He did important geodesic work in Africa, where he surveyed oases. [See Althoen, S.C. and R. McLaughlin, Gauss–Jordan reduction: A brief history. *American Mathematical Monthly*, Vol. **94**, No. 2 (1987), pp. 130–142.]

We do ***not recommend*** it as a method for solving systems of linear equations, since the number of operations in addition to those of the Gauss elimination is larger than that for back substitution, which the Gauss–Jordan elimination avoids. See also Sec. 20.1.

$\mathbf{AX} = \mathbf{I}$, with the unknown matrix $\mathbf{X}$ having the columns $\mathbf{x}_{(1)}, \cdots, \mathbf{x}_{(n)}$. Correspondingly, we combine the n augmented matrices $[\mathbf{A} \quad \mathbf{e}_{(1)}], \cdots, [\mathbf{A} \quad \mathbf{e}_{(n)}]$ into one wide $n \times 2n$ "augmented matrix" $\widetilde{\mathbf{A}} = [\mathbf{A} \quad \mathbf{I}]$. Now multiplication of $\mathbf{AX} = \mathbf{I}$ by $\mathbf{A}^{-1}$ from the left gives $\mathbf{X} = \mathbf{A}^{-1}\mathbf{I} = \mathbf{A}^{-1}$. Hence, to solve $\mathbf{AX} = \mathbf{I}$ for $\mathbf{X}$, we can apply the Gauss elimination to $\widetilde{\mathbf{A}} = [\mathbf{A} \quad \mathbf{I}]$. This gives a matrix of the form $[\mathbf{U} \quad \mathbf{H}]$ with upper triangular $\mathbf{U}$ because the Gauss elimination triangularizes systems. The Gauss–Jordan method reduces $\mathbf{U}$ by further elementary row operations to diagonal form, in fact to the unit matrix $\mathbf{I}$. This is done by eliminating the entries of $\mathbf{U}$ above the main diagonal and making the diagonal entries all 1 by multiplication (see Example 1). Of course, the method operates on the entire matrix $[\mathbf{U} \quad \mathbf{H}]$, transforming $\mathbf{H}$ into some matrix $\mathbf{K}$, hence the entire $[\mathbf{U} \quad \mathbf{H}]$ to $[\mathbf{I} \quad \mathbf{K}]$. This is the "augmented matrix" of $\mathbf{IX} = \mathbf{K}$. Now $\mathbf{IX} = \mathbf{X} = \mathbf{A}^{-1}$, as shown before. By comparison, $\mathbf{K} = \mathbf{A}^{-1}$, so that we can read $\mathbf{A}^{-1}$ directly from $[\mathbf{I} \quad \mathbf{K}]$.

The following example illustrates the practical details of the method.

EXAMPLE 1 Finding the Inverse of a Matrix by Gauss–Jordan Elimination

Determine the inverse $\mathbf{A}^{-1}$ of

$$\mathbf{A} = \begin{bmatrix} -1 & 1 & 2 \\ 3 & -1 & 1 \\ -1 & 3 & 4 \end{bmatrix}.$$

Solution. We apply the Gauss elimination (Sec. 7.3) to the following $n \times 2n = 3 \times 6$ matrix, where BLUE always refers to the previous matrix.

$$[\mathbf{A} \quad \mathbf{I}] = \left[\begin{array}{rrr|rrr} -1 & 1 & 2 & 1 & 0 & 0 \\ 3 & -1 & 1 & 0 & 1 & 0 \\ -1 & 3 & 4 & 0 & 0 & 1 \end{array}\right]$$

$$\left[\begin{array}{rrr|rrr} -1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 7 & 3 & 1 & 0 \\ 0 & 2 & 2 & -1 & 0 & 1 \end{array}\right] \begin{array}{l} \\ \text{Row 2} + 3 \text{ Row 1} \\ \text{Row 3} - \text{Row 1} \end{array}$$

$$\left[\begin{array}{rrr|rrr} -1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 2 & 7 & 3 & 1 & 0 \\ 0 & 0 & -5 & -4 & -1 & 1 \end{array}\right] \begin{array}{l} \\ \\ \text{Row 3} - \text{Row 2} \end{array}$$

This is $[\mathbf{U} \quad \mathbf{H}]$ as produced by the Gauss elimination. Now follow the additional Gauss–Jordan steps, reducing $\mathbf{U}$ to $\mathbf{I}$, that is, to diagonal form with entries 1 on the main diagonal.

$$\left[\begin{array}{rrr|rrr} 1 & -1 & -2 & -1 & 0 & 0 \\ 0 & 1 & 3.5 & 1.5 & 0.5 & 0 \\ 0 & 0 & 1 & 0.8 & 0.2 & -0.2 \end{array}\right] \begin{array}{l} -\text{Row 1} \\ 0.5 \text{ Row 2} \\ -0.2 \text{ Row 3} \end{array}$$

$$\left[\begin{array}{rrr|rrr} 1 & -1 & 0 & 0.6 & 0.4 & -0.4 \\ 0 & 1 & 0 & -1.3 & -0.2 & 0.7 \\ 0 & 0 & 1 & 0.8 & 0.2 & -0.2 \end{array}\right] \begin{array}{l} \text{Row 1} + 2 \text{ Row 3} \\ \text{Row 2} - 3.5 \text{ Row 3} \\ \\ \end{array}$$

$$\left[\begin{array}{rrr|rrr} 1 & 0 & 0 & -0.7 & 0.2 & 0.3 \\ 0 & 1 & 0 & -1.3 & -0.2 & 0.7 \\ 0 & 0 & 1 & 0.8 & 0.2 & -0.2 \end{array}\right] \begin{array}{l} \text{Row 1} + \text{Row 2} \\ \\ \\ \end{array}$$

The last three columns constitute $\mathbf{A}^{-1}$. Check:

$$\begin{bmatrix} -1 & 1 & 2 \\ 3 & -1 & 1 \\ -1 & 3 & 4 \end{bmatrix} \begin{bmatrix} -0.7 & 0.2 & 0.3 \\ -1.3 & -0.2 & 0.7 \\ 0.8 & 0.2 & -0.2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Hence $\mathbf{AA}^{-1} = \mathbf{I}$. Similarly, $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. ■

Formulas for Inverses

Since finding the inverse of a matrix is really a problem of solving a system of linear equations, it is not surprising that Cramer's rule (Theorem 4, Sec. 7.7) might come into play. And similarly, as Cramer's rule was useful for theoretical study but not for computation, so too is the explicit formula (4) in the following theorem useful for theoretical considerations but not recommended for actually determining inverse matrices, except for the frequently occurring 2×2 case as given in (4*).

THEOREM 2

Inverse of a Matrix by Determinants

The inverse of a nonsingular $n \times n$ matrix $\mathbf{A} = [a_{jk}]$ is given by

$$\mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}}[C_{jk}]^{\mathsf{T}} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \cdot & \cdot & \cdots & \cdot \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix}, \tag{4}$$

where C_{jk} is the cofactor of a_{jk} in $\det \mathbf{A}$ (see Sec. 7.7). (CAUTION! Note well that in $\mathbf{A}^{-1}$, the cofactor C_{jk} occupies the same place as a_{kj} (not a_{jk}) does in $\mathbf{A}$.)

In particular, the inverse of

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \textit{is} \quad \mathbf{A}^{-1} = \frac{1}{\det \mathbf{A}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}. \tag{4*}$$

PROOF We denote the right side of (4) by $\mathbf{B}$ and show that $\mathbf{BA} = \mathbf{I}$. We first write

$$\mathbf{BA} = \mathbf{G} = [g_{kl}] \tag{5}$$

and then show that $\mathbf{G} = \mathbf{I}$. Now by the definition of matrix multiplication and because of the form of $\mathbf{B}$ in (4), we obtain (CAUTION! C_{sk}, not C_{ks})

$$g_{kl} = \sum_{s=1}^{n} \frac{C_{sk}}{\det \mathbf{A}} a_{sl} = \frac{1}{\det \mathbf{A}}(a_{1l}C_{1k} + \cdots + a_{nl}C_{nk}). \tag{6}$$

Now (9) and (10) in Sec. 7.7 show that the sum $(\cdots)$ on the right is $D = \det \mathbf{A}$ when $l = k$, and is zero when $l \neq k$. Hence

$$g_{kk} = \frac{1}{\det \mathbf{A}} \det \mathbf{A} = 1,$$

$$g_{kl} = 0 \quad (l \neq k).$$

In particular, for $n = 2$ we have in (4), in the first row, $C_{11} = a_{22}$, $C_{21} = -a_{12}$ and, in the second row, $C_{12} = -a_{21}$, $C_{22} = a_{11}$. This gives (4*). ∎

The special case $n = 2$ occurs quite frequently in geometric and other applications. You may perhaps want to memorize formula (4*). Example 2 gives an illustration of (4*).

EXAMPLE 2 Inverse of a 2 × 2 Matrix by Determinants

$$\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix}, \quad \mathbf{A}^{-1} = \frac{1}{10}\begin{bmatrix} 4 & -1 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} 0.4 & -0.1 \\ -0.2 & 0.3 \end{bmatrix}$$ ∎

EXAMPLE 3 Further Illustration of Theorem 2

Using (4), find the inverse of

$$\mathbf{A} = \begin{bmatrix} -1 & 1 & 2 \\ 3 & -1 & 1 \\ -1 & 3 & 4 \end{bmatrix}.$$

Solution. We obtain $\det \mathbf{A} = -1(-7) - 1 \cdot 13 + 2 \cdot 8 = 10$, and in (4),

$$C_{11} = \begin{vmatrix} -1 & 1 \\ 3 & 4 \end{vmatrix} = -7, \quad C_{21} = -\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = 2, \quad C_{31} = \begin{vmatrix} 1 & 2 \\ -1 & 1 \end{vmatrix} = 3,$$

$$C_{12} = -\begin{vmatrix} 3 & 1 \\ -1 & 4 \end{vmatrix} = -13, \quad C_{22} = \begin{vmatrix} -1 & 2 \\ -1 & 4 \end{vmatrix} = -2, \quad C_{32} = -\begin{vmatrix} -1 & 2 \\ 3 & 1 \end{vmatrix} = 7,$$

$$C_{13} = \begin{vmatrix} 3 & -1 \\ -1 & 3 \end{vmatrix} = 8, \quad C_{23} = -\begin{vmatrix} -1 & 1 \\ -1 & 3 \end{vmatrix} = 2, \quad C_{33} = \begin{vmatrix} -1 & 1 \\ 3 & -1 \end{vmatrix} = -2,$$

so that by (4), in agreement with Example 1,

$$\mathbf{A}^{-1} = \begin{bmatrix} -0.7 & 0.2 & 0.3 \\ -1.3 & -0.2 & 0.7 \\ 0.8 & 0.2 & -0.2 \end{bmatrix}.$$ ∎

Diagonal matrices $\mathbf{A} = [a_{jk}]$, $a_{jk} = 0$ when $j \neq k$, have an inverse if and only if all $a_{jj} \neq 0$. Then $\mathbf{A}^{-1}$ is diagonal, too, with entries $1/a_{11}, \cdots, 1/a_{nn}$.

PROOF For a diagonal matrix we have in (4)

$$\frac{C_{11}}{D} = \frac{a_{22} \cdots a_{nn}}{a_{11}a_{22} \cdots a_{nn}} = \frac{1}{a_{11}}, \qquad \text{etc.}$$ ∎

4.1 – 4.5 (Except Page no. 208 -212)---:)https://github.com/AvinashShrivastav/MATHEMATICS-FOR-COMP

4.1 INTRODUCTION TO VECTOR SPACES

Definition of a Vector Space

In Theorems 1.3 and 1.11, we proved eight properties of addition and scalar multiplication in $\mathbb{R}^n$ and $\mathcal{M}_{mn}$. These properties are important because all other results involving these operations can be derived from them. We now introduce the general class of sets called **vector spaces**,[1] having operations of addition and scalar multiplication with these same eight properties, as well as two **closure properties**.

Definition A **vector space** is a set $\mathcal{V}$ together with an operation called **vector addition** (a rule for adding two elements of $\mathcal{V}$ to obtain a third element of $\mathcal{V}$) and another operation called **scalar multiplication** (a rule for multiplying a real number times an element of $\mathcal{V}$ to obtain a second element of $\mathcal{V}$) on which the following ten properties hold:

For every $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$ in $\mathcal{V}$, and for every a and b in $\mathbb{R}$,

(A) $\mathbf{u} + \mathbf{v} \in \mathcal{V}$	Closure Property of Addition
(B) $a\mathbf{u} \in \mathcal{V}$	Closure Property of Scalar Multiplication
(1) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$	Commutative Law of Addition
(2) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$	Associative Law of Addition
(3) There is an element $\mathbf{0}$ of $\mathcal{V}$ so that for every $\mathbf{y}$ in $\mathcal{V}$ we have $\mathbf{0} + \mathbf{y} = \mathbf{y} = \mathbf{y} + \mathbf{0}$.	Existence of Identity Element for Addition
(4) There is an element $-\mathbf{u}$ in $\mathcal{V}$ such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0} = (-\mathbf{u}) + \mathbf{u}$.	Existence of Additive Inverse
(5) $a(\mathbf{u} + \mathbf{v}) = (a\mathbf{u}) + (a\mathbf{v})$	Distributive Laws for Scalar
(6) $(a + b)\mathbf{u} = (a\mathbf{u}) + (b\mathbf{u})$	Multiplication over Addition
(7) $(ab)\mathbf{u} = a(b\mathbf{u})$	Associativity of Scalar Multiplication
(8) $1\mathbf{u} = \mathbf{u}$	Identity Property for Scalar Multiplication

The elements of a vector space $\mathcal{V}$ are called **vectors**.

The two closure properties require that both the operations of vector addition and scalar multiplication always produce an element of the vector space as a result.

[1] We actually define what are called *real vector spaces*, rather than just vector spaces. The word *real* implies that the scalars involved in the scalar multiplication are real numbers. In Chapter 7, we consider *complex vector spaces*, where the scalars are complex numbers. Other types of vector spaces involving more general sets of scalars are not considered in this book.

The standard plus sign, "+," is used to indicate both vector addition and the sum of real numbers, two different operations. All sums in properties (1), (2), (3), (4), and (5) are vector sums. In property (6), the "+" on the left side of the equation represents addition of real numbers; the "+" on the right side stands for the sum of two vectors. In property (7), the left side of the equation contains one product of real numbers, ab, and one instance of scalar multiplication, (ab) times $\mathbf{u}$. The right side of property (7) involves two scalar multiplications — first, b times $\mathbf{u}$, then, a times the vector $(b\mathbf{u})$. Usually we can tell from the context which type of operation is being used.

In any vector space, the additive identity element in property (3) is unique, and the additive inverse (property (4)) of each vector is unique (see the proof of part (3) of Theorem 4.1 and Exercise 12).

Examples of Vector Spaces

Example 1

Let $\mathcal{V} = \mathbb{R}^n$, with addition and scalar multiplication of n-vectors as defined in Section 1.1. Since these operations always produce vectors in $\mathbb{R}^n$, the closure properties certainly hold for $\mathbb{R}^n$. By Theorem 1.3, the remaining eight properties hold as well. Thus, $\mathcal{V} = \mathbb{R}^n$ is a vector space with these operations.

Similarly, consider $\mathcal{M}_{mn}$, the set of $m \times n$ matrices. The usual operations of matrix addition and scalar multiplication on $\mathcal{M}_{mn}$ always produce $m \times n$ matrices, and so the closure properties certainly hold for $\mathcal{M}_{mn}$. By Theorem 1.11, the remaining eight properties hold as well. Hence, $\mathcal{M}_{mn}$ is a vector space with these operations. ■

$\mathbb{R}^n$ and $\mathcal{M}_{mn}$ (with the usual operations of addition and scalar multiplication) are representative of most of the vector spaces we consider here. Keep $\mathbb{R}^n$ and $\mathcal{M}_{mn}$ in mind as examples later, as we consider theorems involving general vector spaces.

Some vector spaces can have additional operations. For example, $\mathbb{R}^n$ has the dot product, and $\mathcal{M}_{nn}$ has matrix multiplication and the transpose. But these additional structures are not shared by all vector spaces because they are not included in the definition. We cannot assume the existence of any additional operations in a general discussion of vector spaces. In particular, there is no such operation as multiplication or division of one vector by another in general vector spaces. The only general vector space operation that combines two *vectors* is vector addition.

Example 2

The set $\mathcal{V} = \{\mathbf{0}\}$ is a vector space with the rules for addition and multiplication given by $\mathbf{0} + \mathbf{0} = \mathbf{0}$ and $a\mathbf{0} = \mathbf{0}$ for every scalar (real number) a. Since $\mathbf{0}$ is the only possible result of either operation, $\mathcal{V}$ must be closed under both addition and scalar multiplication. A quick check verifies that the remaining eight properties also hold for $\mathcal{V}$. This vector space is called the **trivial vector space**, and no smaller vector space is possible (why?). ■

Example 3

Consider $\mathbb{R}^3$ as the set of 3-vectors in three-dimensional space, all with initial points at the origin. Let $\mathcal{W}$ be any plane containing the origin. $\mathcal{W}$ can also be considered as the set of all 3-vectors whose terminal point lies in this plane (that is, $\mathcal{W}$ is the set of all 3-vectors that lie entirely in the plane when drawn on a graph, since both the initial point and terminal point of each vector lie in the plane). For example, in Figure 4.1, $\mathcal{W}$ is the plane containing the vectors $\mathbf{u}$ and $\mathbf{v}$ (elements of $\mathcal{W}$); $\mathbf{q}$ is not in $\mathcal{W}$ because its terminal point does not lie in the plane. We will prove that $\mathcal{W}$ is a vector space.

To check the closure properties, we must show that the sum of any two vectors in $\mathcal{W}$ is a vector in $\mathcal{W}$ and that any scalar multiple of a vector in $\mathcal{W}$ also lies in $\mathcal{W}$.

If $\mathbf{x}$ and $\mathbf{y}$ are elements of $\mathcal{W}$, then the parallelogram they form lies entirely in the plane, because $\mathbf{x}$ and $\mathbf{y}$ do. Hence, the diagonal $\mathbf{x}+\mathbf{y}$ of this parallelogram also lies in the plane, so $\mathbf{x}+\mathbf{y}$ is in $\mathcal{W}$. This observation verifies that $\mathcal{W}$ is closed under vector addition (that is, the closure property holds for vector addition). Notice that it is not enough to know that the sum of two 3-vectors in $\mathcal{W}$ produces another 3-vector. We have to show that the sum they produce is actually in the set $\mathcal{W}$.

Next consider scalar multiplication. If $\mathbf{x}$ is a vector in $\mathcal{W}$, then any scalar multiple of $\mathbf{x}$, $a\mathbf{x}$, is either parallel to $\mathbf{x}$ or equal to $\mathbf{0}$. Therefore, $a\mathbf{x}$ lies in any plane through the origin that contains $\mathbf{x}$ (in particular, $\mathcal{W}$). Hence, $a\mathbf{x}$ is in $\mathcal{W}$, and $\mathcal{W}$ is closed under scalar multiplication.

We now check that the remaining eight vector space properties hold. Properties (1), (2), (5), (6), (7), and (8) are true for all vectors in $\mathcal{W}$ by Theorem 1.3, since $\mathcal{W} \subseteq \mathbb{R}^3$. However, properties (3) and (4) must be checked separately for $\mathcal{W}$ because they are *existence* properties. We know that the zero vector and additive inverses exist in $\mathbb{R}^3$, but are they in $\mathcal{W}$? Now, $\mathbf{0} = [0,0,0]$ is in $\mathcal{W}$, because the plane $\mathcal{W}$ passes through the origin, thus proving property (3). Also, the opposite (additive inverse) of any vector lying in the plane $\mathcal{W}$ also lies in $\mathcal{W}$, thus proving property (4). Hence, all eight properties and the closure properties are true, so $\mathcal{W}$ is a vector space.

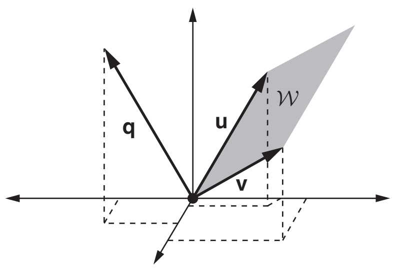


FIGURE 4.1

A plane $\mathcal{W}$ in $\mathbb{R}^3$ containing the origin

Example 4

Let $\mathcal{P}_n$ be the set of polynomials of degree $\leq n$, with real coefficients. The vectors in $\mathcal{P}_n$ have the form $\mathbf{p} = a_n x^n + \cdots + a_1 x + a_0$ for some real numbers $a_0, a_1, \ldots, a_n$. We define addition of polynomials in the usual manner — that is, by adding corresponding coefficients. Then the sum of any two polynomials of degree $\leq n$ also has degree $\leq n$ and so is in $\mathcal{P}_n$. Thus, the closure property of addition holds. Similarly, if b is a real number and $\mathbf{p} = a_n x^n + \cdots + a_1 x + a_0$ is in $\mathcal{P}_n$, we define $b\mathbf{p}$ to be the polynomial $(ba_n)x^n + \cdots + (ba_1)x + ba_0$, which is also in $\mathcal{P}_n$. Hence, the closure property of scalar multiplication holds. Then, if the remaining eight vector space properties hold, $\mathcal{P}_n$ is a vector space under these operations. We verify properties (1), (3), and (4) of the definition and leave the others for you to check.

(1) Commutative Law of Addition: We must show that the order in which two vectors (polynomials) are added makes no difference. Now, by the commutative law of addition for real numbers,

$$\begin{aligned}
&(a_n x^n + \cdots + a_1 x + a_0) + \left(b_n x^n + \cdots + b_1 x + b_0\right) \\
&\quad = (a_n + b_n)x^n + \cdots + (a_1 + b_1)x + (a_0 + b_0) \\
&\quad = (b_n + a_n)x^n + \cdots + (b_1 + a_1)x + (b_0 + a_0) \\
&\quad = (b_n x^n + \cdots + b_1 x + b_0) + (a_n x^n + \cdots + a_1 x + a_0).
\end{aligned}$$

(3) Existence of Identity Element for Addition: The zero-degree polynomial $\mathbf{z} = 0x^n + \cdots + 0x + 0$ acts as the additive identity element $\mathbf{0}$. That is, adding $\mathbf{z}$ to any vector $\mathbf{p} = a_n x^n + \cdots + a_1 x + a_0$ does not change the vector:

$$\mathbf{z} + \mathbf{p} = (0 + a_n)x^n + \cdots + (0 + a_1)x + (0 + a_0) = \mathbf{p}.$$

(4) Existence of Additive Inverse: We must show that each vector $\mathbf{p} = a_n x^n + \cdots + a_1 x + a_0$ in $\mathcal{P}_n$ has an additive inverse in $\mathcal{P}_n$. But, the vector $-\mathbf{p} = -(a_n x^n + \cdots + a_1 x + a_0) = (-a_n)x^n + \cdots + (-a_1)x + (-a_0)$ has the property that $\mathbf{p} + [-\mathbf{p}] = \mathbf{z}$, the zero vector, and so $-\mathbf{p}$ acts as the additive inverse of $\mathbf{p}$. Because $-\mathbf{p}$ is also in $\mathcal{P}_n$, we are done. ■

The vector space in Example 4 is similar to our prototype $\mathbb{R}^n$. For any polynomial in $\mathcal{P}_n$, consider the sequence of its $n + 1$ coefficients. This sequence completely describes that polynomial and can be thought of as an $(n + 1)$-vector. For example, a polynomial $a_2 x^2 + a_1 x + a_0$ in $\mathcal{P}_2$ can be described by the 3-vector $[a_2, a_1, a_0]$. In this way, the vector space $\mathcal{P}_2$ "resembles" the vector space $\mathbb{R}^3$, and in general, $\mathcal{P}_n$ "resembles" $\mathbb{R}^{n+1}$. We will frequently capitalize on this "resemblance" in an informal way throughout the chapter. We will formalize this relationship between $\mathcal{P}_n$ and $\mathbb{R}^{n+1}$ in Section 5.5.

Example 5

The set $\mathcal{P}$ of all polynomials (of all degrees) is a vector space under the usual (term-by-term) operations of addition and scalar multiplication (see Exercise 15). ■

Example 6

Let $\mathcal{V}$ be the set of all real-valued functions defined on $\mathbb{R}$. For example, $\mathbf{f}(x) = \arctan(x)$ is in $\mathcal{V}$. We define addition of functions as usual: $\mathbf{h} = \mathbf{f} + \mathbf{g}$ is the function such that $\mathbf{h}(x) = \mathbf{f}(x) + \mathbf{g}(x)$, for every $x \in \mathbb{R}$. Similarly, if $a \in \mathbb{R}$ and $\mathbf{f}$ is in $\mathcal{V}$, we define the scalar multiple $\mathbf{h} = a\mathbf{f}$ to be the function such that $\mathbf{h}(x) = a\mathbf{f}(x)$, for every $x \in \mathbb{R}$. Now, the closure properties hold for $\mathcal{V}$ because sums and scalar multiples of real-valued functions produce real-valued functions. To finish verifying that $\mathcal{V}$ is a vector space, we must check that the remaining eight vector space properties hold.

Suppose that $\mathbf{f}, \mathbf{g}$, and $\mathbf{h}$ are in $\mathcal{V}$, and a and b are real numbers.

Property (1): For every x in $\mathbb{R}$, $\mathbf{f}(x)$ and $\mathbf{g}(x)$ are both real numbers. Hence, $\mathbf{f}(x) + \mathbf{g}(x) = \mathbf{g}(x) + \mathbf{f}(x)$ for all $x \in \mathbb{R}$, by the commutative law of addition for real numbers, so each represents the same function of x. Hence, $\mathbf{f} + \mathbf{g} = \mathbf{g} + \mathbf{f}$.

Property (2): For every $x \in \mathbb{R}$, $\mathbf{f}(x) + \big(\mathbf{g}(x) + \mathbf{h}(x)\big) = \big(\mathbf{f}(x) + \mathbf{g}(x)\big) + \mathbf{h}(x)$, by the associative law of addition for real numbers. Thus, $\mathbf{f} + (\mathbf{g} + \mathbf{h}) = (\mathbf{f} + \mathbf{g}) + \mathbf{h}$.

Property (3): Let $\mathbf{z}$ be the function given by $\mathbf{z}(x) = 0$ for every $x \in \mathbb{R}$. Then, for each x, $\mathbf{z}(x) + \mathbf{f}(x) = 0 + \mathbf{f}(x) = \mathbf{f}(x)$. Hence, $\mathbf{z} + \mathbf{f} = \mathbf{f}$.

Property (4): Given $\mathbf{f}$ in $\mathcal{V}$, define $-\mathbf{f}$ by $[-\mathbf{f}](x) = -(\mathbf{f}(x))$ for every $x \in \mathbb{R}$. Then, for all x, $[-\mathbf{f}](x) + \mathbf{f}(x) = -(\mathbf{f}(x)) + \mathbf{f}(x) = 0$. Therefore, $[-\mathbf{f}] + \mathbf{f} = \mathbf{z}$, the zero vector, and so the additive inverse of $\mathbf{f}$ is also in $\mathcal{V}$.

Properties (5) and (6): For every $x \in \mathbb{R}$, $a(\mathbf{f}(x) + \mathbf{g}(x)) = a\mathbf{f}(x) + a\mathbf{g}(x)$ and $(a+b)\mathbf{f}(x) = a\mathbf{f}(x) + b\mathbf{f}(x)$ by the distributive laws for real numbers of multiplication over addition. Hence, $a(\mathbf{f} + \mathbf{g}) = a\mathbf{f} + a\mathbf{g}$, and $(a+b)\mathbf{f} = a\mathbf{f} + b\mathbf{f}$.

Property (7): For every $x \in \mathbb{R}$, $(ab)\mathbf{f}(x) = a(b\mathbf{f}(x))$ follows from the associative law of multiplication for real numbers. Hence, $(ab)\mathbf{f} = a(b\mathbf{f})$.

Property (8): Since $1 \cdot \mathbf{f}(x) = \mathbf{f}(x)$ for every real number x, we have $1 \cdot \mathbf{f} = \mathbf{f}$ in $\mathcal{V}$. ∎

Two Unusual Vector Spaces

The next two examples place unusual operations on familiar sets to create new vector spaces. In such cases, regardless of how the operations are defined, we sometimes use the symbols $\oplus$ and $\odot$ to denote addition and scalar multiplication, respectively, in order to remind ourselves that these operations are not the "regular" ones. Note that $\oplus$ is defined differently in Examples 7 and 8 (and similarly for $\odot$).

Example 7

Let $\mathcal{V}$ be the set $\mathbb{R}^+$ of positive real numbers. This set is not a vector space under the usual operations of addition and scalar multiplication (why?). However, we can define new rules for these operations to make $\mathcal{V}$ a vector space. In what follows, we sometimes think of elements of $\mathbb{R}^+$ as abstract vectors (in which case we use boldface type, such as $\mathbf{v}$) or as the values on the positive real number line they represent (in which case we use italics, such as v).

To define "addition" on $\mathcal{V}$, we use *multiplication* of real numbers. That is,

$$\mathbf{v}_1 \oplus \mathbf{v}_2 = v_1 \cdot v_2$$

for every $\mathbf{v}_1$ and $\mathbf{v}_2$ in $\mathcal{V}$, where we use the symbol $\oplus$ for the "addition" operation on $\mathcal{V}$ to emphasize that this is not addition of real numbers. The definition of a vector space states only that vector addition must be a rule for combining two vectors to yield a third vector so that properties (1) through (8) hold. There is no stipulation that vector addition must be at all similar to ordinary addition of real numbers.[2]

We next define "scalar multiplication," $\odot$, on $\mathcal{V}$ by

$$a \odot \mathbf{v} = v^a$$

for every $a \in \mathbb{R}$ and $\mathbf{v} \in \mathcal{V}$.

From the given definitions, we see that if $\mathbf{v}_1$ and $\mathbf{v}_2$ are in $\mathcal{V}$ and a is in $\mathbb{R}$, then both $\mathbf{v}_1 \oplus \mathbf{v}_2$ and $a \odot \mathbf{v}_1$ are in $\mathcal{V}$, thus verifying the two closure properties. To prove the other eight properties, we assume that $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in \mathcal{V}$ and that $a, b \in \mathbb{R}$. We then have the following:

Property (1): $\mathbf{v}_1 \oplus \mathbf{v}_2 = v_1 \cdot v_2 = v_2 \cdot v_1$ (by the commutative law of multiplication for real numbers) $= \mathbf{v}_2 \oplus \mathbf{v}_1$.

Property (2): $\mathbf{v}_1 \oplus (\mathbf{v}_2 \oplus \mathbf{v}_3) = \mathbf{v}_1 \oplus (v_2 \cdot v_3) = v_1 \cdot (v_2 \cdot v_3) = (v_1 \cdot v_2) \cdot v_3$ (by the associative law of multiplication for real numbers) $= (\mathbf{v}_1 \oplus \mathbf{v}_2) \cdot v_3 = (\mathbf{v}_1 \oplus \mathbf{v}_2) \oplus \mathbf{v}_3$.

Property (3): The number 1 in $\mathbb{R}^+$ acts as the zero vector $\mathbf{0}$ in $\mathcal{V}$ (why?).

Property (4): The additive inverse of $\mathbf{v}$ in $\mathcal{V}$ is the positive real number $(1/v)$, because $\mathbf{v} \oplus (1/v) = v \cdot (1/v) = 1$, the zero vector in $\mathcal{V}$.

Property (5): $a \odot (\mathbf{v}_1 \oplus \mathbf{v}_2) = a \odot (v_1 \cdot v_2) = (v_1 \cdot v_2)^a = v_1^a \cdot v_2^a = (a \odot \mathbf{v}_1) \cdot (a \odot \mathbf{v}_2) = (a \odot \mathbf{v}_1) \oplus (a \odot \mathbf{v}_2)$.

Property (6): $(a+b) \odot \mathbf{v} = v^{a+b} = v^a \cdot v^b = (a \odot \mathbf{v}) \cdot (b \odot \mathbf{v}) = (a \odot \mathbf{v}) \oplus (b \odot \mathbf{v})$.

Property (7): $(ab) \odot \mathbf{v} = v^{ab} = \left(v^b\right)^a = (b \odot \mathbf{v})^a = a \odot (b \odot \mathbf{v})$.

Property (8): $1 \odot \mathbf{v} = v^1 = \mathbf{v}$. ∎

Example 8

Let $\mathcal{V} = \mathbb{R}^2$, with addition defined by

$$[x, y] \oplus [w, z] = [x + w + 1,\ y + z - 2]$$

and scalar multiplication defined by

$$a \odot [x, y] = [ax + a - 1,\ ay - 2a + 2].$$

The closure properties hold for these operations (why?). In fact, $\mathcal{V}$ forms a vector space because the eight vector properties also hold. We verify properties (2), (3), (4), and (6) and leave the others for you to check.

[2] You might expect the operation $\oplus$ to be called something other than "addition." However, most of our vector space terminology comes from the motivating example of $\mathbb{R}^n$, so the word *addition* is a natural choice for the name of the operation.

Property (2):
$$\begin{aligned}[x,y]\oplus([u,v]\oplus[w,z]) &= [x,y]\oplus[u+w+1,\ v+z-2]\\ &= [x+u+w+2,\ y+v+z-4]\\ &= [x+u+1,\ y+v-2]\oplus[w,z]\\ &= ([x,y]\oplus[u,v])\oplus[w,z].\end{aligned}$$

Property (3): The vector $[-1,2]$ acts as the zero vector, since

$$[x,y]\oplus[-1,2] = [x+(-1)+1,\ y+2-2] = [x,y].$$

Property (4): The additive inverse of $[x,y]$ is $[-x-2,-y+4]$, because

$$[x,y]\oplus[-x-2,-y+4] = [x-x-2+1,\ y-y+4-2] = [-1,2],$$

the zero vector in $\mathcal{V}$.

Property (6):

$$\begin{aligned}(a+b)\odot[x,y] &= [(a+b)x+(a+b)-1,\ (a+b)y-2(a+b)+2]\\ &= [(ax+a-1)+(bx+b-1)+1,\ (ay-2a+2)+(by-2b+2)-2]\\ &= [ax+a-1,\ ay-2a+2]\oplus[bx+b-1,\ by-2b+2]\\ &= \left(a\odot[x,y]\right)\oplus\left(b\odot[x,y]\right).\end{aligned}$$

■

Some Elementary Properties of Vector Spaces

The next theorem contains several simple results regarding vector spaces. Although these are obviously true in the most familiar examples, we must prove them in general before we know they hold in every possible vector space.

Theorem 4.1 Let $\mathcal{V}$ be a vector space. Then, for every vector $\mathbf{v}$ in $\mathcal{V}$ and every real number a, we have

(1) $a\mathbf{0} = \mathbf{0}$	Any scalar multiple of the zero vector yields the zero vector.
(2) $0\mathbf{v} = \mathbf{0}$	The scalar zero multiplied by any vector yields the zero vector.
(3) $(-1)\mathbf{v} = -\mathbf{v}$	The scalar -1 multiplied by any vector yields the additive inverse of that vector.
(4) If $a\mathbf{v} = \mathbf{0}$, then $a = 0$ or $\mathbf{v} = \mathbf{0}$.	If a scalar multiplication yields the zero vector, then either the scalar is zero, or the vector is the zero vector, or both.

Part (3) justifies the notation for the additive inverse in property (4) of the definition of a vector space and shows we do not need to distinguish between $-\mathbf{v}$ and $(-1)\mathbf{v}$.

This theorem must be proved directly from the properties in the definition of a vector space because at this point we have no other known facts about general vector spaces. We prove parts (1), (3), and (4). The proof of part (2) is similar to the proof of part (1) and is left as Exercise 18.

Proof. (Abridged):

Part (1): By direct proof,

$$\begin{aligned} a\mathbf{0} &= a\mathbf{0} + \mathbf{0} && \text{by property (3)} \\ &= a\mathbf{0} + (a\mathbf{0} + (-[a\mathbf{0}])) && \text{by property (4)} \\ &= (a\mathbf{0} + a\mathbf{0}) + (-[a\mathbf{0}]) && \text{by property (2)} \\ &= a(\mathbf{0} + \mathbf{0}) + (-[a\mathbf{0}]) && \text{by property (5)} \\ &= a\mathbf{0} + (-[a\mathbf{0}]) && \text{by property (3)} \\ &= \mathbf{0}. && \text{by property (4)} \end{aligned}$$

Part (3): First, note that $\mathbf{v} + (-1)\mathbf{v} = 1\mathbf{v} + (-1)\mathbf{v}$ (by property (8)) $= (1 + (-1))\mathbf{v}$ (by property (6)) $= 0\mathbf{v} = \mathbf{0}$ (by part (2) of Theorem 4.1). Therefore, $(-1)\mathbf{v}$ acts as *an* additive inverse for $\mathbf{v}$. We will finish the proof by showing that the additive inverse for $\mathbf{v}$ is unique. Hence, $(-1)\mathbf{v}$ will be *the* additive inverse of $\mathbf{v}$.

Suppose that $\mathbf{x}$ and $\mathbf{y}$ are both additive inverses for $\mathbf{v}$. Thus, $\mathbf{x} + \mathbf{v} = \mathbf{0}$ and $\mathbf{v} + \mathbf{y} = \mathbf{0}$. Hence,

$$\mathbf{x} = \mathbf{x} + \mathbf{0} = \mathbf{x} + (\mathbf{v} + \mathbf{y}) = (\mathbf{x} + \mathbf{v}) + \mathbf{y} = \mathbf{0} + \mathbf{y} = \mathbf{y}.$$

Therefore, any two additive inverses of $\mathbf{v}$ are equal. (Note that this is, in essence, the same proof we gave for Theorem 2.10, the uniqueness of inverse for matrix multiplication. You should compare these proofs.)

Part (4): This is an "If A then B or C" statement. Therefore, we assume that $a\mathbf{v} = \mathbf{0}$ and $a \neq 0$ and show that $\mathbf{v} = \mathbf{0}$. Now,

$$\begin{aligned} \mathbf{v} &= 1\mathbf{v} && \text{by property (8)} \\ &= \left(\tfrac{1}{a} \cdot a\right)\mathbf{v} && \text{because } a \neq 0 \\ &= \left(\tfrac{1}{a}\right)(a\mathbf{v}) && \text{by property (7)} \\ &= \left(\tfrac{1}{a}\right)\mathbf{0} && \text{because } a\mathbf{v} = \mathbf{0} \\ &= \mathbf{0}. && \text{by part (1) of Theorem 4.1} \end{aligned}$$

□

Theorem 4.1 is valid even for unusual vector spaces, such as those in Examples 7 and 8. For instance, part (4) of the theorem claims that, in general, $a\mathbf{v} = \mathbf{0}$ implies $a = 0$ or $\mathbf{v} = \mathbf{0}$. This statement can quickly be verified for the vector space $\mathcal{V} = \mathbb{R}^+$ with operations $\oplus$ and $\odot$ from Example 7. In this case, $a \odot \mathbf{v} = v^a$, and the zero vector $\mathbf{0}$ is the real number 1. Then, part (4) is equivalent here to the true statement that $v^a = 1$ implies $a = 0$ or $v = 1$.

Applying parts (2) and (3) of Theorem 4.1 to an unusual vector space $\mathcal{V}$ gives a quick way of finding the zero vector $\mathbf{0}$ of $\mathcal{V}$ and the additive inverse $-\mathbf{v}$ for any vector $\mathbf{v}$ in $\mathcal{V}$. For instance, in Example 8, we have $\mathcal{V} = \mathbb{R}^2$ with scalar multiplication $a \odot [x,y] = [ax + a - 1, ay - 2a + 2]$. To find the zero vector $\mathbf{0}$ in $\mathcal{V}$, we simply multiply the scalar 0 by any general vector $[x,y]$ in $\mathcal{V}$:

$$\mathbf{0} = 0 \odot [x,y] = [0x + 0 - 1,\ 0y - 2(0) + 2] = [-1, 2].$$

Similarly, if $[x,y] \in \mathcal{V}$, then $-1 \odot [x,y]$ gives the additive inverse of $[x,y]$.

$$\begin{aligned} -[x,y] &= -1 \odot [x,y] = [-1x + (-1) - 1, -1y - 2(-1) + 2] \\ &= [-x - 2, -y + 4]. \end{aligned}$$

Failure of the Vector Space Conditions

We conclude this section by considering some sets that are not vector spaces to see what can go wrong.

Example 9

The set Φ of real-valued functions, f, defined on the interval $[0,1]$ such that $f(\frac{1}{2}) = 1$, is not a vector space under the usual operations of function addition and scalar multiplication because the closure properties do not hold. If f and g are in Φ, then

$$(f+g)\left(\frac{1}{2}\right) = f\left(\frac{1}{2}\right) + g\left(\frac{1}{2}\right) = 1 + 1 = 2 \neq 1,$$

so $f + g$ is not in Φ. Therefore, Φ is not closed under addition and cannot be a vector space. (Is Φ closed under scalar multiplication?)

Example 10

Let Υ be the set $\mathbb{R}^2$ with operations

$$\mathbf{v}_1 \oplus \mathbf{v}_2 = \mathbf{v}_1 + \mathbf{v}_2 \quad \text{and} \quad c \odot \mathbf{v} = c(\mathbf{A}\mathbf{v}), \quad \text{where } \mathbf{A} = \begin{bmatrix} -3 & 1 \\ 5 & -2 \end{bmatrix}.$$

With these operations, Υ is not a vector space. You can verify that Υ is closed under $\oplus$ and $\odot$, but properties (7) and (8) of the definition are not satisfied. For example, property (8) fails since

$$1 \odot \begin{bmatrix} 2 \\ 7 \end{bmatrix} = 1\left(\begin{bmatrix} -3 & 1 \\ 5 & -2 \end{bmatrix}\begin{bmatrix} 2 \\ 7 \end{bmatrix}\right) = 1\begin{bmatrix} 1 \\ -4 \end{bmatrix} = \begin{bmatrix} 1 \\ -4 \end{bmatrix} \neq \begin{bmatrix} 2 \\ 7 \end{bmatrix}.$$

New Vocabulary

closure properties
scalar multiplication (in a general vector space)
trivial vector space
vector addition (in a general vector space)
vector space
vectors (in a general vector space)

Highlights

- Vector spaces have two specified operations: vector addition $(+)$ and scalar multiplication $(\cdot)$. A vector space is closed under these operations and possesses eight additional fundamental properties (as stated in the definition).
- The smallest possible vector space is the trivial vector space.
- Familiar vector spaces (under natural operations) include $\mathbb{R}^n$, $\mathcal{M}_{mn}, \mathcal{P}_n, \mathcal{P}$, a line through the origin, a plane through the origin, all real-valued functions.
- Any scalar multiple of the zero vector equals the zero vector.
- The scalar 0 times any vector equals the zero vector.
- The scalar -1 times any vector gives the additive inverse of the vector.
- If a scalar multiple of a vector equals the zero vector, then either the scalar is zero or the vector is zero.

EXERCISES FOR SECTION 4.1

Remember: To verify that a given set with its operations is a vector space, you must prove the two closure properties as well as the remaining eight properties in the definition. To show that a set with operations is *not* a vector space, you need only find an example showing that one of the closure properties or one of the remaining eight properties is not satisfied.

1. Rewrite properties (2), (5), (6), and (7) in the definition of a vector space using the symbols $\oplus$ for vector addition and $\odot$ for scalar multiplication. (The notations for real number addition and multiplication should not be changed.)

2. Prove that the set of all scalar multiples of the vector $[1,3,2]$ in $\mathbb{R}^3$ forms a vector space with the usual operations on 3-vectors.

3. Verify that the set of polynomials $\mathbf{f}$ in $\mathcal{P}_3$ such that $\mathbf{f}(2) = 0$ forms a vector space with the standard operations.

4. Prove that $\mathbb{R}$ is a vector space using the operations $\oplus$ and $\odot$ given by $\mathbf{x} \oplus \mathbf{y} = \left(x^3 + y^3\right)^{1/3}$ and $a \odot \mathbf{x} = \left(\sqrt[3]{a}\right)x$.

★5. Show that the set of singular 2×2 matrices under the usual operations is *not* a vector space.

6. Prove that the set of nonsingular $n \times n$ matrices under the usual operations is *not* a vector space.

7. Show that $\mathbb{R}$, with ordinary addition but with scalar multiplication replaced by $a \odot \mathbf{x} = \mathbf{0}$ for every real number a, is *not* a vector space.

★8. Show that the set $\mathbb{R}$, with the usual scalar multiplication but with addition given by $x \oplus y = 2(x + y)$, is *not* a vector space.

9. Show that the set $\mathbb{R}^2$, with the usual scalar multiplication but with vector addition replaced by $[x,y] \oplus [w,z] = [x + w, 0]$, does *not* form a vector space.

10. Let $\mathcal{A} = \mathbb{R}$, with the operations $\oplus$ and $\odot$ given by $\mathbf{x} \oplus \mathbf{y} = \left(x^5 + y^5\right)^{1/5}$ and $a \odot \mathbf{x} = a\mathbf{x}$. Determine whether $\mathcal{A}$ is a vector space. Prove your answer.

11. Let $\mathbf{A}$ be a fixed $m \times n$ matrix, and let $\mathbf{B}$ be a fixed m-vector (in $\mathbb{R}^m$). Let $\mathcal{V}$ be the set of solutions $\mathbf{X}$ (in $\mathbb{R}^n$) to the matrix equation $\mathbf{AX} = \mathbf{B}$. Endow $\mathcal{V}$ with the usual n-vector operations.

   (a) Assume $\mathcal{V}$ is nonempty. Show that the closure properties are satisfied in $\mathcal{V}$ if and only if $\mathbf{B} = \mathbf{0}$.

   (b) Explain why properties (1), (2), (5), (6), (7), and (8) in the definition of a vector space have already been proved for $\mathcal{V}$ in Theorem 1.3.

   (c) Prove that property (3) in the definition of a vector space is satisfied if and only if $\mathbf{B} = \mathbf{0}$.

   (d) Explain why property (4) in the definition makes no sense unless property (3) is satisfied. Prove property (4) when $\mathbf{B} = \mathbf{0}$.

   (e) Use parts (a) through (d) of this exercise to determine necessary and sufficient conditions for $\mathcal{V}$ to be a vector space.

12. Let $\mathcal{V}$ be a vector space. Prove that the identity element for vector addition in $\mathcal{V}$ is unique. (Hint: Use a proof by contradiction.)

13. The set $\mathbb{R}^2$ with operations $[x,y] \oplus [w,z] = [x + w - 2, y + z + 3]$ and $a \odot [x,y] = [ax - 2a + 2, ay + 3a - 3]$ is a vector space. Use parts (2) and (3) of Theorem 4.1 to find the zero vector $\mathbf{0}$ and the additive inverse of each vector $\mathbf{v} = [x,y]$ for this vector space. Then check your answers.

14. Let $\mathcal{V}$ be a vector space. Prove the following **cancellation laws**:

   (a) If $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$ are vectors in $\mathcal{V}$ for which $\mathbf{u} + \mathbf{v} = \mathbf{w} + \mathbf{v}$, then $\mathbf{u} = \mathbf{w}$.

   (b) If a and b are scalars and $\mathbf{v} \neq \mathbf{0}$ is a vector in $\mathcal{V}$ with $a\mathbf{v} = b\mathbf{v}$, then $a = b$.

   (c) If $a \neq 0$ is a scalar and $\mathbf{v}, \mathbf{w} \in \mathcal{V}$ with $a\mathbf{v} = a\mathbf{w}$, then $\mathbf{v} = \mathbf{w}$.

15. Prove that the set $\mathcal{P}$ of all polynomials with real coefficients forms a vector space under the usual operations of polynomial (term-by-term) addition and scalar multiplication.

16. Let X be any set, and let $\mathcal{V} = \{$all real-valued functions with domain $X\}$. Prove that $\mathcal{V}$ is a vector space using ordinary addition and scalar multiplication of real-valued functions. (Hint: Alter the proof in Example 6 appropriately.)

17. Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be vectors in a vector space $\mathcal{V}$, and let $a_1, \ldots, a_n$ be any real numbers. Use induction to prove that $\sum_{i=1}^{n} a_i \mathbf{v}_i$ is in $\mathcal{V}$.

18. Prove part (2) of Theorem 4.1.

19. Prove that every nontrivial vector space has an infinite number of distinct elements.

⋆20. True or False:

(a) The set $\mathbb{R}^n$ under any operations of "addition" and "scalar multiplication" is a vector space.

(b) The set of all polynomials of degree 7 is a vector space under the usual operations of addition and scalar multiplication.

(c) The set of all polynomials of degree ≤ 7 is a vector space under the usual operations of addition and scalar multiplication.

(d) If $\mathbf{x}$ is a vector in a vector space $\mathcal{V}$, and c is a nonzero scalar, then $c\mathbf{x} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$.

(e) In a vector space, scalar multiplication by the zero scalar always results in the zero scalar.

(f) In a vector space, scalar multiplication of a vector $\mathbf{x}$ by -1 always results in the additive inverse of $\mathbf{x}$.

(g) The set of all real-valued functions f on $\mathbb{R}$ such that $f(1) = 0$ is a vector space under the usual operations of addition and scalar multiplication.

4.2 SUBSPACES

Section 4.1 presented several examples in which two vector spaces share the same addition and scalar multiplication operations, with one as a subset of the other. In fact, most of these examples involve subsets of either $\mathbb{R}^n$, $\mathcal{M}_{mn}$, or the vector space of real-valued functions defined on some set (see Exercise 16 in Section 4.1). As we will see, when a vector space is a subset of a known vector space and has the same operations, it becomes easier to handle. These subsets, called **subspaces**, also provide additional information about the larger vector space.

Definition of a Subspace and Examples

Definition Let $\mathcal{V}$ be a vector space. Then $\mathcal{W}$ is a **subspace** of $\mathcal{V}$ if and only if $\mathcal{W}$ is a subset of $\mathcal{V}$, and $\mathcal{W}$ is itself a vector space with the same operations as $\mathcal{V}$.

That is, $\mathcal{W}$ is a subspace of $\mathcal{V}$ if and only if $\mathcal{W}$ is a vector space inside $\mathcal{V}$ such that for every a in $\mathbb{R}$ and every $\mathbf{v}$ and $\mathbf{w}$ in $\mathcal{W}$, $\mathbf{v}+\mathbf{w}$ and $a\mathbf{v}$ yield the same vectors when the operations are performed in $\mathcal{W}$ as when they are performed in $\mathcal{V}$.

Example 1

Example 3 of Section 4.1 showed that the set of points lying on a plane $\mathcal{W}$ through the origin in $\mathbb{R}^3$ forms a vector space under the usual addition and scalar multiplication in $\mathbb{R}^3$. $\mathcal{W}$ is certainly a subset of $\mathbb{R}^3$. Hence, the vector space $\mathcal{W}$ is a subspace of $\mathbb{R}^3$.

Example 2

The set $\mathcal{S}$ of scalar multiples of the vector $[1,3,2]$ in $\mathbb{R}^3$ forms a vector space under the usual addition and scalar multiplication in $\mathbb{R}^3$ (see Exercise 2 in Section 4.1). $\mathcal{S}$ is certainly a subset of $\mathbb{R}^3$. Hence, $\mathcal{S}$ is a subspace of $\mathbb{R}^3$. Notice that $\mathcal{S}$ corresponds geometrically to the set of points lying on the line through the origin in $\mathbb{R}^3$ in the direction of the vector $[1,3,2]$ (see Figure 4.2). In the same manner, every line through the origin determines a subspace of $\mathbb{R}^3$ — namely, the set of scalar multiples of a nonzero vector in the direction of that line.

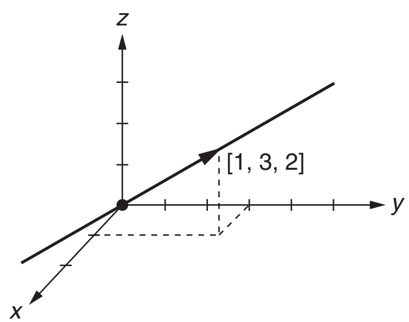


FIGURE 4.2

Line containing all scalar multiples of $[1,3,2]$

Example 3

Let $\mathcal{V}$ be any vector space. Then $\mathcal{V}$ is a subspace of itself (why?). Also, if $\mathcal{W}$ is the subset $\{\mathbf{0}\}$ of $\mathcal{V}$, then $\mathcal{W}$ is a vector space under the same operations as $\mathcal{V}$ (see Example 2 of Section 4.1). Therefore, $\mathcal{W} = \{\mathbf{0}\}$ is a subspace of $\mathcal{V}$.

Although the subspaces $\mathcal{V}$ and $\{\mathbf{0}\}$ of a vector space $\mathcal{V}$ are important, they occasionally complicate matters because they must be considered as special cases in proofs. The subspace $\mathcal{W} = \{\mathbf{0}\}$ is called the **trivial subspace** of $\mathcal{V}$. A vector space containing at least one nonzero vector has at least two distinct subspaces, the trivial subspace and the vector space itself. In fact, under the usual operations, $\mathbb{R}$ has only these two subspaces (see Exercise 16).

All subspaces of $\mathcal{V}$ other than $\mathcal{V}$ itself are called **proper subspaces** of $\mathcal{V}$. If we consider Examples 1 to 3 in the context of $\mathbb{R}^3$, we find at least four different types of subspaces of $\mathbb{R}^3$. These are the trivial subspace $\{[0,0,0]\} = \{\mathbf{0}\}$, subspaces like Example 2 that can be geometrically represented as a line (thus "resembling" $\mathbb{R}$), subspaces like Example 1 that can be represented as a plane (thus "resembling" $\mathbb{R}^2$), and the subspace $\mathbb{R}^3$ itself.[3] All but the last are proper subspaces. Later we will show that each subspace of $\mathbb{R}^3$ is, in fact, one of these four types. Similarly, we will show later that all subspaces of $\mathbb{R}^n$ "resemble" $\{\mathbf{0}\}, \mathbb{R}, \mathbb{R}^2, \mathbb{R}^3, \ldots, \mathbb{R}^{n-1}$, or $\mathbb{R}^n$.

Example 4

Consider the vector spaces (using ordinary function addition and scalar multiplication) in the following chain:

$$\begin{aligned} \mathcal{P}_0 \subset \mathcal{P}_1 \subset \mathcal{P}_2 \subset \cdots \quad &\subset \quad \mathcal{P} \\ &\subset \quad \{\text{differentiable real-valued functions on } \mathbb{R}\} \\ &\subset \quad \{\text{continuous real-valued functions on } \mathbb{R}\} \\ &\subset \quad \{\text{all real-valued functions on } \mathbb{R}\}. \end{aligned}$$

Some of these we encountered in Section 4.1, and the rest are discussed in Exercise 7 of this section. Each of these vector spaces is a proper subspace of every vector space after it in the chain (why?).

When Is a Subset a Subspace?

It is important to note that not every subset of a vector space is a subspace. A subset $\mathcal{S}$ of a vector space $\mathcal{V}$ fails to be a subspace of $\mathcal{V}$ if $\mathcal{S}$ does not satisfy the properties of a vector space in its own right or if $\mathcal{S}$ does not use the same operations as $\mathcal{V}$.

[3] Although some subspaces of $\mathbb{R}^3$ "resemble" $\mathbb{R}$ and $\mathbb{R}^2$ geometrically, note that $\mathbb{R}$ and $\mathbb{R}^2$ are not actually subspaces of $\mathbb{R}^3$ because they are not subsets of $\mathbb{R}^3$.

Example 5

Consider the first quadrant in $\mathbb{R}^2$ — that is, the set Ω of all 2-vectors of the form $[x,y]$ where $x \geq 0$ and $y \geq 0$. This subset Ω of $\mathbb{R}^2$ is not a vector space under the normal operations of $\mathbb{R}^2$ because it is not closed under scalar multiplication. (For example, $[3,4]$ is in Ω, but $-2 \cdot [3,4] = [-6,-8]$ is not in Ω.) Therefore, Ω cannot be a subspace of $\mathbb{R}^2$.

Example 6

Consider the vector space $\mathbb{R}$ under the usual operations. Let $\mathcal{W}$ be the subset $\mathbb{R}^+$. By Example 7 of Section 4.1, we know that $\mathcal{W}$ is a vector space under the unusual operations $\oplus$ and $\odot$, where $\oplus$ represents multiplication and $\odot$ represents exponentiation. Although $\mathcal{W}$ is a nonempty subset of $\mathbb{R}$ and is itself a vector space, $\mathcal{W}$ is not a subspace of $\mathbb{R}$ because $\mathcal{W}$ and $\mathbb{R}$ do not share the same operations.

The following theorem provides a shortcut for verifying that a (nonempty) subset $\mathcal{W}$ of a vector space is a subspace; if the closure properties hold for $\mathcal{W}$, then the remaining eight vector space properties automatically follow as well.

> **Theorem 4.2** Let $\mathcal{V}$ be a vector space, and let $\mathcal{W}$ be a nonempty subset of $\mathcal{V}$ using the same operations. Then $\mathcal{W}$ is a subspace of $\mathcal{V}$ if and only if $\mathcal{W}$ is closed under vector addition and scalar multiplication in $\mathcal{V}$.

Notice that this theorem applies only to *nonempty subsets* of a vector space. Even though the empty set is a subset of every vector space, it is not a subspace of any vector space because it does not contain an additive identity.

Proof. Since this is an "if and only if" statement, the proof has two parts. First we must show that if $\mathcal{W}$ is a subspace of $\mathcal{V}$, then it is closed under the two operations. Now, as a subspace, $\mathcal{W}$ is itself a vector space. Hence, the closure properties hold for $\mathcal{W}$ as they do for any vector space.

For the other part of the proof, we must show that if the closure properties hold for a nonempty subset $\mathcal{W}$ of $\mathcal{V}$, then $\mathcal{W}$ is itself a vector space under the operations in $\mathcal{V}$. That is, we must prove the remaining eight vector space properties hold for $\mathcal{W}$.

Properties (1), (2), (5), (6), (7), and (8) are all true in $\mathcal{W}$ because they are true in $\mathcal{V}$, a known vector space. That is, since these properties hold for all vectors in $\mathcal{V}$, they must be true for all vectors in its subset, $\mathcal{W}$. For example, to prove property (1) for $\mathcal{W}$, let $\mathbf{u}, \mathbf{v} \in \mathcal{W}$. Then,

$$\begin{aligned}
\underbrace{\mathbf{u}+\mathbf{v}}_{\text{addition in } \mathcal{W}} &= \underbrace{\mathbf{u}+\mathbf{v}}_{\text{addition in } \mathcal{V}} && \text{because } \mathcal{W} \text{ and } \mathcal{V} \text{ share the same operations} \\
&= \underbrace{\mathbf{v}+\mathbf{u}}_{\text{addition in } \mathcal{V}} && \text{because } \mathcal{V} \text{ is a vector space and property (1) holds} \\
&= \underbrace{\mathbf{v}+\mathbf{u}.}_{\text{addition in } \mathcal{W}} && \text{because } \mathcal{W} \text{ and } \mathcal{V} \text{ share the same operations}
\end{aligned}$$

Next we prove property (3), the existence of an additive identity in $\mathcal{W}$. Because $\mathcal{W}$ is nonempty, we can choose an element $\mathbf{w}_1$ from $\mathcal{W}$. Now $\mathcal{W}$ is closed under scalar multiplication, so $0\mathbf{w}_1$ is in $\mathcal{W}$. However, since this is the same operation as in $\mathcal{V}$, a known vector space, part (2) of Theorem 4.1 implies that $0\mathbf{w}_1 = \mathbf{0}$. Hence, $\mathbf{0}$ is in $\mathcal{W}$. Because $\mathbf{0}+\mathbf{v}=\mathbf{v}$ for all $\mathbf{v}$ in $\mathcal{V}$, it follows that $\mathbf{0}+\mathbf{w}=\mathbf{w}$ for all $\mathbf{w}$ in $\mathcal{W}$. Therefore, $\mathcal{W}$ contains the same additive identity that $\mathcal{V}$ has.

Finally, we must prove that property (4), the existence of additive inverses, holds for $\mathcal{W}$. Let $\mathbf{w} \in \mathcal{W}$. Then $\mathbf{w} \in \mathcal{V}$. Part (3) of Theorem 4.1 shows $(-1)\mathbf{w}$ is the additive inverse of $\mathbf{w}$ in $\mathcal{V}$. If we can show that this additive inverse is also in $\mathcal{W}$, we will be done. But since $\mathcal{W}$ is closed under scalar multiplication, $(-1)\mathbf{w} \in \mathcal{W}$. □

Checking for Subspaces in $\mathcal{M}_{nn}$ and $\mathbb{R}^n$

In the next three examples, we apply Theorem 4.2 to determine whether several subsets of $\mathcal{M}_{nn}$ and $\mathbb{R}^n$ are subspaces. Assume that $\mathcal{M}_{nn}$ and $\mathbb{R}^n$ have the usual operations.

Example 7

Consider $\mathcal{U}_n$, the set of upper triangular $n \times n$ matrices. Since $\mathcal{U}_n$ is nonempty, we may apply Theorem 4.2 to see whether $\mathcal{U}_n$ is a subspace of $\mathcal{M}_{nn}$. Closure of $\mathcal{U}_n$ under vector addition holds because the sum of any two $n \times n$ upper triangular matrices is again upper triangular. The closure property in $\mathcal{U}_n$ for scalar multiplication also holds, since any scalar multiple of an upper triangular matrix is again upper triangular. Hence, $\mathcal{U}_n$ is a subspace of $\mathcal{M}_{nn}$.

Similar arguments show that $\mathcal{L}_n$ (lower triangular $n \times n$ matrices) and $\mathcal{D}_n$ (diagonal $n \times n$ matrices) are also subspaces of $\mathcal{M}_{nn}$. ■

The subspace $\mathcal{D}_n$ of $\mathcal{M}_{nn}$ in Example 7 is the intersection of the subspaces $\mathcal{U}_n$ and $\mathcal{L}_n$. In fact, the intersection of subspaces of a vector space always produces a subspace under the same operations (see Exercise 18).

If either closure property fails to hold for a subset, it cannot be a subspace. For this reason, none of the following subsets of $\mathcal{M}_{nn}$, $n \geq 2$, is a subspace:

(1) the set of nonsingular $n \times n$ matrices

(2) the set of singular $n \times n$ matrices

(3) the set of $n \times n$ matrices in reduced row echelon form.

You should check that the closure property for vector addition fails in each case and that the closure property for scalar multiplication fails in (1) and (3).

Example 8

Let $\mathcal{Y}$ be the set of vectors in $\mathbb{R}^4$ of the form $[a,0,b,0]$, that is, 4-vectors whose second and fourth coordinates are zero. We prove that $\mathcal{Y}$ is a subspace of $\mathbb{R}^4$ by checking the closure properties.

To prove closure under vector addition, we must add two arbitrary elements of $\mathcal{Y}$ and check that the result has the correct form for a vector in $\mathcal{Y}$. Now, $[a,0,b,0]+[c,0,d,0] = [(a+c),0,(b+d),0]$. The second and fourth coordinates of the sum are zero, so $\mathcal{Y}$ is closed under addition. Similarly, we must prove closure under scalar multiplication. Now,

$k[a,0,b,0] = [ka,0,kb,0]$. Since the second and fourth coordinates of the product are zero, $\mathcal{Y}$ is closed under scalar multiplication. Hence, by Theorem 4.2, $\mathcal{Y}$ is a subspace of $\mathbb{R}^4$. ■

Example 9

Let $\mathcal{W}$ be the set of vectors in $\mathbb{R}^3$ of the form $\left[a,\ b,\ \frac{1}{2}a-2b\right]$, that is, 3-vectors whose third coordinate is half the first coordinate minus twice the second coordinate. We show that $\mathcal{W}$ is a subspace of $\mathbb{R}^3$ by checking the closure properties.

Checking closure under vector addition, we have

$$\left[a,\ b,\ \frac{1}{2}a-2b\right]+\left[c,\ d,\ \frac{1}{2}c-2d\right]=\left[a+c,\ b+d,\ \frac{1}{2}a-2b+\frac{1}{2}c-2d\right]$$
$$=\left[a+c,\ b+d,\ \frac{1}{2}(a+c)-2(b+d)\right],$$

which has the required form, since it equals $\left[A,\ B,\ \frac{1}{2}A-2B\right]$, where $A=a+c$ and $B=b+d$.

Checking closure under scalar multiplication, we get

$$k\left[a,\ b,\ \frac{1}{2}a-2b\right]=\left[ka,\ kb,\ k\left(\frac{1}{2}a-2b\right)\right]=\left[ka,\ kb,\ \frac{1}{2}(ka)-2(kb)\right],$$

which has the required form (why?).

Note that

$$\left[a,\ b,\ \frac{1}{2}a-2b\right]=a\left[1,0,\frac{1}{2}\right]+b[0,1,-2],$$

and so $\mathcal{W}$ consists of the set of all linear combinations of $\left[1,0,\frac{1}{2}\right]$ and $[0,1,-2]$. Geometrically, $\mathcal{W}$ is the plane in $\mathbb{R}^3$ through the origin containing the vectors $\left[1,0,\frac{1}{2}\right]$ and $[0,1,-2]$, shown in Figure 4.3. This plane is the set of all possible "destinations" using these two directions (starting from the origin). This is the type of subspace of $\mathbb{R}^3$ discussed in Example 1.

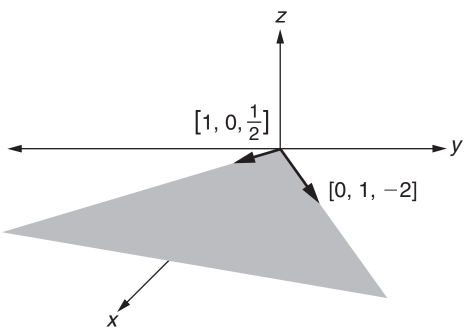


FIGURE 4.3

The plane through the origin containing $\left[1,0,\frac{1}{2}\right]$ and $[0,1,-2]$

■

The following subsets of $\mathbb{R}^n$ are not subspaces. In each case, at least one of the two closure properties fails. (Can you determine which properties?)

(1) The set of n-vectors whose first coordinate is nonnegative (in $\mathbb{R}^2$, this set is a half-plane)

(2) The set of unit n-vectors (in $\mathbb{R}^3$, this set is a sphere)

(3) For $n \geq 2$, the set of n-vectors with a zero in at least one coordinate (in $\mathbb{R}^3$, this set is the union of three planes)

(4) The set of n-vectors having all integer coordinates

(5) For $n \geq 2$, the set of all n-vectors whose first two coordinates add up to 3 (in $\mathbb{R}^2$, this is the line $x + y = 3$)

The subsets (2) and (5), which do not contain the additive identity $\mathbf{0}$ of $\mathbb{R}^n$, can quickly be disqualified as subspaces. In general,

> If a subset $\mathcal{S}$ of a vector space $\mathcal{V}$ does not contain the zero vector $\mathbf{0}$ of $\mathcal{V}$, then $\mathcal{S}$ is not a subspace of $\mathcal{V}$.

Checking for the presence of the additive identity is usually easy and thus is a fast way to show that certain subsets are not subspaces.

Linear Combinations Remain in a Subspace

As in Chapter 1, we define a **linear combination** of vectors in a general vector space to be a sum of scalar multiples of the vectors. The next theorem asserts that if a finite set of vectors is in a given subspace of a vector space, then so is any linear combination of those vectors.

> **Theorem 4.3** Let $\mathcal{W}$ be a subspace of a vector space $\mathcal{V}$, and let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be vectors in $\mathcal{W}$. Then, for any scalars $a_1, a_2, \ldots, a_n$, we have $a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n \in \mathcal{W}$.

Essentially, this theorem points out that a subspace is "closed under linear combinations." That is, when the vectors of a subspace are used to form linear combinations, all possible "destination vectors" remain in the subspace.

Proof. Suppose that $\mathcal{W}$ is a subspace of a vector space $\mathcal{V}$. We give a proof by induction on n.

Base Step: If $n = 1$, then we must show that if $\mathbf{v}_1 \in \mathcal{W}$ and a_1 is a scalar, then $a_1\mathbf{v}_1 \in \mathcal{W}$. But this is certainly true since the subspace $\mathcal{W}$ is closed under scalar multiplication.

Inductive Step: Assume that the theorem is true for any linear combination of n vectors in $\mathcal{W}$. We must prove the theorem holds for a linear combination of $n + 1$ vectors. Suppose $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n, \mathbf{v}_{n+1}$ are vectors in $\mathcal{W}$, and $a_1, a_2, \ldots, a_n, a_{n+1}$ are scalars. We must show that $a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n + a_{n+1}\mathbf{v}_{n+1} \in \mathcal{W}$. However, by the inductive hypothesis, we know that $a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n \in \mathcal{W}$. Also, $a_{n+1}\mathbf{v}_{n+1} \in \mathcal{W}$, since $\mathcal{W}$ is closed under scalar multiplication. But since $\mathcal{W}$ is also closed under addition, the sum of any two vectors in $\mathcal{W}$ is again in $\mathcal{W}$, so $(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n) + (a_{n+1}\mathbf{v}_{n+1}) \in \mathcal{W}$. □

Example 10

In Example 9, we found that the set $\mathcal{W}$ of all vectors of the form $\left[a,\ b,\ \frac{1}{2}a - 2b\right]$ is a subspace of $\mathbb{R}^3$. In particular, $[1, 0, \frac{1}{2}]$ and $[0, 1, -2]$ are in $\mathcal{W}$. By Theorem 4.3, any linear combination of these vectors is also in $\mathcal{W}$. For example, $6[1, 0, \frac{1}{2}] - 5[0, 1, -2] = [6, -5, 13]$ and $-4[1, 0, \frac{1}{2}] + 2[0, 1, -2] = [-4, 2, -6]$ are both in $\mathcal{W}$. Of course, this makes sense geometrically, since $\mathcal{W}$ is a plane through the origin, and any linear combination of vectors in such a plane remains in that plane. ■

An Eigenspace Is a Subspace

We conclude this section by noting that any eigenspace of an $n \times n$ matrix is a subspace of $\mathbb{R}^n$. (In fact, this is why the word "space" appears in the term "eigenspace.")

Theorem 4.4 Let $\mathbf{A}$ be an $n \times n$ matrix, and let λ be an eigenvalue for $\mathbf{A}$, having eigenspace E_λ. Then E_λ is a subspace of $\mathbb{R}^n$.

Proof. Let λ be an eigenvalue for an $n \times n$ matrix $\mathbf{A}$. By definition, the eigenspace E_λ of λ is the set of all n-vectors $\mathbf{X}$ having the property that $\mathbf{AX} = \lambda\mathbf{X}$, including the zero n-vector. We will use Theorem 4.2 to show that E_λ is a subspace of $\mathbb{R}^n$.

Since $\mathbf{0} \in E_\lambda$, E_λ is a nonempty subset of $\mathbb{R}^n$. We must show that E_λ is closed under addition and scalar multiplication.

Let $\mathbf{X}_1, \mathbf{X}_2$ be any two vectors in E_λ. To show that $\mathbf{X}_1 + \mathbf{X}_2 \in E_\lambda$, we need to verify that $\mathbf{A}(\mathbf{X}_1 + \mathbf{X}_2) = \lambda(\mathbf{X}_1 + \mathbf{X}_2)$. But, $\mathbf{A}(\mathbf{X}_1 + \mathbf{X}_2) = \mathbf{AX}_1 + \mathbf{AX}_2 = \lambda\mathbf{X}_1 + \lambda\mathbf{X}_2 = \lambda(\mathbf{X}_1 + \mathbf{X}_2)$.

Similarly, let $\mathbf{X}$ be a vector in E_λ, and let c be a scalar. We must show that $c\mathbf{X} \in E_\lambda$. But, $\mathbf{A}(c\mathbf{X}) = c(\mathbf{AX}) = c(\lambda\mathbf{X}) = \lambda(c\mathbf{X})$, and so $c\mathbf{X} \in E_\lambda$. Hence, E_λ is a subspace of $\mathbb{R}^n$. □

Example 11

Consider

$$\mathbf{A} = \begin{bmatrix} 16 & -4 & -2 \\ 3 & 3 & -6 \\ 2 & -8 & 11 \end{bmatrix}.$$

Computing $|x\mathbf{I}_3 - \mathbf{A}|$ produces $p_{\mathbf{A}}(x) = x^3 - 30x^2 + 225x = x(x-15)^2$. Solving $(0\mathbf{I}_3 - \mathbf{A})\mathbf{X} = \mathbf{0}$ yields $E_0 = \{c[1, 3, 2] \mid c \in \mathbb{R}\}$. Thus, the eigenspace for $\lambda_1 = 0$ is the subspace of $\mathbb{R}^3$ from Example 2. Similarly, solving $(15\mathbf{I}_3 - \mathbf{A}) = \mathbf{0}$ gives $E_{15} = \{a[4, 1, 0] + b[2, 0, 1] \mid a, b \in \mathbb{R}\}$. By Theorem 4.4, E_{15} is also a subspace of $\mathbb{R}^3$. Although it is not obvious, E_{15} is the same subspace of $\mathbb{R}^3$ that we studied in Examples 9 and 10 (see Exercises 14(b) and 14(c)). ■

New Vocabulary

linear combination (of vectors in a vector space)
proper subspace(s)
subspace
trivial subspace

Highlights

- A subset of a vector space is a subspace if it is a vector space itself under the same operations.
- The subset $\{\mathbf{0}\}$ is a trivial subspace of any vector space.
- Any subspace of a vector space $\mathcal{V}$ other than $\mathcal{V}$ itself is considered a proper subspace.
- Familiar proper nontrivial subspaces of $\mathbb{R}^3$ are any line through the origin, any plane through the origin.
- Familiar proper subspaces of the real-valued functions on $\mathbb{R}$ are $\mathcal{P}_n, \mathcal{P}$, all differentiable real-valued functions on $\mathbb{R}$, all continuous real-valued functions on $\mathbb{R}$.
- Familiar proper subspaces of $\mathcal{M}_{nn}$ are $\mathcal{U}_n, \mathcal{L}_n, \mathcal{D}_n$, the symmetric $n \times n$ matrices, the skew-symmetric $n \times n$ matrices.
- A nonempty subset of a vector space is a subspace if it is closed under vector addition and scalar multiplication.
- If a subset of a vector space does not contain the zero vector, it cannot be a subspace.
- If a set of vectors is in a subspace, then any (finite) linear combination of those vectors is also in the subspace.
- If λ is an eigenvalue for an $n \times n$ matrix $\mathbf{A}$, then E_λ (eigenspace for λ) is a subspace of $\mathbb{R}^n$.
- The intersection of subspaces is a subspace.

EXERCISES FOR SECTION 4.2

Note: From this point onward in the book, use a calculator or available software packages to avoid tedious calculations.

1. Prove or disprove that each given subset of $\mathbb{R}^2$ is a subspace of $\mathbb{R}^2$ under the usual vector operations. (In these problems, a and b represent arbitrary real numbers.)

★**(a)** The set of unit 2-vectors

(b) The set of 2-vectors of the form $[1, a]$

★(c) The set of 2-vectors of the form $[a, 2a]$

(d) The set of 2-vectors having a zero in at least one coordinate

★(e) The set $\{[1, 2]\}$

(f) The set of 2-vectors whose second coordinate is zero

★(g) The set of 2-vectors of the form $[a, b]$, where $|a| = |b|$

(h) The set of points in the plane lying on the line $y = -3x$

(i) The set of points in the plane lying on the line $y = 7x - 5$

★(j) The set of points lying on the parabola $y = x^2$

(k) The set of points in the plane lying above the line $y = 2x - 5$

★(l) The set of points in the plane lying inside the circle of radius 1 centered at the origin

2. Prove or disprove that each given subset of $\mathcal{M}_{22}$ is a subspace of $\mathcal{M}_{22}$ under the usual matrix operations. (In these problems, a and b represent arbitrary real numbers.)

★(a) The set of matrices of the form $\begin{bmatrix} a & -a \\ b & 0 \end{bmatrix}$

(b) The set of 2×2 matrices that have at least one row of zeroes

★(c) The set of symmetric 2×2 matrices

(d) The set of nonsingular 2×2 matrices

★(e) The set of 2×2 matrices having the sum of all entries zero

(f) The set of 2×2 matrices having trace zero (Recall that the *trace* of a square matrix is the sum of the main diagonal entries.)

★(g) The set of 2×2 matrices $\mathbf{A}$ such that $\mathbf{A}\begin{bmatrix} 1 & 3 \\ -2 & -6 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

★(h) The set of 2×2 matrices having the product of all entries zero

3. Prove or disprove that each given subset of $\mathcal{P}_5$ is a subspace of $\mathcal{P}_5$ under the usual operations.

★(a) $\{\mathbf{p} \in \mathcal{P}_5 \,|\,$ the coefficient of the first-degree term of $\mathbf{p}$ equals the coefficient of the fifth-degree term of $\mathbf{p}\}$

★(b) $\{\mathbf{p} \in \mathcal{P}_5 \,|\, \mathbf{p}(3) = 0\}$

(c) $\{\mathbf{p} \in \mathcal{P}_5 \,|\,$ the sum of the coefficients of $\mathbf{p}$ is zero$\}$

(d) $\{\mathbf{p} \in \mathcal{P}_5 \,|\, \mathbf{p}(3) = \mathbf{p}(5)\}$

★(e) $\{\mathbf{p} \in \mathcal{P}_5 \,|\, \mathbf{p}$ is an odd-degree polynomial (highest-order nonzero term has odd degree)$\}$

(f) $\{\mathbf{p} \in \mathcal{P}_5 \,|\, \mathbf{p}$ has a relative maximum at $x = 0\}$

★(g) $\{\mathbf{p} \in \mathcal{P}_5 \,|\, \mathbf{p}'(4) = 0$, where $\mathbf{p}'$ is the derivative of $\mathbf{p}\}$

(h) $\{\mathbf{p} \in \mathcal{P}_5 \,|\, \mathbf{p}'(4) = 1$, where $\mathbf{p}'$ is the derivative of $\mathbf{p}\}$

4. Show that the set of vectors of the form $[a, b, 0, c, a - 2b + c]$ in $\mathbb{R}^5$ forms a subspace of $\mathbb{R}^5$ under the usual operations.

5. Show that the set of vectors of the form $[2a - 3b, a - 5c, a, 4c - b, c]$ in $\mathbb{R}^5$ forms a subspace of $\mathbb{R}^5$ under the usual operations.

6. **(a)** Prove that the set of all 3-vectors orthogonal to $[1, -1, 4]$ forms a subspace of $\mathbb{R}^3$.

   (b) Is the subspace from part (a) all of $\mathbb{R}^3$, a plane passing through the origin in $\mathbb{R}^3$, or a line passing through the origin in $\mathbb{R}^3$?

7. Show that each of the following sets is a subspace of the vector space of all real-valued functions on the given domain, under the usual operations of function addition and scalar multiplication:

   (a) The set of continuous real-valued functions with domain $\mathbb{R}$

   (b) The set of differentiable real-valued functions with domain $\mathbb{R}$

   (c) The set of all real-valued functions $\mathbf{f}$ defined on the interval $[0, 1]$ such that $\mathbf{f}\left(\frac{1}{2}\right) = 0$ (Compare this vector space with the set in Example 9 of Section 4.1.)

   (d) The set of all continuous real-valued functions $\mathbf{f}$ defined on the interval $[0, 1]$ such that $\int_0^1 \mathbf{f}(x)\, dx = 0$

8. Let $\mathcal{W}$ be the set of differentiable real-valued functions $y = \mathbf{f}(x)$ defined on $\mathbb{R}$ that satisfy the differential equation $3(dy/dx) - 2y = 0$. Show that, under the usual function operations, $\mathcal{W}$ is a subspace of the vector space of all differentiable real-valued functions. (Do not forget to show $\mathcal{W}$ is nonempty.)

9. Show that the set $\mathcal{W}$ of solutions to the differential equation $y'' + 2y' - 9y = 0$ is a subspace of the vector space of all twice-differentiable real-valued functions defined on $\mathbb{R}$. (Do not forget to show that $\mathcal{W}$ is nonempty.)

10. Prove that the set of discontinuous real-valued functions defined on $\mathbb{R}$ (for example, $\mathbf{f}(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases}$) with the usual function operations is not a subspace of the vector space of all real-valued functions with domain $\mathbb{R}$.

11. Let $\mathbf{A}$ be a fixed $n \times n$ matrix, and let $\mathcal{W}$ be the subset of $\mathcal{M}_{nn}$ of all $n \times n$ matrices that commute with $\mathbf{A}$ under multiplication (that is, $\mathbf{B} \in \mathcal{W}$ if and only if $\mathbf{AB} = \mathbf{BA}$). Show that $\mathcal{W}$ is a subspace of $\mathcal{M}_{nn}$ under the usual vector space operations. (Do not forget to show that $\mathcal{W}$ is nonempty.)

12. **(a)** A careful reading of the proof of Theorem 4.2 reveals that only closure under scalar multiplication (not closure under addition) is sufficient to prove the remaining eight vector space properties for $\mathcal{W}$. Explain, nevertheless, why closure under addition is a necessary condition for $\mathcal{W}$ to be a subspace of $\mathcal{V}$.

    (b) Show that the set of singular $n \times n$ matrices is closed under scalar multiplication in $\mathcal{M}_{nn}$.

(c) Use parts (a) and (b) to determine which of the eight vector space properties are true for the set of singular $n \times n$ matrices.

(d) Show that the set of singular $n \times n$ matrices is not closed under vector addition and hence is not a subspace of $\mathcal{M}_{nn}$ $(n \geq 2)$.

⋆(e) Is the set of nonsingular $n \times n$ matrices closed under scalar multiplication? Why or why not?

13. **(a)** Prove that the set of all points lying on a line passing through the origin in $\mathbb{R}^2$ is a subspace of $\mathbb{R}^2$ (under the usual operations).

(b) Prove that the set of all points in $\mathbb{R}^2$ lying on a line not passing through the origin does not form a subspace of $\mathbb{R}^2$ (under the usual operations).

14. Let $\mathcal{W}$ be the subspace from Examples 9 and 10, and let $\mathbf{A}$ and E_{15} be as given in Example 11.

(a) Use Theorem 4.2 to prove directly that E_{15} is a subspace of $\mathbb{R}^3$.

(b) Show that $E_{15} \subseteq \mathcal{W}$ by proving that every vector in E_{15} has the form $[a, b, \frac{1}{2}a - 2b]$.

(c) Prove that $\mathcal{W} \subseteq E_{15}$ by showing that every nonzero vector of the form $[a, b, \frac{1}{2}a - 2b]$ is an eigenvector for $\mathbf{A}$ corresponding to $\lambda_2 = 15$.

⋆15. Suppose $\mathbf{A}$ is an $n \times n$ matrix and $\lambda \in \mathbb{R}$ is *not* an eigenvalue for $\mathbf{A}$. Determine exactly which vectors are in $S = \{\mathbf{X} \in \mathbb{R}^n \mid \mathbf{AX} = \lambda\mathbf{X}\}$. Is this set a subspace of $\mathbb{R}^n$? Why or why not?

16. Prove that $\mathbb{R}$ (under the usual operations) has no subspaces except $\mathbb{R}$ and $\{\mathbf{0}\}$. (Hint: Let $\mathcal{V}$ be a nontrivial subspace of $\mathbb{R}$, and show that $\mathcal{V} = \mathbb{R}$.)

17. Let $\mathcal{W}$ be a subspace of a vector space $\mathcal{V}$. Show that the set $\mathcal{W}' = \{\mathbf{v} \in \mathcal{V} \mid \mathbf{v} \notin \mathcal{W}\}$ is not a subspace of $\mathcal{V}$.

18. Let $\mathcal{V}$ be a vector space, and let $\mathcal{W}_1$ and $\mathcal{W}_2$ be subspaces of $\mathcal{V}$. Prove that $\mathcal{W}_1 \cap \mathcal{W}_2$ is a subspace of $\mathcal{V}$. (Do not forget to show $\mathcal{W}_1 \cap \mathcal{W}_2$ is nonempty.)

19. Let $\mathcal{V}$ be any vector space, and let $\mathcal{W}$ be a nonempty subset of $\mathcal{V}$.

(a) Prove that $\mathcal{W}$ is a subspace of $\mathcal{V}$ if and only if $a\mathbf{w}_1 + b\mathbf{w}_2$ is an element of $\mathcal{W}$ for every $a, b \in \mathbb{R}$ and every $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{W}$. (Hint: For one half of the proof, first consider the case where $a = b = 1$ and then the case where $b = 0$ and a is arbitrary.)

(b) Prove that $\mathcal{W}$ is a subspace of $\mathcal{V}$ if and only if $a\mathbf{w}_1 + \mathbf{w}_2$ is an element of $\mathcal{W}$ for every real number a and every $\mathbf{w}_1$ and $\mathbf{w}_2$ in $\mathcal{W}$.

20. Let $\mathcal{W}$ be a nonempty subset of a vector space $\mathcal{V}$, and suppose every linear combination of vectors in $\mathcal{W}$ is also in $\mathcal{W}$. Prove that $\mathcal{W}$ is a subspace of $\mathcal{V}$. (This is the converse of Theorem 4.3.)

21. Let λ be an eigenvalue for an $n \times n$ matrix $\mathbf{A}$. Show that if $\mathbf{X}_1, \ldots, \mathbf{X}_k$ are eigenvectors for $\mathbf{A}$ corresponding to λ, then any linear combination of $\mathbf{X}_1, \ldots, \mathbf{X}_k$ is in E_λ.

★22. True or False:

(a) A nonempty subset $\mathcal{W}$ of a vector space $\mathcal{V}$ is always a subspace of $\mathcal{V}$ under the same operations as those in $\mathcal{V}$.

(b) Every vector space has at least one subspace.

(c) Any plane $\mathcal{W}$ in $\mathbb{R}^3$ is a subspace of $\mathbb{R}^3$ (under the usual operations).

(d) The set of all lower triangular 5×5 matrices is a subspace of $\mathcal{M}_{55}$ (under the usual operations).

(e) The set of all vectors of the form $[0, a, b, 0]$ is a subspace of $\mathbb{R}^4$ (under the usual operations).

(f) If a subset $\mathcal{W}$ of a vector space $\mathcal{V}$ contains the zero vector $\mathbf{0}$ of $\mathcal{V}$, then $\mathcal{W}$ must be a subspace of $\mathcal{V}$ (under the same operations).

(g) Any linear combination of vectors from a subspace $\mathcal{W}$ of a vector space $\mathcal{V}$ must also be in $\mathcal{W}$.

(h) If λ is an eigenvalue for a 4×4 matrix $\mathbf{A}$, then E_λ is a subspace of $\mathbb{R}^4$.

4.3 SPAN

In this section, we study the concept of linear combinations in more depth. We show that the set of all linear combinations of the vectors in a subset S of $\mathcal{V}$ forms an important subspace of $\mathcal{V}$, called the span of S in $\mathcal{V}$.

Finite Linear Combinations

In Section 4.2, we introduced linear combinations of vectors in a general vector space. We now extend the concept of linear combination to include the possibility of forming sums of scalar multiples from infinite, as well as finite, sets.

Definition Let S be a nonempty (possibly infinite) subset of a vector space $\mathcal{V}$. Then a vector $\mathbf{v}$ in $\mathcal{V}$ is a **(finite) linear combination of the vectors in** S if and only if there exists a *finite* subset $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of S such that $\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n$ for some real numbers $a_1, \ldots, a_n$.

Examples 1 and 2 below involve a finite set S, while Examples 3 and 4 use an infinite set S. In all these examples, however, only a *finite* number of vectors from S are used at any given time to form linear combinations.

Example 1

Consider the subset $\mathcal{S} = \{[1,-1,0],[1,0,2],[0,-2,5]\}$ of $\mathbb{R}^3$. The vector $[1,-2,-2]$ is a linear combination of the vectors in $\mathcal{S}$ according to the definition, because $[1,-2,-2] = 2[1,-1,0] + (-1)[1,0,2]$. In this case, the (finite) subset of $\mathcal{S}$ used (from the definition) is $\{[1,-1,0],[1,0,2]\}$. However, we could have used all of $\mathcal{S}$ to form the linear combination by placing a zero coefficient in front of the remaining vector $[0,-2,5]$. That is, $[1,-2,-2] = 2[1,-1,0] + (-1)[1,0,2] + 0[0,-2,5]$.

We see from Example 1 that if $\mathcal{S}$ is a *finite* subset of a vector space $\mathcal{V}$, any linear combination $\mathbf{v}$ formed using *some* of the vectors in $\mathcal{S}$ can always be formed using *all* the vectors in $\mathcal{S}$ by placing zero coefficients on the remaining vectors.

A linear combination formed from a set $\{\mathbf{v}\}$ containing a single vector is just a scalar multiple $a\mathbf{v}$ of $\mathbf{v}$, as we see in the next example.

Example 2

Let $\mathcal{S} = \{[1,-2,7]\}$, a subset of $\mathbb{R}^3$ containing a single element. Then the only linear combinations that can be formed from $\mathcal{S}$ are scalar multiples of $[1,-2,7]$, such as $[3,-6,21]$ and $[-4,8,-28]$.

Example 3

Consider $\mathcal{P}$, the vector space of polynomials with real coefficients, and let $\mathcal{S} = \{1, x^2, x^4, \ldots\}$, the infinite subset of $\mathcal{P}$ consisting of all nonnegative even powers of x (since $x^0 = 1$). We can form linear combinations of vectors in $\mathcal{S}$ using any finite subset of $\mathcal{S}$. For example, $\mathbf{p}(x) = 7x^8 - (1/4)x^4 + 10$ is a linear combination formed from $\mathcal{S}$ because it is a sum of scalar multiples of elements of a finite subset $\{x^8,\ x^4,\ 1\}$ of $\mathcal{S}$. In fact, the possible linear combinations of vectors in $\mathcal{S}$ are precisely the polynomials involving only even powers of x.

Notice that we cannot use all of the elements in an infinite set $\mathcal{S}$ when forming a linear combination because an "infinite" sum would result. This is why a linear combination is frequently called a *finite* linear combination in order to stress that only a finite number of vectors are combined at any time.

Example 4

Let $\mathcal{S} = \mathcal{U}_2 \cup \mathcal{L}_2$, an infinite subset of $\mathcal{M}_{22}$. (Recall that $\mathcal{U}_2$ and $\mathcal{L}_2$ are, respectively, the sets of upper and lower triangular 2×2 matrices.) The matrix $\mathbf{A} = \begin{bmatrix} 2 & 3 \\ -1 & \frac{1}{2} \end{bmatrix}$ is a linear combination of the elements in $\mathcal{S}$, because

$$\mathbf{A} = \underbrace{\frac{1}{2}\begin{bmatrix} 4 & 6 \\ 0 & 1 \end{bmatrix}}_{\text{in } \mathcal{U}_2} + \underbrace{(-1)\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}}_{\text{in } \mathcal{L}_2}.$$

But there are many other ways to express $\mathbf{A}$ as a finite linear combination of the elements in $\mathcal{S}$. We can add more elements from $\mathcal{S}$ with zero coefficients, as in Example 1, but in this case there are further possibilities. For example,

$$\mathbf{A} = 2\underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}}_{\text{in } \mathcal{U}_2 \text{ and } \mathcal{L}_2} + 3\underbrace{\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}}_{\text{in } \mathcal{U}_2} + (-1)\underbrace{\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}}_{\text{in } \mathcal{L}_2} + \frac{1}{2}\underbrace{\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}}_{\text{in } \mathcal{U}_2 \text{ and } \mathcal{L}_2}.$$

Definition of the Span of a Set

Definition Let $\mathcal{S}$ be a nonempty subset of a vector space $\mathcal{V}$. Then the **span** of $\mathcal{S}$ in $\mathcal{V}$ is the set of all possible (finite) linear combinations of the vectors in $\mathcal{S}$. We use the notation span($\mathcal{S}$) to denote the span of $\mathcal{S}$ in $\mathcal{V}$.

The span of a set $\mathcal{S}$ is a generalization of the row space of a matrix; each is just the set of all linear combinations of a set of vectors. In fact, from this definition:

The span of the set of rows of a matrix is precisely the row space of the matrix.

We now consider some examples of the span of a subset.

Example 5

In Example 3, we found that for $\mathcal{S} = \{1, x^2, x^4, \ldots\}$ in $\mathcal{P}$, span($\mathcal{S}$) is the set of all polynomials containing only even-degree terms. This consists of all the "destinations" obtainable by traveling in the "directions" $1, x^2, x^4, \ldots$, etc. Thus, we can visualize span($\mathcal{S}$) as the set of "possible destinations" in the same sense as the row space is the set of "possible destinations" obtainable from the rows of a given matrix. Notice that we may only use a finite number of the possible "directions" to obtain a given "destination." That is, span($\mathcal{S}$) only contains polynomials, not infinite series.

Example 6

Let $\mathcal{S} = \mathcal{U}_2 \cup \mathcal{L}_2$ in $\mathcal{M}_{22}$, as in Example 4. Then span($\mathcal{S}$) $= \mathcal{M}_{22}$ because every 2×2 matrix can be expressed as a finite linear combination of upper and lower triangular matrices, as follows:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = a\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + d\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Notice that the span of a given set often (but not always) contains many more vectors than the set itself.

Example 6 shows that, when $\mathcal{S} = \mathcal{U}_2 \cup \mathcal{L}_2$ and $\mathcal{V} = \mathcal{M}_{22}$, every vector in $\mathcal{V}$ is a linear combination of vectors in $\mathcal{S}$. That is, span($\mathcal{S}$) $= \mathcal{V}$ itself. When this happens, we say that $\mathcal{V}$ is **spanned by** $\mathcal{S}$ or that $\mathcal{S}$ **spans** $\mathcal{V}$. Here, we are using span as a *verb* to indicate that the span (noun) of a set $\mathcal{S}$ equals $\mathcal{V}$. Thus, $\mathcal{M}_{22}$ is spanned (verb) by $\mathcal{U}_2 \cup \mathcal{L}_2$, since the span (noun) of $\mathcal{U}_2 \cup \mathcal{L}_2$ is $\mathcal{M}_{22}$.

Example 7

Note that $\mathbb{R}^3$ is spanned by $S_1 = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$, since span($S_1$) $= \mathbb{R}^3$. That is, every 3-vector can be expressed as a linear combination of $\mathbf{i}, \mathbf{j}$, and $\mathbf{k}$ (why?). However, $\mathbb{R}^3$ is not spanned by the smaller set $S_2 = \{\mathbf{i}, \mathbf{j}\}$, since span($S_2$) is the xy-plane in $\mathbb{R}^3$ (why?). More generally, $\mathbb{R}^n$ is spanned by the set of standard unit vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$. Note that no proper subset of $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ will span $\mathbb{R}^n$.

Span($\mathcal{S}$) Is the Minimal Subspace Containing $\mathcal{S}$

The next theorem completely characterizes the span.

Theorem 4.5 Let $\mathcal{S}$ be a nonempty subset of a vector space $\mathcal{V}$. Then:

(1) $\mathcal{S} \subseteq$ span($\mathcal{S}$).

(2) Span($\mathcal{S}$) is a subspace of $\mathcal{V}$ (under the same operations as $\mathcal{V}$).

(3) If $\mathcal{W}$ is a subspace of $\mathcal{V}$ with $\mathcal{S} \subseteq \mathcal{W}$, then span($\mathcal{S}$) $\subseteq \mathcal{W}$.

(4) Span($\mathcal{S}$) is the smallest subspace of $\mathcal{V}$ containing $\mathcal{S}$.

Proof. Part (1): We must show that each vector $\mathbf{w} \in \mathcal{S}$ is also in span($\mathcal{S}$). But if $\mathbf{w} \in \mathcal{S}$, then $\mathbf{w} = 1\mathbf{w}$ is a sum of scalar multiples from the subset $\{\mathbf{w}\}$ of $\mathcal{S}$. Hence, $\mathbf{w} \in$ span($\mathcal{S}$).

Part (2): Since $\mathcal{S}$ is nonempty, part (1) shows that span($\mathcal{S}$) is nonempty. Therefore, by Theorem 4.2, span($\mathcal{S}$) is a subspace of $\mathcal{V}$ if we can prove the closure properties hold for span($\mathcal{S}$).

First, let us verify closure under scalar multiplication. Let $\mathbf{v}$ be in span($\mathcal{S}$), and let c be a scalar. We must show that $c\mathbf{v} \in$ span($\mathcal{S}$). Now, since $\mathbf{v} \in$ span($\mathcal{S}$), a finite subset $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of $\mathcal{S}$ and real numbers $a_1, \ldots, a_n$ exist such that $\mathbf{v} = \sum_{i=1}^{n} a_i\mathbf{v}_i$. Then,

$$c\mathbf{v} = c(a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) = (ca_1)\mathbf{v}_1 + \cdots + (ca_n)\mathbf{v}_n.$$

Hence, $c\mathbf{v}$ is a linear combination of the finite subset $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of $\mathcal{S}$, and so $c\mathbf{v} \in$ span($\mathcal{S}$).

Finally, we show that span($\mathcal{S}$) is closed under vector addition. First we will consider the case in which $\mathcal{S}$ is a finite set. Thus, we suppose that $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. Let $\mathbf{x}$ and $\mathbf{y}$ be two vectors in span($\mathcal{S}$). Hence, there exist real numbers $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ such that

$$\mathbf{x} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n \text{ and } \mathbf{y} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_n\mathbf{v}_n.$$

Therefore,

$$\mathbf{x} + \mathbf{y} = (a_1 + b_1)\mathbf{v}_1 + (a_2 + b_2)\mathbf{v}_2 + \cdots + (a_n + b_n)\mathbf{v}_n,$$

and we have expressed $\mathbf{x}+\mathbf{y}$ as a linear combination of vectors in $\mathcal{S}$. Hence, $\mathbf{x}+\mathbf{y} \in \text{span}(\mathcal{S})$.

The proof of closure under addition in the case in which $\mathcal{S}$ has an infinite number of elements is identical in concept to the finite case. However, the linear combinations for $\mathbf{x}$ and $\mathbf{y}$ might now be formed using two different finite subsets of vectors from $\mathcal{S}$. This complication is remedied by uniting these two subsets into one common finite subset of $\mathcal{S}$ that we use to form the linear combinations for $\mathbf{x}$ and $\mathbf{y}$. Then we place a coefficient of zero in front of any vector in the union that is unneeded when forming the desired linear combination for $\mathbf{x}$, and similarly for $\mathbf{y}$. You are asked to complete the details for this part of the proof in Exercise 28.

Part (3): This part asserts that if $\mathcal{S}$ is a subset of a subspace $\mathcal{W}$, then any (finite) linear combination from $\mathcal{S}$ is also in $\mathcal{W}$. This is merely a rewording of Theorem 4.3 using the "span" concept. The fact that span($\mathcal{S}$) cannot contain vectors outside $\mathcal{W}$ is illustrated in Figure 4.4.

Part (4): This is merely a summary of the other three parts. Parts (1) and (2) assert that span($\mathcal{S}$) is a subspace of $\mathcal{V}$ containing $\mathcal{S}$. But part (3) shows that span($\mathcal{S}$) is the smallest such subspace because span($\mathcal{S}$) must be a subset of, and hence smaller than, any other subspace of $\mathcal{V}$ that contains $\mathcal{S}$. □

Theorem 4.5 implies that span($\mathcal{S}$) is created by appending to $\mathcal{S}$ precisely those vectors needed to make the closure properties hold. In fact, the whole idea behind span is to "close up" a subset of a vector space to create a subspace.

Example 8

Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be any two vectors in $\mathbb{R}^4$. Then, by Theorem 4.5, span($\{\mathbf{v}_1, \mathbf{v}_2\}$) is the smallest subspace of $\mathbb{R}^4$ containing $\mathbf{v}_1$ and $\mathbf{v}_2$. In particular, if $\mathbf{v}_1 = [1,3,-2,5]$ and $\mathbf{v}_2 = [0,-4,3,-1]$, then span($\{\mathbf{v}_1, \mathbf{v}_2\}$) is the subspace of $\mathbb{R}^4$ consisting of all vectors of the form

$$a[1,3,-2,5] + b[0,-4,3,-1] = [a,\ 3a-4b,\ -2a+3b,\ 5a-b].$$

No smaller subspace of $\mathbb{R}^4$ contains $\mathbf{v}_1$ and $\mathbf{v}_2$. ■

The following useful result is left for you to prove in Exercise 21.

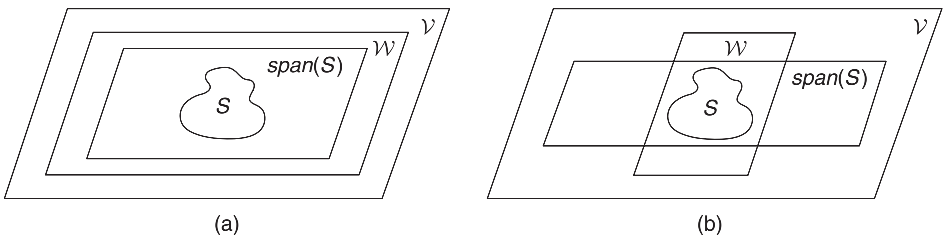


FIGURE 4.4

(a) Situation that *must* occur if $\mathcal{W}$ is a subspace containing $\mathcal{S}$; (b) situation that *cannot* occur if $\mathcal{W}$ is a subspace containing $\mathcal{S}$

Corollary 4.6 Let $\mathcal{V}$ be a vector space, and let S_1 and S_2 be subsets of $\mathcal{V}$ with $S_1 \subseteq S_2$. Then $\text{span}(S_1) \subseteq \text{span}(S_2)$.

Simplifying Span(S) using Row Reduction

Our next goal is to find a simplified form for the vectors in the span of a given set S. The fact that span is a generalization of the row space concept suggests that we can use results from Chapter 2 involving row spaces to help us compute and simplify span(S). If we form the matrix $\mathbf{A}$ whose rows are the vectors in S, the rows of the reduced row echelon form of $\mathbf{A}$ often give a simpler expression for span(S), since row equivalent matrices have the same row space. Hence, we have the following:

Method for Simplifying Span(S) Using Row Reduction (Simplified Span Method)

Suppose that S is a finite subset of $\mathbb{R}^n$ containing k vectors, with $k \geq 2$.

Step 1: Form a $k \times n$ matrix $\mathbf{A}$ by using the vectors in S as the rows of $\mathbf{A}$. (Thus, span(S) is the row space of $\mathbf{A}$).

Step 2: Let $\mathbf{C}$ be the reduced row echelon form matrix for $\mathbf{A}$.

Step 3: Then, a simplified form for span(S) is given by the set of all linear combinations of the *nonzero* rows of $\mathbf{C}$.

Example 9

Let S be the subset $\{[1,4,-1,-5],[2,8,5,4],[-1,-4,2,7],[6,24,-1,-20]\}$ of $\mathbb{R}^4$. By definition, span(S) is the set of all vectors of the form

$$a[1,4,-1,-5]+b[2,8,5,4]+c[-1,-4,2,7]+d[6,24,-1,-20]$$

for $a,b,c,d \in \mathbb{R}$. We want to use the Simplified Span Method to find a simplified form for the vectors in span(S). We first create

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & -1 & -5 \\ 2 & 8 & 5 & 4 \\ -1 & -4 & 2 & 7 \\ 6 & 24 & -1 & -20 \end{bmatrix},$$

whose rows are the vectors in S. Then, span(S) is the row space of $\mathbf{A}$; that is, the set of all linear combinations of the rows of $\mathbf{A}$.

Next, we simplify the form of the row space of $\mathbf{A}$ by obtaining its reduced row echelon form matrix

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

By Theorem 2.8, the row space of $\mathbf{A}$ is the same as the row space of $\mathbf{C}$, which is the set of all 4-vectors of the form

$$a[1,4,0,-3] + b[0,0,1,2] = [a,\ 4a,\ b, -3a + 2b].$$

Therefore, span$(S) = \{[a, 4a, b, -3a + 2b] \mid a, b \in \mathbb{R}\}$, a subspace of $\mathbb{R}^4$. Note, for example, that the vector $[3, 12, -2, -13]$ is in span(S) $(a = 3, b = -2)$. However, the vector $[-2, -8, 4, 6]$ is not in span(S) because the following system has no solutions:

$$\begin{cases} a = -2 \\ 4a = -8 \\ b = 4 \\ -3a + 2b = 6 \end{cases}.$$

■

Example 10

Recall that the eigenspace E_{15} for the matrix $\mathbf{A}$ in Example 11 in Section 4.2 is $E_{15} = \{a[4,1,0] + b[2,0,1] \mid a, b \in \mathbb{R}\}$. Hence, E_{15} is spanned by $\{[4,1,0], [2,0,1]\}$. Although the form of E_{15} is already simple, we can obtain an alternative form by using the Simplified Span Method. Row reducing the matrix

$$A = \begin{bmatrix} 4 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}, \quad \text{we obtain} \quad \mathbf{C} = \begin{bmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & -2 \end{bmatrix}.$$

Hence, an alternative form for the vectors in E_{15} is $\left\{ a\left[1, 0, \frac{1}{2}\right] + b[0,1,-2] \,\middle|\, a, b \in \mathbb{R} \right\} = \left\{ [a, b, \frac{1}{2}a - 2b] \,\middle|\, a, b \in \mathbb{R} \right\}$, just as we claimed in Example 11 in Section 4.2. ■

The method used in Examples 9 and 10 works in vector spaces other than $\mathbb{R}^n$, as we see in the next example. This fact will follow from the discussion of isomorphism in Section 5.5. (However, we will not use this fact in proofs of theorems until after Section 5.5.)

Example 11

Let S be the subset $\{5x^3 + 2x^2 + 4x - 3, -x^2 + 3x - 7, 2x^3 + 4x^2 - 8x + 5,\ x^3 + 2x + 5\}$ of $\mathcal{P}_3$. We use the Simplified Span Method to find a simplified form for the vectors in span(S).

Consider the coefficients of each polynomial as the coordinates of a vector in $\mathbb{R}^4$, yielding the corresponding set of vectors $T = \{[5,2,4,-3],[0,-1,3,-7],[2,4,-8,5],[1,0,2,5]\}$. Using the Simplified Span Method, we create the following matrix, whose rows are the vectors in T.

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

Then span(T) is the row space of the reduced row echelon form of $\mathbf{A}$, which is

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & -3 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Taking each nonzero row of $\mathbf{C}$ as the coefficients of a polynomial in $\mathcal{P}_3$, we see that

$$\begin{aligned} \text{span}(S) &= \{a(x^3+2x) + b(x^2-3x) + c(1) \,|\, a,b,c \in \mathbb{R}\} \\ &= \{ax^3 + bx^2 + (2a-3b)x + c \,|\, a,b,c \in \mathbb{R}\}. \end{aligned}$$

A Spanning Set for an Eigenspace

In Section 3.4, we illustrated a method for diagonalizing an $n \times n$ matrix, when possible. In fact, a set S of fundamental eigenvectors generated for a given eigenvalue λ spans the eigenspace E_λ (see Exercise 27). We illustrate this in the following example:

Example 12

Let

$$\mathbf{A} = \begin{bmatrix} 0 & -6 & 3 \\ 2 & -13 & 6 \\ 4 & -24 & 11 \end{bmatrix}.$$

A little work yields $p_{\mathbf{A}}(x) = x^3 + 2x^2 + x = x(x+1)^2$. We solve the homogeneous system $(-1\mathbf{I}_3 - \mathbf{A})\mathbf{X} = \mathbf{0}$ to find the eigenspace E_{-1} for $\mathbf{A}$.

Row reducing $[(-\mathbf{I}_3 - \mathbf{A})|\mathbf{0}]$ produces

$$\left[\begin{array}{ccc|c} 1 & -6 & 3 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right],$$

giving the solution set

$$E_{-1} = \{[6b-3c, b, c] \,|\, b,c \in \mathbb{R}\} = \{b[6,1,0] + c[-3,0,1] \,|\, b,c \in \mathbb{R}\}.$$

Thus, $E_{-1} = \text{span}(S)$, where $S = \{[6,1,0],[-3,0,1]\}$. The set S is precisely the set of fundamental eigenvectors that we would obtain for $\lambda = -1$ (verify!).

Special Case: The Span of the Empty Set

Until now, our results involving span have specified that the subset S of the vector space $\mathcal{V}$ be nonempty. However, our understanding of span(S) as the smallest subspace of $\mathcal{V}$ containing S allows us to give a meaningful definition for the span of the empty set.

Definition $\text{Span}(\{\ \}) = \{\mathbf{0}\}$.

This definition makes sense because the trivial subspace is the smallest subspace of $\mathcal{V}$, hence the smallest one containing the empty set. Thus, Theorem 4.5 is also true when the set S is empty. Similarly, to maintain consistency, we *define* any linear combination of the empty set of vectors to be $\mathbf{0}$. This ensures that the span of the empty set equals the set of all linear combinations of vectors taken from this set.

New Vocabulary

finite linear combination (of vectors in a vector space)
Simplified Span Method
span (of a set of vectors)
spanned by (as in "$\mathcal{V}$ is spanned by S")
span of the empty set

Highlights

- The span of a set is the collection of all finite linear combinations of vectors from the set.
- A set S spans a vector space $\mathcal{V}$ (i.e., $\mathcal{V}$ is spanned by S) if every vector in $\mathcal{V}$ is a (finite) linear combination of vectors in S.
- The row space of a matrix is the span of the rows of the matrix.
- $\mathbb{R}^3$ is spanned by $\{\mathbf{i},\mathbf{j},\mathbf{k}\}$; $\mathbb{R}^n$ is spanned by $\{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$; $\mathcal{P}_n$ is spanned by $\{1,x,x^2,\ldots,x^n\}$; $\mathcal{M}_{mn}$ is spanned by $\{\Psi_{ij}\}$, where each Ψ_{ij} has a 1 in the (i,j) entry, and zeroes elsewhere.
- The span of a set of vectors is always a subspace of the vector space, and is, in fact, the smallest subspace containing that set.
- If $S_1 \subseteq S_2$, then $\text{span}(S_1) \subseteq \text{span}(S_2)$.
- The Simplified Span Method generally produces a more simplified form of the span of a set of vectors by calculating the reduced row echelon form of the matrix whose *rows* are the given vectors.
- The span of the empty set is $\{\mathbf{0}\}$.

EXERCISES FOR SECTION 4.3

1. In each of the following cases, use the Simplified Span Method to find a simplified general form for all the vectors in span(S), where S is the given subset of $\mathbb{R}^n$:

   ★(a) $S = \{[1,1,0],[2,-3,-5]\}$

   (b) $S = \{[3,1,-2],[-3,-1,2],[6,2,-4]\}$

   ★(c) $S = \{[1,-1,1],[2,-3,3],[0,1,-1]\}$

   (d) $S = \{[1,1,1],[2,1,1],[1,1,2]\}$

   ★(e) $S = \{[1,3,0,1],[0,0,1,1],[0,1,0,1],[1,5,1,4]\}$

   (f) $S = \{[2,-1,3,1],[1,-2,0,-1],[3,-3,3,0],[5,-4,6,1],[1,-5,-3,-4]\}$

2. In each case, use the Simplified Span Method to find a simplified general form for all the vectors in span(S), where S is the given subset of $\mathcal{P}_3$:

   ★(a) $S = \{x^3 - 1, x^2 - x, x - 1\}$

   (b) $S = \{x^3 + 2x^2, 1 - 4x^2, 12 - 5x^3, x^3 - x^2\}$

   ★(c) $S = \{x^3 - x + 5, 3x^3 - 3x + 10, 5x^3 - 5x - 6, 6x - 6x^3 - 13\}$

3. In each case, use the Simplified Span Method to find a simplified general form for all the vectors in span(S), where S is the given subset of $\mathcal{M}_{22}$. (Hint: Rewrite each matrix as a 4-vector.)

   ★(a) $S = \left\{\begin{bmatrix}-1 & 1\\ 0 & 0\end{bmatrix}, \begin{bmatrix}0 & 0\\ 1 & -1\end{bmatrix}, \begin{bmatrix}-1 & 0\\ 0 & 1\end{bmatrix}\right\}$

   (b) $S = \left\{\begin{bmatrix}1 & 3\\ -2 & 1\end{bmatrix}, \begin{bmatrix}-2 & -5\\ 3 & 1\end{bmatrix}, \begin{bmatrix}1 & 4\\ -3 & 4\end{bmatrix}\right\}$

   ★(c) $S = \left\{\begin{bmatrix}1 & -1\\ 3 & 0\end{bmatrix}, \begin{bmatrix}2 & -1\\ 8 & -1\end{bmatrix}, \begin{bmatrix}-1 & 4\\ 4 & -1\end{bmatrix}, \begin{bmatrix}3 & -4\\ 5 & 6\end{bmatrix}\right\}$

★4. (a) Express the subspace $\mathcal{W}$ of $\mathbb{R}^4$ of all 4 -vectors of the form $[a+b, a+c, b+c, c]$ as the row space of a matrix $\mathbf{A}$.

   (b) Find the reduced row echelon form matrix $\mathbf{B}$ for $\mathbf{A}$.

   (c) Use the matrix $\mathbf{B}$ from part (b) to find a simplified form for the vectors in $\mathcal{W}$.

5. (a) Express the subspace $\mathcal{W}$ of $\mathbb{R}^5$ of all 5-vectors of the form $[2a+3b-4c, a+b-c, -b+7c, 3a+4b, 4a+2b]$ as the row space of a matrix $\mathbf{A}$.

   (b) Find the reduced row echelon form matrix $\mathbf{B}$ for $\mathbf{A}$.

   (c) Use the matrix $\mathbf{B}$ from part (b) to find a simplified form for the vectors in $\mathcal{W}$.

6. Prove that the set $S = \{[1,3,-1],[2,7,-3],[4,8,-7]\}$ spans $\mathbb{R}^3$.

7. Prove that the set $S = \{[1,-2,2],[3,-4,-1],[1,-4,9],[0,2,-7]\}$ does not span $\mathbb{R}^3$.

8. Show that the set $\{x^2+x+1, x+1, 1\}$ spans $\mathcal{P}_2$.

9. Prove that the set $\{x^2+4x-3, 2x^2+x+5, 7x-11\}$ does not span $\mathcal{P}_2$.

10. **(a)** Let $S = \{[1,-2,-2],[3,-5,1],[-1,1,-5]\}$. Show that $[-4,5,-13] \in \text{span}(S)$ by expressing it as a linear combination of the vectors in S.

(b) Prove that the set S in part (a) does not span $\mathbb{R}^3$.

⋆11. Consider the subset $S = \{x^3-2x^2+x-3, 2x^3-3x^2+2x+5, 4x^2+x-3, 4x^3-7x^2+4x-1\}$ of $\mathcal{P}$. Show that $3x^3-8x^2+2x+16$ is in span(S) by expressing it as a linear combination of the elements of S.

12. Prove that the set S of all vectors in $\mathbb{R}^4$ that have zeroes in exactly two coordinates spans $\mathbb{R}^4$. (Hint: Find a subset of S that spans $\mathbb{R}^4$.)

13. Let $\mathbf{a}$ be any nonzero element of $\mathbb{R}$. Prove that $\text{span}(\{\mathbf{a}\}) = \mathbb{R}$.

14. ⋆**(a)** Suppose that S_1 is the set of symmetric 2×2 matrices and that S_2 is the set of skew-symmetric 2×2 matrices. Prove that $\text{span}(S_1 \cup S_2) = \mathcal{M}_{22}$.

(b) State and prove the corresponding statement for $n \times n$ matrices.

15. Consider the subset $S = \{1+x^2, x+x^3, 3-2x+3x^2-12x^3\}$ of $\mathcal{P}$, and let $\mathcal{W} = \left\{ax^3+bx^2+cx+b \,\middle|\, a,b,c \in \mathbb{R}\right\}$. Show that $\mathcal{W} = \text{span}(S)$.

16. Let $\mathbf{A} = \begin{bmatrix} -9 & -15 & 8 \\ -10 & -14 & 8 \\ -30 & -45 & 25 \end{bmatrix}$.

⋆**(a)** Find a set S of two fundamental eigenvectors for $\mathbf{A}$ corresponding to the eigenvalue $\lambda = 1$. Multiply by a scalar to eliminate any fractions in your answers.

(b) Verify that the set S from part (a) spans E_1.

17. Let $S_1 = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a nonempty subset of a vector space $\mathcal{V}$. Let $S_2 = \{-\mathbf{v}_1, -\mathbf{v}_2, \ldots, -\mathbf{v}_n\}$. Show that $\text{span}(S_1) = \text{span}(S_2)$.

18. Let $\mathbf{u}$ and $\mathbf{v}$ be two nonzero vectors in $\mathbb{R}^3$, and let $S = \{\mathbf{u}, \mathbf{v}\}$. Show that span($S$) is a line through the origin if $\mathbf{u} = a\mathbf{v}$ for some real number a, but otherwise span(S) is a plane through the origin.

19. Let $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$ be three vectors in $\mathbb{R}^3$ and let $\mathbf{A}$ be the matrix whose rows are $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$. Show that $S = \{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ spans $\mathbb{R}^3$ if and only if $|\mathbf{A}| \neq 0$. (Hint: To prove that $\text{span}(S) = \mathbb{R}^3$ implies $|\mathbf{A}| \neq 0$, suppose $\mathbf{x} \in \mathbb{R}^3$ such that $\mathbf{A}\mathbf{x} = \mathbf{0}$. First, show that $\mathbf{x}$ is orthogonal to $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$. Then, express $\mathbf{x}$ as a linear combination of $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$. Prove that $\mathbf{x} \cdot \mathbf{x} = 0$, and then use Theorem 2.5 and Corollary 3.6. To prove that $|\mathbf{A}| \neq 0$ implies $\text{span}(S) = \mathbb{R}^3$, show that $\mathbf{A}$ is row equivalent to $\mathbf{I}_3$ and apply Theorem 2.8.)

20. Let $S = \{\mathbf{p}_1, \ldots, \mathbf{p}_k\}$ be a finite subset of $\mathcal{P}$. Prove that there is some positive integer n such that $\text{span}(S) \subseteq \mathcal{P}_n$.

▶**21.** Prove Corollary 4.6.

22. **(a)** Prove that if S is a nonempty subset of a vector space $\mathcal{V}$, then S is a subspace of $\mathcal{V}$ if and only if $\text{span}(S) = S$.

(b) Use part (a) to show that every subspace $\mathcal{W}$ of a vector space $\mathcal{V}$ has a set of vectors that spans $\mathcal{W}$ — namely, the set $\mathcal{W}$ itself.

(c) Describe the span of the set of the skew-symmetric matrices in $\mathcal{M}_{33}$.

23. Let S_1 and S_2 be subsets of a vector space $\mathcal{V}$. Prove that $\text{span}(S_1) = \text{span}(S_2)$ if and only if $S_1 \subseteq \text{span}(S_2)$ and $S_2 \subseteq \text{span}(S_1)$.

24. Let S_1 and S_2 be two subsets of a vector space $\mathcal{V}$.

(a) Prove that $\text{span}(S_1 \cap S_2) \subseteq \text{span}(S_1) \cap \text{span}(S_2)$.

⋆**(b)** Give an example of distinct subsets S_1 and S_2 of $\mathbb{R}^3$ for which the inclusion in part (a) is actually an equality.

⋆**(c)** Give an example of subsets S_1 and S_2 of $\mathbb{R}^3$ for which the inclusion in part (a) is not an equality.

25. Let S_1 and S_2 be subsets of a vector space $\mathcal{V}$.

(a) Show that $\text{span}(S_1) \cup \text{span}(S_2) \subseteq \text{span}(S_1 \cup S_2)$.

(b) Prove that if $S_1 \subseteq S_2$, then the inclusion in part (a) is an equality.

⋆**(c)** Give an example of subsets S_1 and S_2 in $\mathcal{P}_5$ for which the inclusion in part (a) is not an equality.

26. Let S be a subset of a vector space $\mathcal{V}$, and let $\mathbf{v} \in \mathcal{V}$. Show that $\text{span}(S) = \text{span}(S \cup \{\mathbf{v}\})$ if and only if $\mathbf{v} \in \text{span}(S)$.

27. Let $\mathbf{A}$ be an $n \times n$ matrix and λ be an eigenvalue for $\mathbf{A}$. Suppose S is a set of fundamental eigenvectors for $\mathbf{A}$ corresponding to λ. Prove that S spans E_λ.

▶**28.** Finish the proof of Theorem 4.5 by providing the details necessary to show that $\text{span}(S)$ is closed under addition if S is an infinite subset of a vector space $\mathcal{V}$.

⋆**29.** True or False:

(a) Span(S) is only defined if S is a finite subset of a vector space.

(b) If S is a subset of a vector space $\mathcal{V}$, then $\text{span}(S)$ contains every finite linear combination of vectors in S.

(c) If S is a subset of a vector space $\mathcal{V}$, then $\text{span}(S)$ is the smallest set in $\mathcal{V}$ containing S.

(d) If S is a subset of a vector space $\mathcal{V}$, and $\mathcal{W}$ is a subspace of $\mathcal{V}$ containing S, then we must have $\mathcal{W} \subseteq \text{span}(S)$.

(e) The row space of a 4×5 matrix $\mathbf{A}$ is a subspace of $\mathbb{R}^4$.

(f) A simplified form for the span of a finite set S of vectors in $\mathbb{R}^n$ can be found by row reducing the matrix whose rows are the vectors of S.

(g) The eigenspace E_λ for an eigenvalue λ of an $n \times n$ matrix $\mathbf{A}$ is the row space of $\lambda \mathbf{I}_n - \mathbf{A}$.

4.4 LINEAR INDEPENDENCE

In this section, we explore the concept of a linearly independent set of vectors and examine methods for determining whether or not a given set of vectors is linearly independent. We will also see that there are important connections between the concepts of span and linear independence.

Linear Independence and Dependence

At first, we will define linear independence and linear dependence only for finite sets of vectors. We will extend the definition to infinite sets at the end of this section.

Definition Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a finite nonempty subset of a vector space $\mathcal{V}$. Then S is **linearly dependent** if and only if there exist real numbers $a_1, \ldots, a_n$, not all zero, such that $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$. That is, S is linearly dependent if and only if the zero vector can be expressed as a nontrivial linear combination of the vectors in S.

S is **linearly independent** if and only if it is *not* linearly dependent.

The empty set, $\{\ \}$, is linearly independent.

To understand this definition, we begin first with the simplest cases: sets having one or two elements.

Suppose $S = \{\mathbf{v}\}$, a one-element set. Then, by part (4) of Theorem 4.1, $a\mathbf{v} = \mathbf{0}$ implies that either $a = 0$ or $\mathbf{v} = \mathbf{0}$. Now, for S to be linearly dependent, we would have to have some nonzero a satisfy $a\mathbf{v} = \mathbf{0}$. This would imply that $\mathbf{v} = \mathbf{0}$. We conclude that if $S = \{\mathbf{v}\}$, a one-element set, then S is linearly dependent if and only if $\mathbf{v} = \mathbf{0}$. Equivalently, $S = \{\mathbf{v}\}$ is linearly independent if and only if $\mathbf{v} \neq \mathbf{0}$.

Example 1

Let $S_1 = \{[3, -1, 4]\}$. Since S_1 contains a single vector and this vector is nonzero, S_1 is a linearly independent subset of $\mathbb{R}^3$. On the other hand, $S_2 = \{[0, 0, 0, 0]\}$ is a linearly dependent subset of $\mathbb{R}^4$.

Next, suppose $S = \{\mathbf{v}_1, \mathbf{v}_2\}$ is a linearly dependent set with two elements. Then there exist real numbers a_1 and a_2, not both zero, such that $a_1\mathbf{v}_1 + a_2\mathbf{v}_2 = \mathbf{0}$. If $a_1 \neq 0$, this implies that $\mathbf{v}_1 = -\frac{a_2}{a_1}\mathbf{v}_2$. That is, $\mathbf{v}_1$ is a scalar multiple of $\mathbf{v}_2$. Similarly, if $a_2 \neq 0$, we see that $\mathbf{v}_2$ is a scalar multiple of $\mathbf{v}_1$. Thus, linearly dependent sets containing exactly two vectors are precisely those for which at least one of the vectors is a scalar multiple of the other. So, a set of exactly two vectors is linearly independent precisely when neither of the vectors is a scalar multiple of the other. That is, two linearly independent vectors are not parallel. They represent two different directions.

Example 2

The set of vectors $S_1 = \{[1,-1,2],[-3,3,-6]\}$ in $\mathbb{R}^3$ is linearly dependent since one of the vectors is a scalar multiple (and hence a linear combination) of the other. For example, $[1,-1,2] = (-\frac{1}{3})[-3,3,-6]$.

Also, the set $S_2 = \{[3,-8],[2,5]\}$ is a linearly independent subset of $\mathbb{R}^2$ because neither of these vectors is a scalar multiple of the other. These two vectors are not parallel. They represent two different directions.

In general, because linear independence is defined as the negation of linear dependence, we can express linear independence as follows:

Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a finite nonempty subset of a vector space $\mathcal{V}$. Then S is **linearly independent** if and only if for any set of real numbers $a_1, \ldots, a_n$, the equation $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$ implies $a_1 = a_2 = \cdots = a_n = 0$.

Example 3

The set of vectors $\{\mathbf{i},\mathbf{j},\mathbf{k}\}$ in $\mathbb{R}^3$ is linearly independent because $a\mathbf{i} + b\mathbf{j} + c\mathbf{k} = [a,b,c] = [0,0,0]$ if and only if $a = b = c = 0$. More generally, the set $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ in $\mathbb{R}^n$ is linearly independent.

Example 4

Let S be any subset of a vector space $\mathcal{V}$ containing the zero vector $\mathbf{0}$. If S contains no vector other than $\mathbf{0}$, then we have already seen that S is linearly dependent. If $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ contains at least two distinct vectors with one of them $\mathbf{0}$ (say $\mathbf{v}_1 = \mathbf{0}$), then $\mathbf{0}$ can be expressed as a nontrivial linear combination of the vectors in S since $1\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_n = 1 \cdot \mathbf{0} + \mathbf{0} + \cdots + \mathbf{0} = \mathbf{0}$. Hence, by the definition, S is linearly dependent. Therefore, in all cases, any finite subset of a vector space that contains the zero vector $\mathbf{0}$ is linearly dependent.

The result we obtained in Example 4 is important enough to highlight:

Any finite subset of a vector space that contains the zero vector $\mathbf{0}$ is linearly dependent.

Example 5

Let $S = \{[2,5],[3,-2],[4,-9]\}$. Notice that $[4,-9] = -[2,5] + 2[3,-2]$. This shows that some vector in S can be expressed as a linear combination of other vectors in S. In other words, the vector $[4,-9]$ is a "destination" that can be reached using a linear combination of the other vectors in S. It does not strike out in a new, independent, direction. Notice that we can subtract $[4,-9]$ from both sides of the equation $[4,-9] = -[2,5] + 2[3,-2]$ to obtain

$$\mathbf{0} = -[2,5] + 2[3,-2] - [4,-9].$$

We have thus expressed the zero vector as a nontrivial linear combination of the vectors in S, and this implies that S is linearly dependent.

Example 6

Consider the subset $S = \{[1,-1,0,2],[0,-2,1,0],[2,0,-1,1]\}$ of $\mathbb{R}^4$. We will investigate whether S is linearly independent.

We proceed by assuming that $a[1,-1,0,2] + b[0,-2,1,0] + c[2,0,-1,1] = [0,0,0,0]$ and solve for $a, b,$ and c to see whether all these coefficients must be zero. That is, we determine whether the following homogeneous system has only the trivial solution:

$$\left\{\begin{array}{rrrl} a & & +\ 2c & = 0 \\ -a & -\ 2b & & = 0 \\ & b & -\ \ c & = 0 \\ 2a & & +\ \ c & = 0 \end{array}\right. .$$

Row reducing

$$\begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \left[\begin{array}{rrr|r} 1 & 0 & 2 & 0 \\ -1 & -2 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 2 & 0 & 1 & 0 \end{array}\right] \end{array}, \quad \text{we obtain} \quad \begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \end{array},$$

which shows that this system has only the trivial solution $a = b = c = 0$. Hence, S is linearly independent.

Using Row Reduction to Test for Linear Independence

Notice that in Example 6, the columns of the matrix to the left of the augmentation bar are just the vectors in S. In general, to test a finite set of vectors in $\mathbb{R}^n$ for linear independence, we row reduce the matrix whose *columns* are the vectors in the set, and then check whether the associated homogeneous system has only the trivial solution. In practice it is not necessary to include the augmentation bar and the column of zeroes to its right, since this column never changes in the row reduction process. Thus, we have

Method to Test for Linear Independence Using Row Reduction (Independence Test Method)
Let S be a finite nonempty set of vectors in $\mathbb{R}^n$. To determine whether S is linearly independent, perform the following steps:

Step 1: Create the matrix $\mathbf{A}$ whose *columns* are the vectors in S.

Step 2: Find $\mathbf{B}$, the reduced row echelon form of $\mathbf{A}$.

Step 3: If there is a pivot in every column of $\mathbf{B}$, then S is linearly independent. Otherwise, S is linearly dependent.

Example 7

Consider the subset $S = \{[3,1,-1],[-5,-2,2],[2,2,-1]\}$ of $\mathbb{R}^3$. Using the Independence Test Method, we row reduce

$$\begin{bmatrix} 3 & -5 & 2 \\ 1 & -2 & 2 \\ -1 & 2 & -1 \end{bmatrix} \text{ to obtain } \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Since we found a pivot in every column, the set S is linearly independent. ■

Example 8

Consider the subset $S = \{[2,5],[3,7],[4,-9],[-8,3]\}$ of $\mathbb{R}^2$. Using the Independence Test Method, we row reduce

$$\begin{bmatrix} 2 & 3 & 4 & -8 \\ 5 & 7 & -9 & 3 \end{bmatrix} \text{ to obtain } \begin{bmatrix} 1 & 0 & -55 & 65 \\ 0 & 1 & 38 & -46 \end{bmatrix}.$$

Since we have no pivots in columns 3 and 4, the set S is linearly dependent. ■

In the last example, there are more columns than rows in the matrix we row reduced. Hence, there must ultimately be some column without a pivot, since each pivot is in a different row. In such cases, the original set of vectors must be linearly dependent. This motivates the following result, which we ask you to formally prove as Exercise 16:

Theorem 4.7 If S is any set in $\mathbb{R}^n$ containing k distinct vectors, where $k > n$, then S is linearly dependent.

The Independence Test Method can be adapted for use on vector spaces other than $\mathbb{R}^n$, as in the next example. We will prove that the Independence Test Method is actually valid in such cases in Section 5.5.

Example 9

Consider the following subset of $\mathcal{M}_{22}$:

$$S = \left\{ \begin{bmatrix} 2 & 3 \\ -1 & 4 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 6 & -1 \\ 3 & 2 \end{bmatrix}, \begin{bmatrix} -11 & 3 \\ -2 & 2 \end{bmatrix} \right\}.$$

We determine whether S is linearly independent using the Independence Test Method. First, we represent the 2×2 matrices in S as 4-vectors. Placing them in a matrix, using each 4-vector as a column, we get

$$\begin{bmatrix} 2 & -1 & 6 & -11 \\ 3 & 0 & -1 & 3 \\ -1 & 1 & 3 & -2 \\ 4 & 1 & 2 & 2 \end{bmatrix}, \text{ which reduces to } \begin{bmatrix} 1 & 0 & 0 & \frac{1}{2} \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -\frac{3}{2} \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

There is no pivot in column 4. Hence, S is linearly dependent.

Alternate Characterizations of Linear Independence

We have already seen that a set of two vectors is linearly dependent if one vector is a linear combination of the other. We now generalize this to larger sets as well. Notice in the last two examples that the final columns of the row reduced matrix indicate how to obtain the original vectors in the nonpivot columns from earlier columns. In Example 8, the third column of the row reduced matrix is $[-55, 38]$. The entries -55 and 38 represent the coefficients for a linear combination of the original first and second columns that produces the original third column; that is, $[4, -9] = -55[2, 5] + 38[3, 7]$. Similarly, the fourth column $[65, -46]$ of the row reduced matrix implies $[-8, 3] = 65[2, 5] - 46[3, 7]$. In Example 9, the entries of the fourth column of the row reduced matrix are $\frac{1}{2}, 3, -\frac{3}{2}, 0$, respectively. The first three of these are the coefficients for a linear combination of the first three matrices in S that produces the fourth matrix; that is,

$$\begin{bmatrix} -11 & 3 \\ -2 & 2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 2 & 3 \\ -1 & 4 \end{bmatrix} + 3\begin{bmatrix} -1 & 0 \\ 1 & 1 \end{bmatrix} - \frac{3}{2}\begin{bmatrix} 6 & -1 \\ 3 & 2 \end{bmatrix}.$$

We see that when vectors are linearly dependent, the Independence Test Method gives a natural way of expressing certain vectors as linear combinations of the others. More generally, we have

Theorem 4.8 Suppose S is a finite set of vectors having at least two elements. Then S is linearly dependent if and only if some vector in S can be expressed as a linear combination of the other vectors in S.

Proof. We start by assuming that S is linearly dependent. Therefore, we have coefficients $a_1,\ldots,a_n$ such that $\mathbf{0} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$, with $a_i \neq 0$ for some i. Then,

$$\mathbf{v}_i = \left(-\frac{a_1}{a_i}\right)\mathbf{v}_1 + \cdots + \left(-\frac{a_{i-1}}{a_i}\right)\mathbf{v}_{i-1} + \left(-\frac{a_{i+1}}{a_i}\right)\mathbf{v}_{i+1} + \cdots + \left(-\frac{a_n}{a_i}\right)\mathbf{v}_n,$$

which expresses $\mathbf{v}_i$ as a linear combination of the other vectors in S.

For the second half of the proof, we assume that there is a vector $\mathbf{v}_i$ in S that is a linear combination of the other vectors in S. Without loss of generality, assume $\mathbf{v}_i = \mathbf{v}_1$; that is, $i = 1$. Therefore, there are real numbers $a_2,\ldots,a_n$ such that

$$\mathbf{v}_1 = a_2\mathbf{v}_2 + a_3\mathbf{v}_3 + \cdots + a_n\mathbf{v}_n.$$

Letting $a_1 = -1$, we get $\mathbf{0} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$. Since $a_1 \neq 0$, this shows that S is linearly dependent, completing the proof of the theorem. □

Example 10

The set of vectors $S = \{[1,2,-1],[0,1,2],[2,7,4]\}$ in $\mathbb{R}^3$ is linearly dependent because it is possible to express some vector in the set S as a linear combination of the others. For example, $[2,7,4] = 2[1,2,-1] + 3[0,1,2]$. From a geometric point of view, the fact that $[2,7,4]$ can be expressed as a linear combination of the vectors $[1,2,-1]$ and $[0,1,2]$ means that $[2,7,4]$ lies in the plane spanned by $[1,2,-1]$ and $[0,1,2]$, assuming that all three vectors have their initial points at the origin (see Figure 4.5).

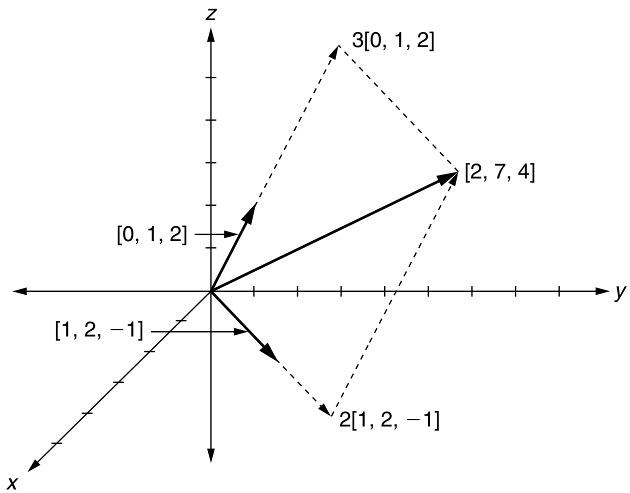


FIGURE 4.5

The vector $[2,7,4]$ in the plane spanned by $[1,2,-1]$ and $[0,1,2]$

Example 11

Consider the subset $S = \{[1,2,-1,1],[2,1,0,1],[2,-2,1,0],[11,1,1,4]\}$ of $\mathbb{R}^4$. Using the Independence Test Method, we row reduce

$$\begin{bmatrix} 1 & 2 & 2 & 11 \\ 2 & 1 & -2 & 1 \\ -1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 4 \end{bmatrix} \text{ to obtain } \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Because there is no pivot in column 4, S is linearly dependent. This means that at least one vector in S is a linear combination of the others. In particular, the first three entries of the fourth column of the row reduced matrix represent coefficients that express $[11,1,1,4]$ as a linear combination of the other vectors:

$$[11,1,1,4] = 1\cdot[1,2,-1,1] + 3\cdot[2,1,0,1] + 2\cdot[2,-2,1,0].$$

The characterization of linear dependence and linear independence in Theorem 4.8 can be expressed in alternate notation using the concept of span.

If $\mathbf{v}$ is a vector in a set S, we use the notation $S - \{\mathbf{v}\}$ to represent the set of all (other) vectors in S except $\mathbf{v}$. Of course, in the special case where $S = \{\mathbf{v}\}$ itself, the set $S - \{\mathbf{v}\} = \{\ \}$, the empty set. Theorem 4.8 implies that a subset S of two or more vectors in a vector space $\mathcal{V}$ is linearly independent precisely when no vector $\mathbf{v}$ in S is in the span of the remaining vectors. That is,

> A set S in a vector space $\mathcal{V}$ is linearly independent if and only if there is no vector $\mathbf{v} \in S$ such that $\mathbf{v} \in \text{span}(S - \{\mathbf{v}\})$.

This statement holds even in the special cases when $S = \{\mathbf{v}\}$ or $S = \{\ \}$. You are asked to prove this in Exercise 21.

Equivalently, we have

> A set S in a vector space $\mathcal{V}$ is linearly dependent if and only if there is some vector $\mathbf{v} \in S$ such that $\mathbf{v} \in \text{span}(S - \{\mathbf{v}\})$.

Another useful characterization of linear independence is the following:

> A nonempty set of vectors $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly independent if and only if
>
> (1) $\mathbf{v}_1 \neq \mathbf{0}$, and
>
> (2) for each k, $2 \le k \le n$, $\mathbf{v}_k \notin \text{span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}\})$.

This states that $\mathcal{S}$ is linearly independent if each vector in $\mathcal{S}$ can not be expressed as a linear combination of those vectors listed before it. You are asked to prove this in Exercise 22.

Uniqueness of Expression of a Vector as a Linear Combination

The next theorem serves as the foundation for the rest of this chapter because it gives an even more powerful connection between the concepts of span and linear independence.

> **Theorem 4.9** Let $\mathcal{S}$ be a nonempty finite subset of a vector space $\mathcal{V}$. Then $\mathcal{S}$ is linearly independent if and only if every vector $\mathbf{v} \in \text{span}(\mathcal{S})$ can be expressed *uniquely* as a linear combination of the elements of $\mathcal{S}$.

Proof. Let $\mathcal{S} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$.

Suppose first that $\mathcal{S}$ is linearly independent. Assume that $\mathbf{v} \in \text{span}(\mathcal{S})$ can be expressed both as $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$ and as $\mathbf{v} = b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n$. In order to show that the linear combination for $\mathbf{v}$ is unique, we need to prove that $a_i = b_i$ for all i. But $\mathbf{0} = \mathbf{v} - \mathbf{v} = (a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n) - (b_1\mathbf{v}_1 + \cdots + b_n\mathbf{v}_n) = (a_1 - b_1)\mathbf{v}_1 + \cdots + (a_n - b_n)\mathbf{v}_n$. Since $\mathcal{S}$ is a linearly independent set, each $a_i - b_i = 0$, by the definition of linear independence, and thus $a_i = b_i$ for all i.

Conversely, assume every vector in span($\mathcal{S}$) can be uniquely expressed as a linear combination of elements of $\mathcal{S}$. Since $\mathbf{0} \in \text{span}(\mathcal{S})$, there is exactly one linear combination $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$ of elements of $\mathcal{S}$ that equals $\mathbf{0}$. But the fact that $\mathbf{0} = 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_n$ together with the uniqueness of expression for $\mathbf{0}$ means $a_1, \ldots, a_n$ are all zero. Thus, by the definition of linear independence, $\mathcal{S}$ is linearly independent. $\square$

By Theorem 4.9, $\mathcal{S}$ is linearly independent if there is precisely one way of reaching any "destination" in span($\mathcal{S}$) using the given "directions" in $\mathcal{S}$!

Example 12

Recall the subset $\mathcal{S} = \{[1, -1, 0, 2], [0, -2, 1, 0], [2, 0, -1, 1]\}$ of $\mathbb{R}^4$ from Example 6. In that example, we proved that $\mathcal{S}$ is linearly independent. Now

$$[11, 1, -6, 10] = 3[1, -1, 0, 2] + (-2)[0, -2, 1, 0] + 4[2, 0, -1, 1]$$

so $[11, 1, -6, 10]$ is in span($\mathcal{S}$). Then by Theorem 4.9, this is the *only* possible way to express $[11, 1, -6, 10]$ as a linear combination of the elements in $\mathcal{S}$.

Recall the subset $\mathcal{S} = \{[2, 5], [3, 7], [4, -9], [-8, 3]\}$ of $\mathbb{R}^2$ from Example 8. In that example, we proved that $\mathcal{S}$ is linearly dependent. Just before Theorem 4.8 we showed that $[4, -9] = -55[2, 5] + 38[3, 7]$. This means that $[4, 9] = -55[2, 5] + 38[3, 7] + 0[4, -9] + 0[-8, 3]$, but we can also express this vector as $[4, 9] = 0[2, 5] + 0[3, 7] + 1[4, -9] + 0[-8, 3]$. Since $[4, 9]$ is obviously in span($\mathcal{S}$), we have found a vector in span($\mathcal{S}$) for which the linear combination of elements in $\mathcal{S}$ is not unique, just as Theorem 4.9 asserts. ■

Linear Independence of Eigenvectors

We will prove in Section 5.6 that any set of fundamental eigenvectors for an $n \times n$ matrix produced by the Diagonalization Method is always linearly independent (also see Exercise 25). Let us assume this for the moment. Now, if the method produces n eigenvectors, then the matrix $\mathbf{P}$ whose columns are these eigenvectors must row reduce to $\mathbf{I}_n$, by the Independence Test Method. This will establish the claim in Section 3.4 that $\mathbf{P}$ is nonsingular.

Example 13

Consider the 3×3 matrix

$$\mathbf{A} = \begin{bmatrix} -2 & 12 & -4 \\ -2 & 8 & -2 \\ -3 & 9 & -1 \end{bmatrix}.$$

You are asked to show in Exercise 14 that $[4,2,3]$ is a fundamental eigenvector for the eigenvalue $\lambda_1 = 1$, and that $[3,1,0]$ and $[-1,0,1]$ are fundamental eigenvectors for the eigenvalue $\lambda_2 = 2$. We test their linear independence by row reducing

$$\mathbf{P} = \begin{bmatrix} 4 & 3 & -1 \\ 2 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \quad \text{to obtain} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

thus illustrating that this set of fundamental eigenvectors is indeed linearly independent *and* that $\mathbf{P}$ is nonsingular. ■

Linear Independence of Infinite Sets

Most cases in which we check for linear independence involve a *finite* set S. However, we will occasionally want to discuss linear independence for infinite sets of vectors.

Definition An infinite subset S of a vector space $\mathcal{V}$ is **linearly dependent** if and only if there is some finite subset T of S such that T is linearly dependent. S is **linearly independent** if and only if S is *not* linearly dependent.

Example 14

Consider the subset S of $\mathcal{M}_{22}$ consisting of all nonsingular 2×2 matrices. We will show that S is linearly dependent.

Let $T = \{\mathbf{I}_2, \ 2\mathbf{I}_2\}$, a subset of S. Clearly, since the second element of T is a scalar multiple of the first element of T, T is a linearly dependent set. Hence, S is linearly dependent, since one of its finite subsets is linearly dependent. ■

We can also express the definition of linear independence using the negation of the definition of linear dependence:

An infinite subset $\mathcal{S}$ of a vector space $\mathcal{V}$ is linearly independent if and only if every finite subset T of $\mathcal{S}$ is linearly independent.

From this, Theorem 4.8 implies that an infinite subset $\mathcal{S}$ of a vector space $\mathcal{V}$ is linearly independent if and only if no vector in $\mathcal{S}$ is a finite linear combination of other vectors in $\mathcal{S}$.

These characterizations of linear independence are obviously valid as well when $\mathcal{S}$ is a finite set.

Example 15

Let $\mathcal{S} = \{1, 1+x, 1+x+x^2, 1+x+x^2+x^3, \ldots\}$, an infinite subset of $\mathcal{P}$. We will show that $\mathcal{S}$ is linearly independent.

Suppose $T = \{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$ is a finite subset of $\mathcal{S}$, with the polynomials written in order of increasing degree. Also suppose that

$$a_1\mathbf{p}_1 + \cdots + a_n\mathbf{p}_n = \mathbf{0}.$$

We need to show that $a_1 = a_2 = \cdots = a_n = 0$. We prove this by contradiction.

Suppose at least one a_i is nonzero. Let a_k be the last nonzero coefficient in the series. Then,

$$a_1\mathbf{p}_1 + \cdots + a_k\mathbf{p}_k = \mathbf{0}, \text{ with } a_k \neq 0.$$

Hence,

$$\mathbf{p}_k = -\frac{a_1}{a_k}\mathbf{p}_1 - \frac{a_2}{a_k}\mathbf{p}_2 - \cdots - \frac{a_{k-1}}{a_k}\mathbf{p}_{k-1}.$$

Because all the degrees of the polynomials in T are different and they were listed in order of increasing degree, this equation expresses $\mathbf{p}_k$ as a linear combination of polynomials whose degrees are lower than that of $\mathbf{p}_k$. This can not happen, and so we get our desired contradiction. ■

The next theorem generalizes Theorem 4.9 to include both finite and infinite sets. You are asked to prove this in Exercise 27.

Theorem 4.10 Let $\mathcal{S}$ be a nonempty subset of a vector space $\mathcal{V}$. Then $\mathcal{S}$ is linearly independent if and only if every vector $\mathbf{v} \in \text{span}(\mathcal{S})$ can be expressed *uniquely* as a finite linear combination of the elements of $\mathcal{S}$, if terms with zero coefficients are ignored.

Remember: a *finite* linear combination from an infinite set $\mathcal{S}$ involves only a finite number of vectors from $\mathcal{S}$. The phrase "if terms with zero coefficients are ignored" means that two finite linear combinations from a set S are considered the same when all their terms with nonzero coefficients agree. Adding more terms with zero coefficients to a linear combination is not considered to produce a different linear combination.

Example 16

Recall the set $\mathcal{S}$ of nonsingular 2×2 matrices discussed in Example 14. Because $\mathcal{S}$ is linearly dependent, some vector in span($\mathcal{S}$) can be expressed in more than one way as a linear combination of vectors in $\mathcal{S}$. For example,

$$\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} = 2\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + 2\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = 1\begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix} + (-1)\begin{bmatrix} 1 & -1 \\ -2 & 1 \end{bmatrix}.$$

Summary of Results

This section includes several different, but equivalent, descriptions of linearly independent and linearly dependent sets of vectors. Several additional characterizations are described in the exercises. The most important results from both the section and the exercises are summarized in Table 4.1.

New Vocabulary

Independence Test Method
linearly dependent (set of vectors)
linearly independent (set of vectors)
redundant vector

Highlights

- A set of vectors is linearly dependent if there is a nontrivial linear combination of the vectors that equals $\mathbf{0}$.
- A set of vectors is linearly independent if the only linear combination of the vectors that equals $\mathbf{0}$ is the trivial linear combination (i.e., all coefficients $= 0$).
- A single element set $\{\mathbf{v}\}$ is linearly independent if and only if $\mathbf{v} \neq \mathbf{0}$.
- A two-element set $\{\mathbf{v}_1, \mathbf{v}_2\}$ is linearly independent if and only if neither vector is a scalar multiple of the other.
- The vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ are linearly independent in $\mathbb{R}^n$, and the vectors $\{1, x, x^2, \ldots, x^n\}$ are linearly independent in $\mathcal{P}_n$.
- Any set containing the zero vector is linearly dependent.
- The Independence Test Method determines whether a finite set is linearly independent by calculating the reduced row echelon form of the matrix whose *columns* are the given vectors.

Table 4.1 Equivalent conditions for a subset S of a vector space to be linearly independent or linearly dependent

Linear Independence of S	Linear Dependence of S	Source
If $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$, then $a_1 = a_2 = \cdots = a_n = 0$. (The zero vector requires zero coefficients.)	If $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, then $a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$ for some scalars $a_1, a_2, \ldots, a_n$, with some $a_i \neq 0$. (The zero vector does not require all coefficients to be zero.)	Definition
No vector in S is a finite linear combination of other vectors in S.	*Some* vector in S is a finite linear combination of other vectors in S.	Theorem 4.8 and Remarks after Example 14
For every $\mathbf{v} \in S$, we have $\mathbf{v} \notin \text{span}(S - \{\mathbf{v}\})$.	There is a $\mathbf{v} \in S$ such that $\mathbf{v} \in \text{span}(S - \{\mathbf{v}\})$.	Alternate characterization
For every $\mathbf{v} \in S$, span$(S - \{\mathbf{v}\})$ does not contain all the vectors of span(S).	There is some $\mathbf{v} \in S$ such that span$(S - \{\mathbf{v}\}) =$ span(S).	Exercise 12
If $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, then for each k, $\mathbf{v}_k \notin \text{span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}\})$. (Each $\mathbf{v}_k$ is not a linear combination of the previous vectors in S.)	If $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, some $\mathbf{v}_k$ can be expressed as $\mathbf{v}_k = a_1\mathbf{v}_1 + \cdots + a_{k-1}\mathbf{v}_{k-1}$. (Some $\mathbf{v}_k$ is a linear combination of the previous vectors in S.)	Exercise 22
Every vector in span(S) can be uniquely expressed as a linear combination of the vectors in S.	*Some* vector in span(S) can be expressed in more than one way as a linear combination of the vectors in S.	Theorem 4.9 and Theorem 4.10
Every finite subset of S is linearly independent.	*Some* finite subset of S is linearly dependent.	Definition when S is infinite

- If a subset of $\mathbb{R}^n$ contains more than n vectors, then the subset is linearly dependent.
- A set of vectors is linearly dependent if some vector can be expressed as a linear combination of the others (i.e., is in the span of the other vectors). (Such a vector is said to be redundant.)
- A set of vectors is linearly independent if no vector can be expressed as a linear combination of the others (i.e., is in the span of the other vectors).
- A set of vectors is linearly independent if no vector can be expressed as a linear combination of those listed before it in the set.
- A set of fundamental eigenvectors produced by the Diagonalization Method is linearly independent (this will be justified in Section 5.6).

- An infinite set of vectors is linearly dependent if some finite subset is linearly dependent.
- An infinite set of vectors is linearly independent if every finite subset is linearly independent.
- A set S of vectors is linearly independent if and only if every vector in span(S) is produced by a unique linear combination of the vectors in S.

EXERCISES FOR SECTION 4.4

★**1.** In each part, determine by quick inspection whether the given set of vectors is linearly independent. State a reason for your conclusion.

(a) $\{[0,1,1]\}$

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

(c) $\{[1,2,-5],[-2,-4,10]\}$

(d) $\{[4,2,1],[-1,3,7],[0,0,0]\}$

(e) $\{[2,-5,1],[1,1,-1],[0,2,-3],[2,2,6]\}$

2. Use the Independence Test Method to determine which of the following sets of vectors are linearly independent:

★**(a)** $\{[1,9,-2],[3,4,5],[-2,5,-7]\}$

★**(b)** $\{[2,-1,3],[4,-1,6],[-2,0,2]\}$

(c) $\{[-2,4,2][-1,5,2],[3,5,1]\}$

(d) $\{[5,-2,3],[-4,1,-7],[7,-4,-5]\}$

★**(e)** $\{[2,5,-1,6],[4,3,1,4],[1,-1,1,-1]\}$

(f) $\{[1,3,-2,4],[3,11,-2,-2],[2,8,3,-9],[3,11,-8,5]\}$

3. Use the Independence Test Method to determine which of the following subsets of $\mathcal{P}_2$ are linearly independent:

★**(a)** $\left\{x^2+x+1,\ x^2-1,\ x^2+1\right\}$

(b) $\{x^2-x+3, 2x^2-3x-1, 5x^2-9x-7\}$

★**(c)** $\{2x-6, 7x+2, 12x-7\}$

(d) $\left\{x^2+ax+b \,\middle|\, |a|=|b|=1\right\}$

4. Determine which of the following subsets of $\mathcal{P}$ are linearly independent:

★**(a)** $\{x^2-1, x^2+1, x^2+x\}$

(b) $\{1+x^2-x^3, 2x-1, x+x^3\}$

★**(c)** $\{4x^2+2, x^2+x-1, x, x^2-5x-3\}$

(d) $\{3x^3+2x+1, x^3+x, x-5, x^3+x-10\}$

★(e) $\{1, x, x^2, x^3, \ldots\}$

(f) $\{1, 1+2x, 1+2x+3x^2, 1+2x+3x^2+4x^3, \ldots\}$

5. Show that the following is a linearly dependent subset of $\mathcal{M}_{22}$:

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

6. Prove that the following is linearly independent in $\mathcal{M}_{32}$:

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

7. Let $S = \{[1,1,0], [-2,0,1]\}$.

(a) Show that S is a linearly independent subset of $\mathbb{R}^3$.

★(b) Find a vector $\mathbf{v}$ in $\mathbb{R}^3$ such that $S \cup \{\mathbf{v}\}$ is also linearly independent.

★(c) Is the vector $\mathbf{v}$ from part (b) unique, or could some other choice for $\mathbf{v}$ have been made? Why or why not?

★(d) Find a nonzero vector $\mathbf{u}$ in $\mathbb{R}^3$ such that $S \cup \{\mathbf{u}\}$ is linearly dependent.

8. Suppose that S is the subset $\{[2,-1,0,5], [1,-1,2,0], [-1,0,1,1]\}$ of $\mathbb{R}^4$.

(a) Show that S is linearly independent.

(b) Find a linear combination of vectors in S that produces $[-2,0,3,-4]$ (an element of span(S)).

(c) Is there a different linear combination of the elements of S that yields $[-2,0,3,-4]$? If so, find one. If not, why not?

9. Consider $S = \{2x^3 - x + 3, 3x^3 + 2x - 2, x^3 - 4x + 8, 4x^3 + 5x - 7\} \subseteq \mathcal{P}_3$.

(a) Show that S is linearly dependent.

(b) Show that every three-element subset of S is linearly dependent.

(c) Explain why every subset of S containing exactly two vectors is linearly independent. (Note: There are six possible two-element subsets.)

10. Let $\mathbf{u} = [u_1, u_2, u_3], \mathbf{v} = [v_1, v_2, v_3], \mathbf{w} = [w_1, w_2, w_3]$ be three vectors in $\mathbb{R}^3$. Show that $S = \{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$ is linearly independent if and only if

$$\begin{vmatrix}u_1 & u_2 & u_3\\v_1 & v_2 & v_3\\w_1 & w_2 & w_3\end{vmatrix} \neq 0.$$

(Hint: Consider the transpose and use the Independence Test Method.) (Compare this exercise with Exercise 19 in Section 4.3.)

11. For each of the following vector spaces, find a linearly independent subset S containing exactly four elements:

★**(a)** $\mathbb{R}^4$

(b) $\mathbb{R}^5$

★**(c)** $\mathcal{P}_3$

(d) $\mathcal{M}_{23}$

★**(e)** $\mathcal{V}$ = set of all symmetric matrices in $\mathcal{M}_{33}$.

12. Let S be a (possibly infinite) subset of a vector space $\mathcal{V}$. Prove that S is linearly dependent if and only if there is a vector $\mathbf{v} \in S$ such that span$(S - \{\mathbf{v}\}) =$ span(S). (We say that such a vector $\mathbf{v}$ is **redundant** in S because the same set of linear combinations is obtained after $\mathbf{v}$ is removed from S; that is, $\mathbf{v}$ is not needed.)

13. Find a redundant vector in each given linearly dependent set, and show that it satisfies the definition of a redundant vector given in Exercise 12.

(a) $\{[4,-2,6,1],[1,0,-1,2],[0,0,0,0],[6,-2,5,5]\}$

★**(b)** $\{[1,1,0,0],[1,1,1,0],[0,0,-6,0]\}$

(c) $\left\{[x_1,x_2,x_3,x_4] \in \mathbb{R}^4 \,\middle|\, x_i = \pm 1, \text{ for each } i\right\}$

14. Verify that the Diagonalization Method of Section 3.4 produces the fundamental eigenvectors given in the text for the matrix $\mathbf{A}$ of Example 13.

15. Let $S_1 = \{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ be a subset of a vector space $\mathcal{V}$, let c be a nonzero real number, and let $S_2 = \{c\mathbf{v}_1,\ldots,c\mathbf{v}_n\}$. Show that S_1 is linearly independent if and only if S_2 is linearly independent.

▶**16.** Prove Theorem 4.7. (Hint: Use the definition of linear dependence. Construct an appropriate homogeneous system of linear equations, and show that the system has a nontrivial solution.)

17. Let $\mathbf{f}$ be a polynomial with at least two nonzero terms having different degrees. Prove that the set $\{\mathbf{f}(x), x\mathbf{f}'(x)\}$ (where $\mathbf{f}'$ is the derivative of $\mathbf{f}$) is linearly independent in $\mathcal{P}$.

18. Let $\mathcal{V}$ be a vector space, $\mathcal{W}$ a subspace of $\mathcal{V}$, S a linearly independent subset of $\mathcal{W}$, and $\mathbf{v} \in \mathcal{V} - \mathcal{W}$. Prove that $S \cup \{\mathbf{v}\}$ is linearly independent.

19. Let $\mathbf{A}$ be an $n \times m$ matrix, let $S = \{\mathbf{v}_1,\ldots,\mathbf{v}_k\}$ be a finite subset of $\mathbb{R}^m$, and let $T = \{\mathbf{A}\mathbf{v}_1,\ldots,\mathbf{A}\mathbf{v}_k\}$, a subset of $\mathbb{R}^n$.

(a) Prove that if T is a linearly independent subset of $\mathbb{R}^n$ containing k distinct vectors, then S is a linearly independent subset of $\mathbb{R}^m$.

★**(b)** Find a matrix $\mathbf{A}$ for which the converse to part (a) is false.

(c) Show that the converse to part (a) is true if $\mathbf{A}$ is square and nonsingular.

20. Prove that every subset of a linearly independent set is linearly independent.

21. Let S be a subset of a vector space $\mathcal{V}$. If $S = \{\mathbf{a}\}$ or $S = \{\ \}$, prove that S is linearly independent if and only if there is no vector $\mathbf{v} \in S$ such that $\mathbf{v} \in \text{span}(S - \{\mathbf{v}\})$.

22. Suppose $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a finite subset of a vector space $\mathcal{V}$. Prove that S is linearly independent if and only if $\mathbf{v}_1 \neq \mathbf{0}$ and, for each k with $2 \leq k \leq n$, $\mathbf{v}_k \notin \text{span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}\})$. (Hint: Half of the proof is done by contrapositive. For this half, assume that S is linearly dependent, and use an argument similar to the first half of the proof of Theorem 4.8 to show some $\mathbf{v}_k$ is in $\text{span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}\})$. For the other half, assume S is linearly independent and show $\mathbf{v}_1 \neq \mathbf{0}$ and each $\mathbf{v}_k \notin \text{span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}\})$.)

23. Let $\mathbf{f}$ be an nth-degree polynomial in $\mathcal{P}$, and let $\mathbf{f}^{(i)}$ be the ith derivative of $\mathbf{f}$. Show that $\left\{\mathbf{f}, \mathbf{f}^{(1)}, \mathbf{f}^{(2)}, \ldots, \mathbf{f}^{(n)}\right\}$ is a linearly independent subset of $\mathcal{P}$. (Hint: Reverse the order of the elements, and use Exercise 22.)

24. Let S be a nonempty (possibly infinite) subset of a vector space $\mathcal{V}$.

(a) Prove that S is linearly independent if and only if *some* vector $\mathbf{v}$ in $\text{span}(S)$ has a unique expression as a linear combination of the vectors in S (ignoring zero coefficients).

(b) The contrapositive of both halves of the "if and only if" statement in part (a), when combined, gives a necessary and sufficient condition for S to be linearly dependent. What is this condition?

25. Suppose $\mathbf{A}$ is an $n \times n$ matrix and that λ is an eigenvalue for $\mathbf{A}$. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a set of fundamental eigenvectors for $\mathbf{A}$ corresponding to λ. Prove that S is linearly independent. (Hint: Consider that each $\mathbf{v}_i$ has a 1 in a coordinate in which all the other vectors in S have a 0.)

26. Suppose T is a linearly independent subset of a vector space $\mathcal{V}$ and that $\mathbf{v} \in \mathcal{V}$.

(a) Prove that if $T \cup \{\mathbf{v}\}$ is linearly dependent, then $\mathbf{v} \in \text{span}(T)$.

(b) Prove that if $\mathbf{v} \in \text{span}(T)$, then $T \cup \{\mathbf{v}\}$ is linearly independent. (Compare this to Exercise 18.)

▶27. Prove Theorem 4.10. (Hint: Generalize the proof of Theorem 4.9. In the first half of the proof, suppose that $\mathbf{v} \in \text{span}(S)$ and that $\mathbf{v}$ can be expressed as both $a_1\mathbf{u}_1 + \cdots + a_k\mathbf{u}_k$ and $b_1\mathbf{v}_1 + \cdots + b_l\mathbf{v}_l$ for distinct $\mathbf{u}_1, \ldots, \mathbf{u}_k$ and distinct $\mathbf{v}_1, \ldots, \mathbf{v}_l$ in S. Consider the union $W = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\} \cup \{\mathbf{v}_1, \ldots, \mathbf{v}_l\}$, and label the distinct vectors in the union as $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$. Then use the given linear combinations to express $\mathbf{v}$ in two ways as a linear combination of the vectors in W. Finally, use the fact that W is a linearly independent set.)

★28. True or False:

(a) The set $\{[2,-3,1], [-8,12,-4]\}$ is a linearly independent subset of $\mathbb{R}^3$.

(b) A set $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ in a vector space $\mathcal{V}$ is linearly dependent if $\mathbf{v}_2$ is a linear combination of $\mathbf{v}_1$ and $\mathbf{v}_3$.

(c) A subset $\mathcal{S} = \{\mathbf{v}\}$ of a vector space $\mathcal{V}$ is linearly dependent if $\mathbf{v} = \mathbf{0}$.

(d) A subset $\mathcal{S}$ of a vector space $\mathcal{V}$ is linearly independent if there is a vector $\mathbf{v} \in \mathcal{S}$ such that $\mathbf{v} \in \text{span}(\mathcal{S} - \{\mathbf{v}\})$.

(e) If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a linearly independent set of vectors in a vector space $\mathcal{V}$, and $a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$, then $a_1 = a_2 = \cdots = a_n = 0$.

(f) If $\mathcal{S}$ is a subset of $\mathbb{R}^4$ containing six vectors, then $\mathcal{S}$ is linearly dependent.

(g) Let $\mathcal{S}$ be a finite nonempty set of vectors in $\mathbb{R}^n$. If the matrix $\mathbf{A}$ whose rows are the vectors in $\mathcal{S}$ has n pivots after row reduction, then $\mathcal{S}$ is linearly independent.

(h) If $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a linearly independent set of a vector space $\mathcal{V}$, then no vector in span($\mathcal{S}$) can be expressed as two different linear combinations of $\mathbf{v}_1, \mathbf{v}_2$, and $\mathbf{v}_3$.

(i) If $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2\}$ is a subset of a vector space $\mathcal{V}$, and $\mathbf{v}_3 = 3\mathbf{v}_1 - 2\mathbf{v}_2$, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is linearly dependent.

4.5 BASIS AND DIMENSION

Suppose that $\mathcal{S}$ is a subset of a vector space $\mathcal{V}$ and that $\mathbf{v}$ is some vector in $\mathcal{V}$. We can ask the following two fundamental questions about $\mathcal{S}$ and $\mathbf{v}$:

Existence: Is there a linear combination of vectors in $\mathcal{S}$ equal to $\mathbf{v}$?

Uniqueness: If so, is this the only such linear combination?

The interplay between existence and uniqueness questions is a pervasive theme throughout mathematics. Answering the existence question is equivalent to determining whether $\mathbf{v} \in \text{span}(\mathcal{S})$. Answering the uniqueness question is equivalent (by Theorem 4.10) to determining whether $\mathcal{S}$ is linearly independent.

We are most interested in cases where both existence and uniqueness occur. In this section, we tie together these concepts by examining those subsets of vector spaces that simultaneously span and are linearly independent. Such a subset is called a **basis**.

Definition of Basis

Definition Let $\mathcal{V}$ be a vector space, and let B be a subset of $\mathcal{V}$. Then B is a **basis** for $\mathcal{V}$ if and only if both of the following are true:

(1) B spans $\mathcal{V}$.

(2) B is linearly independent.

Example 1

We show that $B = \{[1,2,1],[2,3,1],[-1,2,-3]\}$ is a basis for $\mathbb{R}^3$ by showing that it both spans $\mathbb{R}^3$ and is linearly independent.

First, we use the Simplified Span Method in Section 4.3 to show that B spans $\mathbb{R}^3$. Expressing the vectors in B as rows and row reducing the matrix

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 1 \\ -1 & 2 & -3 \end{bmatrix} \quad \text{yields} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

which proves that $\text{span}(B) = \{a[1,0,0] + b[0,1,0] + c[0,0,1] \,|\, a,b,c \in \mathbb{R}\} = \mathbb{R}^3$.

Next, we must show that B is linearly independent. Expressing the vectors in B as columns, and using the Independence Test Method in Section 4.4, we row reduce

$$\begin{bmatrix} 1 & 2 & -1 \\ 2 & 3 & 2 \\ 1 & 1 & -3 \end{bmatrix} \quad \text{to obtain} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Hence, B is also linearly independent.

Since B spans $\mathbb{R}^3$ and is linearly independent, B is a basis for $\mathbb{R}^3$. (B is not the only basis for $\mathbb{R}^3$, as we show in the next example.)

Example 2

The vector space $\mathbb{R}^n$ has $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ as a basis. Although $\mathbb{R}^n$ has other bases as well, the basis $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is the most useful for general applications and is therefore referred to as the **standard basis** for $\mathbb{R}^n$. Thus, we refer to $\{\mathbf{i}, \mathbf{j}\}$ and $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ as the standard bases for $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively.

Each of our fundamental examples of vector spaces also has a "standard basis."

Example 3

The standard basis in $\mathcal{M}_{32}$ is defined as the set

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}.$$

More generally, we define the **standard basis** in $\mathcal{M}_{mn}$ to be the set of $m \cdot n$ distinct matrices

$$\{\Psi_{ij} \,|\, 1 \le i \le m, 1 \le j \le n\},$$

where Ψ_{ij} is the $m \times n$ matrix with 1 in the (i,j) position and zeroes elsewhere. You should check that these $m \cdot n$ matrices are linearly independent and span $\mathcal{M}_{mn}$. In addition to the standard basis, $\mathcal{M}_{mn}$ has many other bases as well.

Example 4

We define $\{1,x,x^2,x^3\}$ to be the standard basis for $\mathcal{P}_3$. More generally, the **standard basis** for $\mathcal{P}_n$ is defined to be the set $\{1,x,x^2,\ldots,x^n\}$, containing $n+1$ elements. Similarly, we define the infinite set $\{1,x,x^2,\ldots\}$ to be the **standard basis** for $\mathcal{P}$. Again, note that in each case these sets both span and are linearly independent.

Of course, the polynomial spaces have other bases. For example, the following is also a basis for $\mathcal{P}_4$:

$$\left\{x^4, x^4 - x^3, x^4 - x^3 + x^2, x^4 - x^3 + x^2 - x, x^3 - 1\right\}.$$

In Exercise 3, you are asked to verify that this is a basis.

Example 5

The empty set, { }, is a basis for the trivial vector space, $\{\mathbf{0}\}$. At the end of Section 4.3, we defined the span of the empty set to be the trivial vector space. That is, { } spans $\{\mathbf{0}\}$. Similarly, at the beginning of Section 4.4, we defined { } to be linearly independent.

A Technical Lemma

In Examples 1 through 4 we saw that $\mathbb{R}^n, \mathcal{P}_n$, and $\mathcal{M}_{mn}$ each have some *finite* set for a basis, while $\mathcal{P}$ has an infinite basis. We will mostly be concerned with those vector spaces that have finite bases. To begin our study of such vector spaces, we need to show that if a vector space has *one* basis that is finite, then *all* of its bases are finite, and all have the same size. Proving this requires some effort. We begin with Lemma 4.11.

In Lemma 4.11, and throughout the remainder of the text, we use the notation $|S|$ to represent the number of elements in a set S. For example, if B is the standard basis for $\mathbb{R}^3$, $|B| = 3$.

Lemma 4.11 Let S and T be subsets of a vector space $\mathcal{V}$ such that S spans $\mathcal{V}$, S is finite, and T is linearly independent. Then T is finite and $|T| \le |S|$.

Proof. If S is empty, then $\mathcal{V} = \{\mathbf{0}\}$. Since $\{\mathbf{0}\}$ is not linearly independent, T is also empty, and so $|T| = |S|$.

Assume that $|S| = n \ge 1$. We will proceed with a proof by contradiction. Suppose that either T is infinite or $|T| > |S| = n$. Then, since every finite subset of T is also linearly independent (see Table 4.1 in Section 4.4), there is a linearly independent set $Y \subseteq T$ such that $|Y| = n+1$. Let $S = \{\mathbf{v}_1,\ldots,\mathbf{v}_n\}$ and let $Y = \{\mathbf{w}_1,\ldots,\mathbf{w}_n,\mathbf{w}_{n+1}\}$. We will show that $\mathbf{w}_{n+1} \in \text{span}(\{\mathbf{w}_1,\ldots,\mathbf{w}_n\})$, which will contradict the linear independence of Y.

Now since $\mathcal{S}$ spans $\mathcal{V}$, there are scalars $a_1, a_2, \ldots, a_n$ such that

$$\mathbf{w}_{n+1} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n.$$

Also, there are scalars c_{ij}, for $1 \le i \le n$ and $1 \le j \le n$, such that

$$\begin{aligned}
\mathbf{w}_1 &= c_{11}\mathbf{v}_1 + c_{12}\mathbf{v}_2 + \cdots + c_{1n}\mathbf{v}_n \\
\mathbf{w}_2 &= c_{21}\mathbf{v}_1 + c_{22}\mathbf{v}_2 + \cdots + c_{2n}\mathbf{v}_n \\
\vdots & \qquad\qquad \vdots \\
\mathbf{w}_n &= c_{n1}\mathbf{v}_1 + c_{n2}\mathbf{v}_2 + \cdots + c_{nn}\mathbf{v}_n.
\end{aligned}$$

Let $\mathbf{C}$ be the $n \times n$ matrix whose (i,j) entry is c_{ij}. Our first step is to prove that $\mathbf{C}^T$ is nonsingular. To do this, we show that the homogeneous system $\mathbf{C}^T\mathbf{x} = \mathbf{0}$ has only the trivial solution. So, let $\mathbf{u}$ represent a solution to the system $\mathbf{C}^T\mathbf{x} = \mathbf{0}$; that is, suppose $\mathbf{C}^T\mathbf{u} = \mathbf{0}$. Then, with $\mathbf{u} = [u_1, \ldots, u_n]$, we have

$$\begin{aligned}
u_1\mathbf{w}_1 + u_2\mathbf{w}_2 + \cdots + u_n\mathbf{w}_n &= u_1(c_{11}\mathbf{v}_1 + c_{12}\mathbf{v}_2 + \cdots + c_{1n}\mathbf{v}_n) \\
&\quad + u_2(c_{21}\mathbf{v}_1 + c_{22}\mathbf{v}_2 + \cdots + c_{2n}\mathbf{v}_n) \\
&\qquad \vdots \\
&\quad + u_n(c_{n1}\mathbf{v}_1 + c_{n2}\mathbf{v}_2 + \cdots + c_{nn}\mathbf{v}_n) \\
&= c_{11}u_1\mathbf{v}_1 + c_{12}u_1\mathbf{v}_2 + \cdots + c_{1n}u_1\mathbf{v}_n \\
&\quad + c_{21}u_2\mathbf{v}_1 + c_{22}u_2\mathbf{v}_2 + \cdots + c_{2n}u_2\mathbf{v}_n \\
&\qquad \vdots \\
&\quad + c_{n1}u_n\mathbf{v}_1 + c_{n2}u_n\mathbf{v}_2 + \cdots + c_{nn}u_n\mathbf{v}_n \\
&= (c_{11}u_1 + c_{21}u_2 + \cdots + c_{n1}u_n)\mathbf{v}_1 \\
&\quad + (c_{12}u_1 + c_{22}u_2 + \cdots + c_{n2}u_n)\mathbf{v}_2 \\
&\qquad \vdots \\
&\quad + (c_{1n}u_1 + c_{2n}u_2 + \cdots + c_{nn}u_n)\mathbf{v}_n.
\end{aligned}$$

But the coefficient of each $\mathbf{v}_i$ in the last expression is just the ith entry of $\mathbf{C}^T\mathbf{u}$. Hence, the coefficient of each $\mathbf{v}_i$ equals 0. Therefore,

$$u_1\mathbf{w}_1 + u_2\mathbf{w}_2 + \cdots + u_n\mathbf{w}_n = 0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_n = \mathbf{0}.$$

Now, $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$, a subset of Y, is linearly independent. Hence, $u_1 = u_2 = \cdots = u_n = 0$. Thus, $\mathbf{u} = \mathbf{0}$, proving that the system $\mathbf{C}^T\mathbf{x} = \mathbf{0}$ has only the trivial solution. From this we conclude that $\mathbf{C}^T$ is nonsingular.

Let $\mathbf{a} = [a_1, \ldots, a_n]$, where $a_1, \ldots, a_n$ are as previously defined. Since $\mathbf{C}^T$ is nonsingular, the system $\mathbf{C}^T\mathbf{x} = \mathbf{a}$ has a unique solution $\mathbf{b}$; that is, there is a vector $\mathbf{b} = [b_1, \ldots, b_n]$ such that $\mathbf{C}^T\mathbf{b} = \mathbf{a}$. Using a computation similar to the above, we get

$$\begin{aligned}
b_1\mathbf{w}_1 + b_2\mathbf{w}_2 + \cdots + b_n\mathbf{w}_n &= b_1(c_{11}\mathbf{v}_1 + c_{12}\mathbf{v}_2 + \cdots + c_{1n}\mathbf{v}_n) \\
&\quad + b_2(c_{21}\mathbf{v}_1 + c_{22}\mathbf{v}_2 + \cdots + c_{2n}\mathbf{v}_n) \\
&\quad\ \vdots \\
&\quad + b_n(c_{n1}\mathbf{v}_1 + c_{n2}\mathbf{v}_2 + \cdots + c_{nn}\mathbf{v}_n) \\
&= (c_{11}b_1 + c_{21}b_2 + \cdots + c_{n1}b_n)\mathbf{v}_1 \\
&\quad + (c_{12}b_1 + c_{22}b_2 + \cdots + c_{n2}b_n)\mathbf{v}_2 \\
&\quad\ \vdots \\
&\quad + (c_{1n}b_1 + c_{2n}b_2 + \cdots + c_{nn}b_n)\mathbf{v}_n.
\end{aligned}$$

Now, the coefficient of each $\mathbf{v}_i$ in the last expression equals the ith coordinate of $\mathbf{C}^T\mathbf{b}$, which equals a_i. Hence,

$$b_1\mathbf{w}_1 + b_2\mathbf{w}_2 + \cdots + b_n\mathbf{w}_n = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n = \mathbf{w}_{n+1}.$$

This proves that $\mathbf{w}_{n+1} \in \text{span}(\{\mathbf{w}_1, \ldots, \mathbf{w}_n\})$, the desired contradiction, completing the proof of the lemma. □

Example 6

Let $T = \{[1,4,3],[2,-7,6],[5,5,-5],[0,3,19]\}$, a subset of $\mathbb{R}^3$. We already know from Theorem 4.7 that because $|T| > 3$, T is linearly dependent. However, Lemma 4.11 gives us the same conclusion because $\{\mathbf{i},\mathbf{j},\mathbf{k}\}$ is a spanning set for $\mathbb{R}^3$ containing three elements, and so the fact that $|T| > 3$ again shows that T is linearly dependent. ■

Dimension

We can now prove the main result of this section.

Theorem 4.12 Let $\mathcal{V}$ be a vector space, and let B_1 and B_2 be bases for $\mathcal{V}$ such that B_1 has finitely many elements. Then B_2 also has finitely many elements, and $|B_1| = |B_2|$.

Proof. Because B_1 and B_2 are bases for $\mathcal{V}$, B_1 spans $\mathcal{V}$ and B_2 is linearly independent. Hence, Lemma 4.11 shows that B_2 has finitely many elements and $|B_2| \le |B_1|$. Now, since B_2 is finite, we can reverse the roles of B_1 and B_2 in this argument to show that $|B_1| \le |B_2|$. Therefore, $|B_1| = |B_2|$. □

It follows from Theorem 4.12 that if a vector space $\mathcal{V}$ has *one* basis containing a finite number of elements, then *every* basis for $\mathcal{V}$ is finite, and all bases for $\mathcal{V}$ have the same number of elements. This allows us to unambiguously define the **dimension** of such a vector space, as follows:

Definition Let $\mathcal{V}$ be a vector space. If $\mathcal{V}$ has a basis B containing a finite number of elements, then $\mathcal{V}$ is said to be **finite dimensional**. In this case, the **dimension** of $\mathcal{V}$, $\dim(\mathcal{V})$, is the number of elements in any basis for $\mathcal{V}$. In particular, $\dim(\mathcal{V}) = |B|$.

If $\mathcal{V}$ has no finite basis, then $\mathcal{V}$ is **infinite dimensional**.

Example 7

Because $\mathbb{R}^3$ has the (standard) basis $\{\mathbf{i},\mathbf{j},\mathbf{k}\}$, the dimension of $\mathbb{R}^3$ is 3. Theorem 4.12 then implies that every other basis for $\mathbb{R}^3$ also has exactly three elements. More generally, $\dim(\mathbb{R}^n) = n$, since $\mathbb{R}^n$ has the basis $\{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$.

Example 8

Because the standard basis $\{1,x,x^2,x^3\}$ for $\mathcal{P}_3$ has four elements, $\dim(\mathcal{P}_3) = 4$. Every other basis for $\mathcal{P}_3$, such as $\left\{x^3 - x, x^2 + x + 1, x^3 + x - 5, 2x^3 + x^2 + x - 3\right\}$, also has four elements. (Verify that this set is a basis for $\mathcal{P}_3$.)

Also, $\dim(\mathcal{P}_n) = n + 1$, since $\mathcal{P}_n$ has the basis $\left\{1,x,x^2,\ldots,x^n\right\}$, containing $n + 1$ elements. Be careful! Many students *erroneously* believe that the dimension of $\mathcal{P}_n$ is n because of the subscript n.

Example 9

The standard basis for $\mathcal{M}_{22}$ contains four elements. Hence, $\dim(\mathcal{M}_{22}) = 4$. In general, from the size of the standard basis for $\mathcal{M}_{mn}$, we see that $\dim(\mathcal{M}_{mn}) = m \cdot n$.

Example 10

Let $\mathcal{V} = \{\mathbf{0}\}$ be the trivial vector space. Then $\dim(\mathcal{V}) = 0$ because the empty set, which contains no elements, is a basis for $\mathcal{V}$.

Example 11

Consider the following subsets of $\mathbb{R}^4$:

$$S_1 = \{[1,3,1,2],[3,11,5,10],[-2,4,4,4]\} \text{ and}$$

$$S_2 = \{[1,5,-2,3],[-2,-8,8,8],[1,1,-10,-2],[0,2,4,-9],[3,13,-10,-8]\}.$$

Since $\dim(\mathbb{R}^4) = 4$, $|S_1| = 3$, and $|S_2| = 5$, Theorem 4.12 shows us that neither S_1 nor S_2 is a basis for $\mathbb{R}^4$. In particular, S_1 cannot span $\mathbb{R}^4$ because the standard basis for $\mathbb{R}^4$ would then be a linearly independent set that is larger than S_1, contradicting Lemma 4.11. Similarly, S_2 cannot be linearly independent because the standard basis would be a spanning set that is smaller than S_2, again contradicting Lemma 4.11.

Notice, however, that in this case we can make no conclusions regarding whether S_1 is linearly independent or whether S_2 spans $\mathbb{R}^4$ based solely on the size of these sets. We must check for these properties separately using the techniques of Sections 4.3 and 4.4. ■

Sizes of Spanning Sets and Linearly Independent Sets

Example 11 illustrates the next result, which summarizes much of what we have learned regarding the sizes of spanning sets and linearly independent sets.

Theorem 4.13 Let $\mathcal{V}$ be a finite dimensional vector space.

(1) Suppose S is a finite subset of $\mathcal{V}$ that spans $\mathcal{V}$. Then $\dim(\mathcal{V}) \leq |S|$. Moreover, $|S| = \dim(\mathcal{V})$ if and only if S is a basis for $\mathcal{V}$.

(2) Suppose T is a linearly independent subset of $\mathcal{V}$. Then T is finite and $|T| \leq \dim(\mathcal{V})$. Moreover, $|T| = \dim(\mathcal{V})$ if and only if T is a basis for $\mathcal{V}$.

Proof. Let B be a basis for $\mathcal{V}$ with $|B| = n$. Then $\dim(\mathcal{V}) = |B|$, by definition.

Part (1): Since S is a finite spanning set and B is linearly independent, Lemma 4.11 implies that $|B| \leq |S|$, and so $\dim(\mathcal{V}) \leq |S|$.

If $|S| = \dim(\mathcal{V})$, we prove that S is a basis for $\mathcal{V}$ by contradiction. If S is not a basis, then it is not linearly independent (because it spans). So, by Exercise 12 in Section 4.4 (see Table 4.1), there is a redundant vector in S—that is, a vector $\mathbf{v}$ such that $\text{span}(S - \{\mathbf{v}\}) = \text{span}(S) = \mathcal{V}$. But then $S - \{\mathbf{v}\}$ is a spanning set for $\mathcal{V}$ having fewer than n elements, contradicting the fact that we just observed that the size of a spanning set is never less than the dimension.

Finally, suppose S is a basis for $\mathcal{V}$. By Theorem 4.12, S is finite, and $|S| = \dim(\mathcal{V})$ by the definition of dimension.

Part (2): Using B as the spanning set S in Lemma 4.11 proves that T is finite and $|T| \leq \dim(\mathcal{V})$.

If $|T| = \dim(\mathcal{V})$, we prove that T is a basis for $\mathcal{V}$ by contradiction. If T is not a basis for $\mathcal{V}$, then T does not span $\mathcal{V}$ (because it is linearly independent). Therefore, there is a vector $\mathbf{v} \in \mathcal{V}$ such that $\mathbf{v} \notin \text{span}(T)$. Hence, by part (b) of Exercise 26 in Section 4.4, $T \cup \{\mathbf{v}\}$ is also linearly independent. But $T \cup \{\mathbf{v}\}$ has $n + 1$ elements, contradicting the fact we just proved — that a linearly independent subset must have size $\leq \dim(\mathcal{V})$.

Finally, if T is a basis for $\mathcal{V}$, then $|T| = \dim(\mathcal{V})$, by the definition of dimension. □

Example 12

Recall the subset $B = \{[1,2,1],[2,3,1],[-1,2,-3]\}$ of $\mathbb{R}^3$ from Example 1. In that example, after showing that B spans $\mathbb{R}^3$, we could have immediately concluded that B is a basis for $\mathbb{R}^3$ without having proved linear independence by using part (1) of Theorem 4.13 because B is a spanning set with $\dim(\mathbb{R}^3) = 3$ elements.

Similarly, consider $T = \{3, x+5, x^2 - 7x + 12, x^3 + 4\}$, a subset of $\mathcal{P}_3$. T is linearly independent from Exercise 22 in Section 4.4 (see Table 4.1) because each vector in T is not in the span of those before it. Since $|T| = 4 = \dim(\mathcal{P}_3)$, part (2) of Theorem 4.13 shows that T is a basis for $\mathcal{P}_3$. ■

Maximal Linearly Independent Sets and Minimal Spanning Sets

Theorem 4.13 shows that in a finite dimensional vector space, a large enough linearly independent set is a basis, as is a small enough spanning set. The "borderline" size is the dimension of the vector space. No linearly independent sets are larger than this, and no spanning sets are smaller. The next two results illustrate this same principle without explicitly using the dimension. Thus, they are useful in cases in which the dimension is not known or for infinite dimensional vector spaces. Outlines of their proofs are given in Exercises 18 and 19.

Theorem 4.14 Let $\mathcal{V}$ be a vector space with spanning set S (so, span(S) = $\mathcal{V}$), and let B be a maximal linearly independent subset of S. Then B is a basis for $\mathcal{V}$.

The phrase "B is a **maximal linearly independent subset** of S" means that both of the following are true:

- B is a linearly independent subset of S.
- If $B \subset C \subseteq S$ and $B \neq C$, then C is linearly dependent.

Theorem 4.14 asserts that if there is no way to include another vector from S in B without making B linearly dependent, then B is a basis for span(S) $= \mathcal{V}$. The converse to Theorem 4.14 is also true (see Exercise 20).

Example 13

Consider the subset $S = \{[1,-2,1], [3,1,-2], [5,-3,0], [5,4,-5], [0,0,0]\}$ of $\mathbb{R}^3$ and the subset $B = \{[1,-2,1], [5,-3,0]\}$ of S. We show that B is a maximal linearly independent subset of S and hence, by Theorem 4.14, it is a basis for $\mathcal{V} = $ span(S).

Now, B is a linearly independent subset of S. The following equations show that if any of the remaining vectors of S are added to B, the set is no longer linearly independent:

$$\begin{aligned} [3,1,-2] &= -2[1,-2,1] + [5,-3,0] \\ [5,4,-5] &= -5[1,-2,1] + 2[5,-3,0] \\ [0,0,0] &= 0[1,-2,1] + 0[5,-3,0]. \end{aligned}$$

Thus, B is a maximal linearly independent subset of S and so is a basis for span(S). ■

Another consequence of Theorem 4.14 is that any vector space $\mathcal{V}$ having a finite spanning set S must be finite dimensional. This is because a maximal linearly independent subset of S, which must also be finite, is a basis for $\mathcal{V}$ (see Exercise 24).

We also have the following result for spanning sets:

Theorem 4.15 Let $\mathcal{V}$ be a vector space, and let B be a minimal spanning set for $\mathcal{V}$. Then B is a basis for $\mathcal{V}$.

The phrase "B is a **minimal spanning set** for $\mathcal{V}$" means that both of the following are true:

- B is a subset of $\mathcal{V}$ that spans $\mathcal{V}$.
- If $C \subset B$ and $C \neq B$, then C does not span $\mathcal{V}$.

The converse of Theorem 4.15 is true as well (see Exercise 21).

Example 14

Consider the subsets S and B of $\mathbb{R}^3$ given in Example 13. We can use Theorem 4.15 to give another justification that B is a basis for $\mathcal{V} = \text{span}(S)$. Recall from Example 13 that every vector in S is a linear combination of vectors in B, so $S \subseteq \text{span}(B)$. This fact along with $B \subseteq S$ and Corollary 4.6 shows that $\text{span}(B) = \text{span}(S) = \mathcal{V}$. Also, neither vector in B is a scalar multiple of the other, so that neither vector alone can span $\mathcal{V}$ (why?). Hence, B is a minimal spanning set for $\mathcal{V}$, and by Theorem 4.15, B is a basis for $\text{span}(S)$. ■

Dimension of a Subspace

We conclude this section with the result that every subspace of a finite dimensional vector space is also finite dimensional. This is important because it tells us that the theorems we have developed about finite dimension apply to all *subspaces* of our basic examples $\mathbb{R}^n$, $\mathcal{M}_{mn}$, and $\mathcal{P}_n$.

Theorem 4.16 Let $\mathcal{V}$ be a finite dimensional vector space, and let $\mathcal{W}$ be a subspace of $\mathcal{V}$. Then $\mathcal{W}$ is also finite dimensional with $\dim(\mathcal{W}) \leq \dim(\mathcal{V})$. Moreover, $\dim(\mathcal{W}) = \dim(\mathcal{V})$ if and only if $\mathcal{W} = \mathcal{V}$.

The proof of Theorem 4.16 is left for you to do, with hints, in Exercise 22. The only subtle part of this proof involves showing that $\mathcal{W}$ actually has a basis.[4]

[4] Although it is true that *every* vector space has a basis, we must be careful here, because we have not proven this. In fact, Theorem 4.16 establishes that every subspace of a finite dimensional vector space *does* have a basis and that this basis is finite. Although every finite dimensional vector space has a finite basis by definition, the proof that every infinite dimensional vector space has a basis requires advanced set theory and is beyond the scope of this text.

Example 15

Consider the nested sequence of subspaces of $\mathbb{R}^3$ given by $\{\mathbf{0}\} \subset$ {scalar multiples of $[4, -7, 0]$} $\subset$ xy-plane $\subset \mathbb{R}^3$. Their respective dimensions are $0, 1, 2$, and 3 (why?). Hence, the dimensions of each successive pair of these subspaces satisfy the inequality given in Theorem 4.16.

Example 16

It can be shown that $B = \{x^3 + 2x^2 - 4x + 18, 3x^2 + 4x - 4, x^3 + 5x^2 - 3, 3x + 2\}$ is a linearly independent subset of $\mathcal{P}_3$. Therefore, by part (2) of Theorem 4.13, B is a basis for $\mathcal{P}_3$. However, we can also reach the same conclusion from Theorem 4.16. For, $\mathcal{W} = \text{span}(B)$ has B as a basis (why?), and hence, $\dim(\mathcal{W}) = 4$. But since $\mathcal{W}$ is a subspace of $\mathcal{P}_3$ and $\dim(\mathcal{P}_3) = 4$, Theorem 4.16 implies that $\mathcal{W} = \mathcal{P}_3$. Hence, B is a basis for $\mathcal{P}_3$.

New Vocabulary

basis
dimension
finite dimensional (vector space)
infinite dimensional (vector space)
maximal linearly independent set
minimal spanning set
standard basis (for $\mathbb{R}^n, \mathcal{M}_{mn}, \mathcal{P}_n$)

Highlights

- A basis is a subset of a vector space that both spans and is linearly independent.
- If a finite basis exists for a vector space, the vector space is said to be finite dimensional.
- For a finite dimensional vector space, all bases have the same number of vectors, and this number is known as the dimension of the vector space.
- The standard basis for $\mathbb{R}^n$ is $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$; $\dim(\mathbb{R}^n) = n$.
- The standard basis for $\mathcal{P}_n$ is $\{1, x, x^2, \ldots, x^n\}$; $\dim(\mathcal{P}_n) = n + 1$.
- The standard basis for $\mathcal{M}_{mn}$ is $\{\Psi_{ij}\}$, where each Ψ_{ij} has a 1 in the (i,j) entry, and zeroes elsewhere; $\dim(\mathcal{M}_{mn}) = m \cdot n$.
- The basis for the trivial vector space $\{\mathbf{0}\}$ is the empty set $\{\ \}$; $\dim(\{\mathbf{0}\}) = 0$.
- If no finite basis exists for a vector space, the vector space is said to be infinite dimensional. $\mathcal{P}$ is an infinite dimensional vector space, as is the set of all real-valued functions (under normal operations).
- In a vector space $\mathcal{V}$ with dimension n, the size of a spanning set S is always $\geq n$. If $|S| = n$, then S is a basis for $\mathcal{V}$.
- In a vector space $\mathcal{V}$ with dimension n, the size of a linearly independent set T is always $\leq n$. If $|T| = n$, then T is a basis for $\mathcal{V}$.

- A maximal linearly independent set in a vector space is a basis.
- A minimal spanning set in a vector space is a basis.
- In a vector space $\mathcal{V}$ with dimension n, the dimension of a subspace $\mathcal{W}$ is always $\leq n$. If $\dim(\mathcal{W}) = n$, then $\mathcal{W} = \mathcal{V}$.

EXERCISES FOR SECTION 4.5

1. Prove that each of the following subsets of $\mathbb{R}^4$ is a basis for $\mathbb{R}^4$ by showing both that it spans $\mathbb{R}^4$ and that is linearly independent:

(a) $\{[2,1,0,0],[0,1,1,-1],[0,-1,2,-2],[3,1,0,-2]\}$

(b) $\{[6,1,1,-1],[1,0,0,9],[-2,3,2,4],[2,2,5,-5]\}$

(c) $\{[1,1,1,1],[1,1,1,-1],[1,1,-1,-1],[1,-1,-1,-1]\}$

(d) $\left\{\left[\frac{15}{2},5,\frac{12}{5},1\right],\left[2,\frac{1}{2},\frac{3}{4},1\right],\left[-\frac{13}{2},1,0,4\right],\left[\frac{18}{5},0,\frac{1}{5},-\frac{1}{5}\right]\right\}$

2. Prove that the following set is a basis for $\mathcal{M}_{22}$ by showing that it spans $\mathcal{M}_{22}$ and is linearly independent:

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

3. Show that the subset $\left\{x^4,\ x^4-x^3,\ x^4-x^3+x^2,\ x^4-x^3+x^2-x,\ x^3-1\right\}$ of $\mathcal{P}_4$ is a basis for $\mathcal{P}_4$.

4. Determine which of the following subsets of $\mathbb{R}^4$ form a basis for $\mathbb{R}^4$:

★**(a)** $S = \{[7,1,2,0],[8,0,1,-1],[1,0,0,-2]\}$

(b) $S = \{[1,3,2,0],[-2,0,6,7],[0,6,10,7]\}$

★**(c)** $S = \{[7,1,2,0],[8,0,1,-1],[1,0,0,-2],[3,0,1,-1]\}$

(d) $S = \{[1,3,2,0],[-2,0,6,7],[0,6,10,7],[2,10,-3,1]\}$

★**(e)** $S = \{[1,2,3,2],[1,4,9,3],[6,-2,1,4],[3,1,2,1],[10,-9,-15,6]\}$

5. **(a)** Show that $B = \{[2,3,0,-1],[-1,1,1,-1]\}$ is a maximal linearly independent subset of $S = \{[1,4,1,-2],[-1,1,1,-1],[3,2,-1,0],[2,3,0,-1]\}$.

★**(b)** Calculate $\dim(\text{span}(S))$.

★**(c)** Does $\text{span}(S) = \mathbb{R}^4$? Why or why not?

(d) Is B a minimal spanning set for $\text{span}(S)$? Why or why not?

6. **(a)** Show that $B = \{x^3-x^2+2x+1,\ 2x^3+4x-7,\ 3x^3-x^2-6x+6\}$ is a maximal linearly independent subset of $S = \{x^3-x^2+2x+1,\ x-1,\ 2x^3+4x-7, x^3-3x^2-22x+34, 3x^3-x^2-6x+6\}$.

(b) Calculate $\dim(\text{span}(S))$.

(c) Does span(S) $= \mathcal{P}_3$? Why or why not?

(d) Is B a minimal spanning set for span(S)? Why or why not?

7. Let $\mathcal{W}$ be the solution set to the matrix equation $\mathbf{AX} = \mathbf{O}$, where

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

(a) Show that $\mathcal{W}$ is a subspace of $\mathbb{R}^5$.

(b) Find a basis for $\mathcal{W}$.

(c) Show that $\dim(\mathcal{W}) + \text{rank}(\mathbf{A}) = 5$.

8. Prove that every proper nontrivial subspace of $\mathbb{R}^3$ can be thought of, from a geometric point of view, as either a line through the origin or a plane through the origin.

9. Let $\mathbf{f}$ be a polynomial of degree n. Show that the set $\{\mathbf{f}, \mathbf{f}^{(1)}, \mathbf{f}^{(2)}, \ldots, \mathbf{f}^{(n)}\}$ is a basis for $\mathcal{P}_n$ (where $\mathbf{f}^{(i)}$ denotes the ith derivative of $\mathbf{f}$). (Hint: See Exercise 23 in Section 4.4.)

10. (a) Let $\mathbf{A}$ be a 2×2 matrix. Prove that there are real numbers $a_0, a_1, \ldots, a_4$, not all zero, such that $a_4\mathbf{A}^4 + a_3\mathbf{A}^3 + a_2\mathbf{A}^2 + a_1\mathbf{A} + a_0\mathbf{I}_2 = \mathbf{O}_2$. (Hint: You can assume that $\mathbf{A}, \mathbf{A}^2, \mathbf{A}^3, \mathbf{A}^4$, and $\mathbf{I}_2$ are all distinct because if they are not, opposite nonzero coefficients can be chosen for any identical pair to demonstrate that the given statement holds.)

(b) Suppose $\mathbf{B}$ is an $n \times n$ matrix. Show that there must be a nonzero polynomial $\mathbf{p} \in \mathcal{P}_{n^2}$ such that $\mathbf{p}(\mathbf{B}) = \mathbf{O}_n$.

11. (a) Show that $B = \{(x-2), x(x-2), x^2(x-2), x^3(x-2), x^4(x-2)\}$ is a basis for $\mathcal{V} = \{\mathbf{p} \in \mathcal{P}_5 \,|\, \mathbf{p}(2) = 0\}$.

★(b) What is $\dim(\mathcal{V})$?

★(c) Find a basis for $\mathcal{W} = \{\mathbf{p} \in \mathcal{P}_5 \mid \mathbf{p}(2) = \mathbf{p}(3) = 0\}$.

★(d) Calculate $\dim(\mathcal{W})$.

★12. Let $\mathcal{V}$ be a finite dimensional vector space.

(a) Let S be a subset of $\mathcal{V}$ with $\dim(\mathcal{V}) \le |S|$. Find an example to show that S need not span $\mathcal{V}$.

(b) Let T be a subset of $\mathcal{V}$ with $|T| \le \dim(\mathcal{V})$. Find an example to show that T need not be linearly independent.

13. Let S be a subset of a finite dimensional vector space $\mathcal{V}$ such that $|S| = \dim(\mathcal{V})$. If S is not a basis for $\mathcal{V}$, prove that S neither spans $\mathcal{V}$ nor is linearly independent.

14. Let $\mathcal{V}$ be an n-dimensional vector space, and let S be a subset of $\mathcal{V}$ containing exactly n elements. Prove that S spans $\mathcal{V}$ if and only if S is linearly independent.

15. Let $\mathbf{A}$ be a nonsingular $n \times n$ matrix, and let B be a basis for $\mathbb{R}^n$.

(a) Show that $B_1 = \{\mathbf{Av} | \mathbf{v} \in B\}$ is also a basis for $\mathbb{R}^n$. (Treat the vectors in B as column vectors.)

(b) Show that $B_2 = \{\mathbf{vA} | \mathbf{v} \in B\}$ is also a basis for $\mathbb{R}^n$. (Treat the vectors in B as row vectors.)

(c) Letting B be the standard basis for $\mathbb{R}^n$, use the result of part (a) to show that the columns of $\mathbf{A}$ form a basis for $\mathbb{R}^n$.

(d) Prove that the rows of $\mathbf{A}$ form a basis for $\mathbb{R}^n$.

16. Prove that $\mathcal{P}$ is infinite dimensional by showing that no finite subset S of $\mathcal{P}$ can span $\mathcal{P}$, as follows:

(a) Let S be a finite subset of $\mathcal{P}$. Show that $S \subseteq \mathcal{P}_n$, for some n.

(b) Use part (a) to prove that $\text{span}(S) \subseteq \mathcal{P}_n$.

(c) Conclude that S cannot span $\mathcal{P}$.

17. **(a)** Prove that if a vector space $\mathcal{V}$ has an infinite linearly independent subset, then $\mathcal{V}$ is not finite dimensional.

(b) Use part (a) to prove that any vector space having $\mathcal{P}$ as a subspace is not finite dimensional.

18. The purpose of this exercise is to prove Theorem 4.14. Let $\mathcal{V}$, S, and B be as given in the statement of the theorem. Suppose $B \neq S$, and $\mathbf{w} \in S$ with $\mathbf{w} \notin B$.

(a) Explain why it is sufficient to prove that B spans $\mathcal{V}$.

▶**(b)** Prove that if $S \subseteq \text{span}(B)$, then B spans $\mathcal{V}$.

▶**(c)** Let $C = B \cup \{\mathbf{w}\}$. Prove that C is linearly dependent.

(d) Use part (c) to prove that $\mathbf{w} \in \text{span}(B)$. (Also see part (a) of Exercise 26 in Section 4.4.)

(e) Tie together all parts to finish the proof.

19. The purpose of this exercise is to prove Theorem 4.15.

(a) Explain why it is sufficient to prove the following statement: Let S be a spanning set for a vector space $\mathcal{V}$. If S is a minimal spanning set for $\mathcal{V}$, then S is linearly independent.

▶**(b)** State the contrapositive of the statement in part (a).

▶**(c)** Prove the statement from part (b). (Hint: Use Exercise 12 from Section 4.4.)

20. Let B be a basis for a vector space $\mathcal{V}$. Prove that B is a maximal linearly independent subset of $\mathcal{V}$. (Note: You may *not* use $\dim(\mathcal{V})$ in your proof, since $\mathcal{V}$ could be infinite dimensional.)

21. Let B be a basis for a vector space $\mathcal{V}$. Prove that B is a minimal spanning set for $\mathcal{V}$. (Note: You may *not* use $\dim(\mathcal{V})$ in your proof, since $\mathcal{V}$ could be infinite dimensional.)

22. The purpose of this exercise is to prove Theorem 4.16. Let $\mathcal{V}$ and $\mathcal{W}$ be as given in the theorem. Consider the set A of nonnegative integers defined by $A = \{k \,|\, \text{a set } T \text{ exists with } T \subseteq \mathcal{W}, |T| = k, \text{and } T \text{ linearly independent}\}$.

- **(a)** Prove that $0 \in A$. (Hence, A is nonempty.)
- **(b)** Prove that $k \in A$ implies $k \leq \dim(\mathcal{V})$. (Hint: Use Theorem 4.13.) (Hence, A is finite.)
- ▶**(c)** Let n be the largest element of A. Let T be a linearly independent subset of $\mathcal{W}$ such that $|T| = n$. Prove T is a maximal linearly independent subset of $\mathcal{W}$.
- ▶**(d)** Use part (c) and Theorem 4.14 to prove that T is a basis for $\mathcal{W}$.
- **(e)** Conclude that $\mathcal{W}$ is finite dimensional and use part (b) to show $\dim(\mathcal{W}) \leq \dim(\mathcal{V})$.
- **(f)** Prove that if $\dim(\mathcal{W}) = \dim(\mathcal{V})$, then $\mathcal{W} = \mathcal{V}$. (Hint: Let T be a basis for $\mathcal{W}$ and use part (2) of Theorem 4.13 to show that T is also a basis for $\mathcal{V}$.)
- **(g)** Prove the converse of part (f).

23. Let $\mathcal{V}$ be a subspace of $\mathbb{R}^n$ with $\dim(\mathcal{V}) = n - 1$. (Such a subspace is called a **hyperplane** in $\mathbb{R}^n$.) Prove that there is a nonzero $\mathbf{x} \in \mathbb{R}^n$ such that $\mathcal{V} = \{\mathbf{v} \in \mathbb{R}^n \,|\, \mathbf{x} \cdot \mathbf{v} = 0\}$. (Hint: Set up a homogeneous system of equations whose coefficient matrix has a basis for $\mathcal{V}$ as its rows. Then notice that this $(n-1) \times n$ system has at least one nontrivial solution, say $\mathbf{x}$.)

24. Let $\mathcal{V}$ be a vector space and let S be a finite spanning set for $\mathcal{V}$. Prove that $\mathcal{V}$ is finite dimensional.

★**25.** True or False:

- **(a)** A set B of vectors in a vector space $\mathcal{V}$ is a basis for $\mathcal{V}$ if B spans $\mathcal{V}$ and B is linearly independent.
- **(b)** All bases for $\mathcal{P}_4$ have four elements.
- **(c)** $\dim(\mathcal{M}_{43}) = 7$.
- **(d)** If S is a spanning set for $\mathcal{W}$ and $\dim(\mathcal{W}) = n$, then $|S| \leq n$.
- **(e)** If T is a linearly independent set in $\mathcal{W}$ and $\dim(\mathcal{W}) = n$, then $|T| = n$.
- **(f)** If T is a linearly independent set in a finite dimensional vector space $\mathcal{W}$ and S is a finite spanning set for $\mathcal{W}$, then $|T| \leq |S|$.
- **(g)** If $\mathcal{W}$ is a subspace of a finite dimensional vector space $\mathcal{V}$, then $\dim(\mathcal{W}) < \dim(\mathcal{V})$.
- **(h)** Every subspace of an infinite dimensional vector space is infinite dimensional.

(i) If T is a maximal linearly independent set for a vector space $\mathcal{V}$ and S is a minimal spanning set for $\mathcal{V}$, then $S = T$.

(j) If $\mathbf{A}$ is a nonsingular 4×4 matrix, then the rows of $\mathbf{A}$ are a basis for $\mathbb{R}^4$.

4.6 CONSTRUCTING SPECIAL BASES

In this section, we present additional methods for finding a basis for a given finite dimensional vector space, starting with either a spanning set or a linearly independent subset.

Using Row Reduction to Construct a Basis

Recall the Simplified Span Method from Section 4.3. Using that method, we were able to simplify the form of span(S) for a subset S of $\mathbb{R}^n$. This was done by creating a matrix $\mathbf{A}$ whose rows are the vectors in S, and then row reducing $\mathbf{A}$ to obtain a reduced row echelon form matrix $\mathbf{C}$. We discovered that a simplified form of span(S) is given by the set of all linear combinations of the nonzero rows of $\mathbf{C}$. Now, each nonzero row of the matrix $\mathbf{C}$ has a (pivot) 1 in a column in which all other rows have zeroes, so the nonzero rows of $\mathbf{C}$ must be linearly independent. Thus, the nonzero rows of $\mathbf{C}$ not only span S but are linearly independent as well, and so they form a basis for span(S). Therefore, whenever we use the Simplified Span Method on a subset S of $\mathbb{R}^n$, we are actually creating a basis for span(S).

Example 1

Let $S = \{[2,-2,3,5,5], [-1,1,4,14,-8], [4,-4,-2,-14,18], [3,-3,-1,-9,13]\}$, a subset of $\mathbb{R}^5$. We can use the Simplified Span Method to find a basis B for $\mathcal{V} = \text{span}(S)$. We construct the matrix

$$\mathbf{A} = \begin{bmatrix} 2 & -2 & 3 & 5 & 5 \\ -1 & 1 & 4 & 14 & -8 \\ 4 & -4 & -2 & -14 & 18 \\ 3 & -3 & -1 & -9 & 13 \end{bmatrix},$$

whose rows are the vectors in S. The reduced row echelon form matrix for $\mathbf{A}$ is

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

Therefore, the desired basis for $\mathcal{V}$ is the set $B = \{[1,-1,0,-2,4], [0,0,1,3,-1]\}$ of nonzero rows of $\mathbf{C}$, and $\dim(\mathcal{V}) = 2$. ■

CHAPTER

Linear Transformations

5

TRANSFORMING SPACE

Although a vector can be used to indicate a particular type of movement, actual vectors themselves are essentially static, unchanging objects. For example, if we represent the edges of a particular image on a computer screen by vectors, then these vectors are fixed in place. However, when we want to move or alter the image in some way, such as rotating it about a point on the screen, we need a function to calculate the new position for each of the original vectors.

This suggests that we need another "tool" in our arsenal: functions that move a given set of vectors in a prescribed "linear" manner. Such functions are called linear transformations. Just as we saw in Chapter 4 that general vector spaces are abstract generalizations of $\mathbb{R}^n$, we will find in this chapter that linear transformations are the corresponding abstract generalization of matrix multiplication.

In this chapter, we study functions that map the vectors in one vector space to those in another. We concentrate on a special class of these functions, known as linear transformations. The formal definition of a linear transformation is introduced in Section 5.1 along with several of its fundamental properties. In Section 5.2, we show that the effect of any linear transformation is equivalent to multiplication by a corresponding matrix. In Section 5.3, we examine an important relationship between the dimensions of the domain and the range of a linear transformation, known as the Dimension Theorem. In Section 5.4, we introduce two special types of linear transformations: one-to-one and onto. In Section 5.5, these two types of linear transformations are combined to form isomorphisms, which are used to establish that all n-dimensional vector spaces are in some sense equivalent. Finally, in Section 5.6, we return to the topic of eigenvalues and eigenvectors to study them in the context of linear transformations.

5.1 INTRODUCTION TO LINEAR TRANSFORMATIONS

In this section, we introduce linear transformations and examine their elementary properties.

Functions

If you are not familiar with the terms *domain*, *codomain*, *range*, *image*, and *pre-image* in the context of functions, read Appendix B before proceeding. The following example illustrates some of these terms:

Example 1

Let $f\colon \mathcal{M}_{23} \to \mathcal{M}_{22}$ be given by

$$f\left(\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}\right) = \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix}.$$

Then f is a function that maps one vector space to another. The domain of f is $\mathcal{M}_{23}$, the codomain of f is $\mathcal{M}_{22}$, and the range of f is the set of all 2×2 matrices with second row entries equal to zero. The image of $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ under f is $\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$. The matrix $\begin{bmatrix} 1 & 2 & 10 \\ 11 & 12 & 13 \end{bmatrix}$ is one of the pre-images of $\begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$ under f. Also, the image under f of the set S of all matrices of the form $\begin{bmatrix} 7 & * & * \\ * & * & * \end{bmatrix}$ (where "$*$" represents any real number) is the set $f(S)$ containing all matrices of the form $\begin{bmatrix} 7 & * \\ 0 & 0 \end{bmatrix}$. Finally, the pre-image under f of the set T of all matrices of the form $\begin{bmatrix} a & a+2 \\ 0 & 0 \end{bmatrix}$ is the set $f^{-1}(T)$ consisting of all matrices of the form $\begin{bmatrix} a & a+2 & * \\ * & * & * \end{bmatrix}$. ■

Linear Transformations

Definition Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces, and let $f\colon \mathcal{V} \to \mathcal{W}$ be a function from $\mathcal{V}$ to $\mathcal{W}$. (That is, for each vector $\mathbf{v} \in \mathcal{V}$, $f(\mathbf{v})$ denotes exactly one vector of $\mathcal{W}$.) Then f is a **linear transformation** if and only if both of the following are true:

(1) $f(\mathbf{v}_1 + \mathbf{v}_2) = f(\mathbf{v}_1) + f(\mathbf{v}_2)$, for all $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}$

(2) $f(c\mathbf{v}) = cf(\mathbf{v})$, for all $c \in \mathbb{R}$ and all $\mathbf{v} \in \mathcal{V}$.

Properties (1) and (2) insist that the operations of addition and scalar multiplication give the same result on vectors whether the operations are performed before

f is applied (in $\mathcal{V}$) or after f is applied (in $\mathcal{W}$). Thus, a linear transformation is a function between vector spaces that "preserves" the operations that give structure to the spaces.

To determine whether a given function f from a vector space $\mathcal{V}$ to a vector space $\mathcal{W}$ is a linear transformation, we need only verify properties (1) and (2) in the definition, as in the next three examples.

Example 2

Consider the mapping $f\colon \mathcal{M}_{mn} \to \mathcal{M}_{nm}$, given by $f(\mathbf{A}) = \mathbf{A}^T$ for any $m \times n$ matrix $\mathbf{A}$. We will show that f is a linear transformation.

(1) We must show that $f(\mathbf{A}_1 + \mathbf{A}_2) = f(\mathbf{A}_1) + f(\mathbf{A}_2)$, for matrices $\mathbf{A}_1, \mathbf{A}_2 \in \mathcal{M}_{mn}$. However, $f(\mathbf{A}_1 + \mathbf{A}_2) = (\mathbf{A}_1 + \mathbf{A}_2)^T = \mathbf{A}_1^T + \mathbf{A}_2^T$ (by part (2) of Theorem 1.12) $= f(\mathbf{A}_1) + f(\mathbf{A}_2)$.

(2) We must show that $f(c\mathbf{A}) = cf(\mathbf{A})$, for all $c \in \mathbb{R}$ and for all $\mathbf{A} \in \mathcal{M}_{mn}$. However, $f(c\mathbf{A}) = (c\mathbf{A})^T = c(\mathbf{A}^T)$ (by part (3) of Theorem 1.12) $= cf(\mathbf{A})$.

Hence, f is a linear transformation. ■

Example 3

Consider the function $g: \mathcal{P}_n \to \mathcal{P}_{n-1}$ given by $g(\mathbf{p}) = \mathbf{p}'$, the derivative of $\mathbf{p}$. We will show that g is a linear transformation.

(1) We must show that $g(\mathbf{p}_1 + \mathbf{p}_2) = g(\mathbf{p}_1) + g(\mathbf{p}_2)$, for all $\mathbf{p}_1, \mathbf{p}_2 \in \mathcal{P}_n$. Now, $g(\mathbf{p}_1 + \mathbf{p}_2) = \left(\mathbf{p}_1 + \mathbf{p}_2\right)'$. From calculus we know that the derivative of a sum is the sum of the derivatives, so $\left(\mathbf{p}_1 + \mathbf{p}_2\right)' = \mathbf{p}_1' + \mathbf{p}_2' = g(\mathbf{p}_1) + g(\mathbf{p}_2)$.

(2) We must show that $g(c\mathbf{p}) = cg(\mathbf{p})$, for all $c \in \mathbb{R}$ and $\mathbf{p} \in \mathcal{P}_n$. Now, $g(c\mathbf{p}) = \left(c\mathbf{p}\right)'$. Again, from calculus we know that the derivative of a constant times a function is equal to the constant times the derivative of the function, so $(c\mathbf{p})' = c(\mathbf{p}') = cg(\mathbf{p})$.

Hence, g is a linear transformation. ■

Example 4

Let $\mathcal{V}$ be a finite dimensional vector space, and let B be an ordered basis for $\mathcal{V}$. Then every element $\mathbf{v} \in \mathcal{V}$ has its coordinatization $[\mathbf{v}]_B$ with respect to B. Consider the mapping $f\colon \mathcal{V} \to \mathbb{R}^n$ given by $f(\mathbf{v}) = [\mathbf{v}]_B$. We will show that f is a linear transformation.

Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}$. By Theorem 4.20, $[\mathbf{v}_1 + \mathbf{v}_2]_B = [\mathbf{v}_1]_B + [\mathbf{v}_2]_B$. Hence,

$$f(\mathbf{v}_1 + \mathbf{v}_2) = [\mathbf{v}_1 + \mathbf{v}_2]_B = [\mathbf{v}_1]_B + [\mathbf{v}_2]_B = f(\mathbf{v}_1) + f(\mathbf{v}_2).$$

Next, let $c \in \mathbb{R}$ and $\mathbf{v} \in \mathcal{V}$. Again by Theorem 4.20, $[c\mathbf{v}]_B = c\,[\mathbf{v}]_B$. Hence,

$$f(c\mathbf{v}) = [c\mathbf{v}]_B = c\,[\mathbf{v}]_B = cf(\mathbf{v}).$$

Thus, f is a linear transformation from $\mathcal{V}$ to $\mathbb{R}^n$. ■

Not every function between vector spaces is a linear transformation. For example, consider the function $h: \mathbb{R}^2 \to \mathbb{R}^2$ given by $h([x,y]) = [x+1, y-2] = [x,y] + [1,-2]$. In this case, h merely adds $[1,-2]$ to each vector $[x,y]$ (see Figure 5.1). This type of mapping is called a **translation**. However, h is not a linear transformation. To show that it is not, we have to produce a counterexample to verify that either property (1) or property (2) of the definition fails. Property (1) fails, since $h([1,2]+[3,4]) = h([4,6]) = [5,4]$, while $h([1,2]) + h([3,4]) = [2,0] + [4,2] = [6,2]$.

In general, when given a function f between vector spaces, we do not always know right away whether f is a linear transformation. If we suspect that either property (1) or (2) does not hold for f, then we look for a counterexample.

Linear Operators and Some Geometric Examples

An important type of linear transformation is one that maps a vector space to itself.

Definition Let $\mathcal{V}$ be a vector space. A **linear operator** on $\mathcal{V}$ is a linear transformation whose domain and codomain are both $\mathcal{V}$.

Example 5

If $\mathcal{V}$ is any vector space, then the mapping $\boldsymbol{i}: \mathcal{V} \to \mathcal{V}$ given by $\boldsymbol{i}(\mathbf{v}) = \mathbf{v}$ for all $\mathbf{v} \in \mathcal{V}$ is a linear operator, known as the **identity linear operator**. Also, the constant mapping $\boldsymbol{z}: \mathcal{V} \to \mathcal{V}$ given by $\boldsymbol{z}(\mathbf{v}) = \mathbf{0}_{\mathcal{V}}$ is a linear operator known as the **zero linear operator** (see Exercise 2).

The next few examples exhibit important geometric operators. In these examples, assume that all vectors begin at the origin.

Example 6

Reflections: Consider the mapping $f: \mathbb{R}^3 \to \mathbb{R}^3$ given by $f([a_1,a_2,a_3]) = [a_1,a_2,-a_3]$. This mapping "reflects" the vector $[a_1,a_2,a_3]$ through the xy-plane, which acts like a "mirror" (see

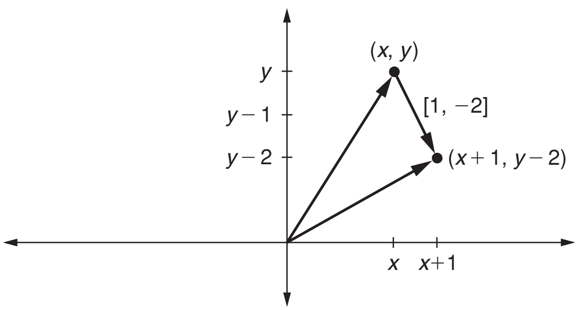


FIGURE 5.1

A translation in $\mathbb{R}^2$

Figure 5.2). Now, since

$$\begin{aligned} f([a_1,a_2,a_3]+[b_1,b_2,b_3]) &= f([a_1+b_1,a_2+b_2,a_3+b_3]) \\ &= [a_1+b_1,a_2+b_2,-(a_3+b_3)] \\ &= [a_1,a_2,-a_3]+[b_1,b_2,-b_3] \\ &= f([a_1,a_2,a_3])+f([b_1,b_2,b_3]), \quad \text{and} \end{aligned}$$

$$f(c[a_1,a_2,a_3]) = f([ca_1,ca_2,ca_3]) = [ca_1,ca_2,-ca_3] = c[a_1,a_2,-a_3] = cf([a_1,a_2,a_3]),$$

we see that f is a linear operator. Similarly, reflection through the xz-plane or the yz-plane is also a linear operator on $\mathbb{R}^3$ (see Exercise 4).

Example 7

Contractions and Dilations: Consider the mapping $g\colon \mathbb{R}^n \to \mathbb{R}^n$ given by scalar multiplication by k, where $k \in \mathbb{R}$; that is, $g(\mathbf{v}) = k\mathbf{v}$, for $\mathbf{v} \in \mathbb{R}^n$. The function g is a linear operator (see Exercise 3). If $|k| > 1$, g represents a **dilation** (lengthening) of the vectors in $\mathbb{R}^n$; if $|k| < 1$, g represents a **contraction** (shrinking).

Example 8

Projections: Consider the mapping $h\colon \mathbb{R}^3 \to \mathbb{R}^3$ given by $h([a_1,a_2,a_3]) = [a_1,a_2,0]$. This mapping takes each vector in $\mathbb{R}^3$ to a corresponding vector in the xy-plane (see Figure 5.3). Similarly,

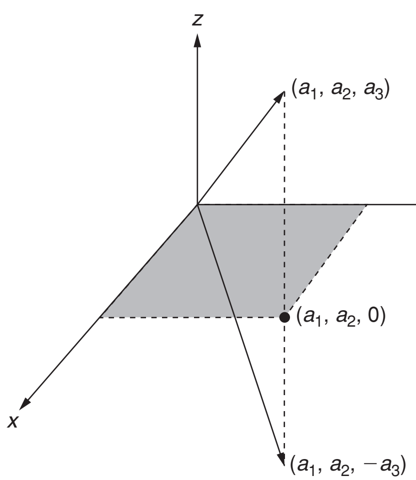


FIGURE 5.2

Reflection in $\mathbb{R}^3$ through the xy-plane

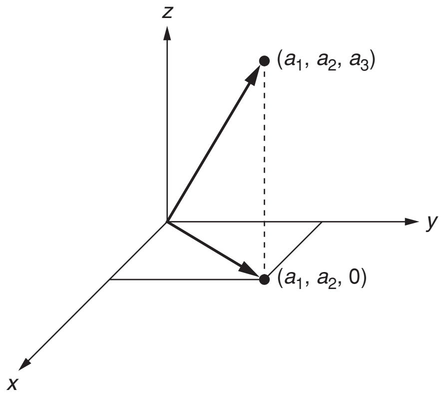


FIGURE 5.3

Projection of $[a_1, a_2, a_3]$ to the xy-plane

consider the mapping $j\colon \mathbb{R}^4 \to \mathbb{R}^4$ given by $j([a_1, a_2, a_3, a_4]) = [0, a_2, 0, a_4]$. This mapping takes each vector in $\mathbb{R}^4$ to a corresponding vector whose first and third coordinates are zero. The functions h and j are both linear operators (see Exercise 5). Such mappings, where at least one of the coordinates is "zeroed out," are examples of **projection mappings**. You can verify that all such mappings are linear operators. (Other types of projection mappings are illustrated in Exercises 6 and 7.)

Example 9

Rotations: Let θ be a fixed angle in $\mathbb{R}^2$, and let $l\colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$l\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{bmatrix}.$$

In Exercise 9 you are asked to show that l rotates $[x, y]$ counterclockwise through the angle θ (see Figure 5.4).

Now, let $\mathbf{v}_1 = [x_1, y_1]$ and $\mathbf{v}_2 = [x_2, y_2]$ be two vectors in $\mathbb{R}^2$. Then,

$$\begin{aligned} l(\mathbf{v}_1 + \mathbf{v}_2) &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}(\mathbf{v}_1 + \mathbf{v}_2) \\ &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\mathbf{v}_1 + \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\mathbf{v}_2 \\ &= l(\mathbf{v}_1) + l(\mathbf{v}_2). \end{aligned}$$

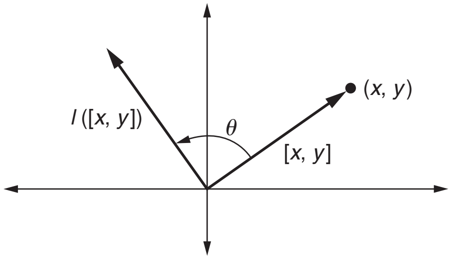


FIGURE 5.4

Counterclockwise rotation of $[x, y]$ through an angle θ in $\mathbb{R}^2$

Similarly, $l(c\mathbf{v}) = cl(\mathbf{v})$, for any $c \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^2$. Hence, l is a linear operator. ■

Beware! Not all geometric operations are linear operators. Recall that the translation function is not a linear operator!

Multiplication Transformation

The linear operator in Example 9 is actually a special case of the next example, which shows that multiplication by an $m \times n$ matrix is always a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$.

Example 10

Let $\mathbf{A}$ be a given $m \times n$ matrix. We show that the function $f: \mathbb{R}^n \to \mathbb{R}^m$ defined by $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$, for all $\mathbf{x} \in \mathbb{R}^n$, is a linear transformation. Let $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$. Then $f(\mathbf{x}_1 + \mathbf{x}_2) = \mathbf{A}(\mathbf{x}_1 + \mathbf{x}_2) = \mathbf{A}\mathbf{x}_1 + \mathbf{A}\mathbf{x}_2 = f(\mathbf{x}_1) + f(\mathbf{x}_2)$. Also, let $\mathbf{x} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. Then, $f(c\mathbf{x}) = \mathbf{A}(c\mathbf{x}) = c(\mathbf{A}\mathbf{x}) = cf(\mathbf{x})$. ■

For a specific example of the multiplication transformation, consider the matrix $\mathbf{A} = \begin{bmatrix} -1 & 4 & 2 \\ 5 & 6 & -3 \end{bmatrix}$. The mapping given by

$$f\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} -1 & 4 & 2 \\ 5 & 6 & -3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -x_1 + 4x_2 + 2x_3 \\ 5x_1 + 6x_2 - 3x_3 \end{bmatrix}$$

is a linear transformation from $\mathbb{R}^3$ to $\mathbb{R}^2$. In the next section, we will show that the converse of the result in Example 10 also holds; every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is equivalent to multiplication by an appropriate $m \times n$ matrix.

Elementary Properties of Linear Transformations

We now prove some basic properties of linear transformations. From here on, we usually use italicized capital letters, such as "L," to represent linear transformations.

Theorem 5.1 Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces, and let $L: \mathcal{V} \to \mathcal{W}$ be a linear transformation. Let $\mathbf{0}_{\mathcal{V}}$ be the zero vector in $\mathcal{V}$ and $\mathbf{0}_{\mathcal{W}}$ be the zero vector in $\mathcal{W}$. Then

(1) $L(\mathbf{0}_{\mathcal{V}}) = \mathbf{0}_{\mathcal{W}}$

(2) $L(-\mathbf{v}) = -L(\mathbf{v})$, for all $\mathbf{v} \in \mathcal{V}$

(3) $L(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n) = a_1L(\mathbf{v}_1) + a_2L(\mathbf{v}_2) + \cdots + a_nL(\mathbf{v}_n)$, for all $a_1, \ldots, a_n \in \mathbb{R}$, and $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathcal{V}$, for $n \geq 2$.

Proof.

Part (1):

$$\begin{aligned} L(\mathbf{0}_{\mathcal{V}}) &= L(0\mathbf{0}_{\mathcal{V}}) && \text{part (2) of Theorem 4.1, in } \mathcal{V} \\ &= 0L(\mathbf{0}_{\mathcal{V}}) && \text{property (2) of linear transformation} \\ &= \mathbf{0}_{\mathcal{W}} && \text{part (2) of Theorem 4.1, in } \mathcal{W} \end{aligned}$$

Part (2):

$$\begin{aligned} L(-\mathbf{v}) &= L(-1\mathbf{v}) && \text{part (3) of Theorem 4.1, in } \mathcal{V} \\ &= -1(L(\mathbf{v})) && \text{property (2) of linear transformation} \\ &= -L(\mathbf{v}) && \text{part (3) of Theorem 4.1, in } \mathcal{W} \end{aligned}$$

Part (3): (Abridged) This part is proved by induction. We prove the Base Step ($n = 2$) here and leave the Inductive Step as Exercise 29. For the Base Step, we must show that $L(a_1\mathbf{v}_1 + a_2\mathbf{v}_2) = a_1L(\mathbf{v}_1) + a_2L(\mathbf{v}_2)$. But,

$$\begin{aligned} L(a_1\mathbf{v}_1 + a_2\mathbf{v}_2) &= L(a_1\mathbf{v}_1) + L(a_2\mathbf{v}_2) && \text{property (1) of linear transformation} \\ &= a_1L(\mathbf{v}_1) + a_2L(\mathbf{v}_2) && \text{property (2) of linear transformation.} \end{aligned}$$ □

The next theorem asserts that the composition $L_2 \circ L_1$ of linear transformations L_1 and L_2 is again a linear transformation (see Appendix B for a review of composition of functions).

Theorem 5.2 Let $\mathcal{V}_1, \mathcal{V}_2$, and $\mathcal{V}_3$ be vector spaces. Let $L_1: \mathcal{V}_1 \to \mathcal{V}_2$ and $L_2: \mathcal{V}_2 \to \mathcal{V}_3$ be linear transformations. Then $L_2 \circ L_1: \mathcal{V}_1 \to \mathcal{V}_3$ given by $(L_2 \circ L_1)(\mathbf{v}) = L_2(L_1(\mathbf{v}))$, for all $\mathbf{v} \in \mathcal{V}_1$, is a linear transformation.

Proof. (Abridged) To show that $L_2 \circ L_1$ is a linear transformation, we must show that for all $c \in \mathbb{R}$ and $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}$,

$$\begin{aligned} (L_2 \circ L_1)(\mathbf{v}_1 + \mathbf{v}_2) &= (L_2 \circ L_1)(\mathbf{v}_1) + (L_2 \circ L_1)(\mathbf{v}_2) \\ \text{and} \quad (L_2 \circ L_1)(c\mathbf{v}) &= c(L_2 \circ L_1)(\mathbf{v}). \end{aligned}$$

The first property holds since

$$\begin{aligned} (L_2 \circ L_1)(\mathbf{v}_1 + \mathbf{v}_2) &= L_2(L_1(\mathbf{v}_1 + \mathbf{v}_2)) \\ &= L_2(L_1(\mathbf{v}_1) + L_1(\mathbf{v}_2)) && \text{because } L_1 \text{ is a linear transformation} \\ &= L_2(L_1(\mathbf{v}_1)) + L_2(L_1(\mathbf{v}_2)) && \text{because } L_2 \text{ is a linear transformation} \\ &= (L_2 \circ L_1)(\mathbf{v}_1) + (L_2 \circ L_1)(\mathbf{v}_2). \end{aligned}$$

We leave the proof of the second property as Exercise 33. □

Example 11

Let L_1 represent the rotation of vectors in $\mathbb{R}^2$ through a fixed angle θ (as in Example 9), and let L_2 represent the reflection of vectors in $\mathbb{R}^2$ through the x-axis. That is, if $\mathbf{v} = [v_1, v_2]$, then

$$L_1(\mathbf{v}) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad \text{and} \quad L_2(\mathbf{v}) = \begin{bmatrix} v_1 \\ -v_2 \end{bmatrix}.$$

Because L_1 and L_2 are both linear transformations, Theorem 5.2 asserts that

$$L_2(L_1(\mathbf{v})) = L_2\left(\begin{bmatrix} v_1\cos\theta - v_2\sin\theta \\ v_1\sin\theta + v_2\cos\theta \end{bmatrix}\right) = \begin{bmatrix} v_1\cos\theta - v_2\sin\theta \\ -v_1\sin\theta - v_2\cos\theta \end{bmatrix}$$

is also a linear transformation. $L_2 \circ L_1$ represents a rotation of $\mathbf{v}$ through θ followed by a reflection through the x-axis.

Theorem 5.2 generalizes naturally to more than two linear transformations. That is, if $L_1, L_2, \ldots, L_k$ are linear transformations and the composition $L_k \circ \cdots \circ L_2 \circ L_1$ makes sense, then $L_k \circ \cdots \circ L_2 \circ L_1$ is also a linear transformation.

Linear Transformations and Subspaces

The final theorem of this section assures us that, under a linear transformation $L: \mathcal{V} \to \mathcal{W}$, subspaces of $\mathcal{V}$ "correspond" to subspaces of $\mathcal{W}$, and vice versa.

Theorem 5.3 Let $L: \mathcal{V} \to \mathcal{W}$ be a linear transformation.

(1) If $\mathcal{V}'$ is a subspace of $\mathcal{V}$, then $L(\mathcal{V}') = \{L(\mathbf{v}) \mid \mathbf{v} \in \mathcal{V}'\}$, the image of $\mathcal{V}'$ in $\mathcal{W}$, is a subspace of $\mathcal{W}$. In particular, the range of L is a subspace of $\mathcal{W}$.

(2) If $\mathcal{W}'$ is a subspace of $\mathcal{W}$, then $L^{-1}(\mathcal{W}') = \{\mathbf{v} \mid L(\mathbf{v}) \in \mathcal{W}'\}$, the pre-image of $\mathcal{W}'$ in $\mathcal{V}$, is a subspace of $\mathcal{V}$.

We prove part (1) and leave part (2) as Exercise 31.

Proof. Part (1): Suppose that $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation and that $\mathcal{V}'$ is a subspace of $\mathcal{V}$. Now, $L(\mathcal{V}')$, the image of $\mathcal{V}'$ in $\mathcal{W}$ (see Figure 5.5), is certainly nonempty (why?). Hence, to show that $L(\mathcal{V}')$ is a subspace of $\mathcal{W}$, we must prove that $L(\mathcal{V}')$ is closed under addition and scalar multiplication.

First, suppose that $\mathbf{w}_1, \mathbf{w}_2 \in L(\mathcal{V}')$. Then, by definition of $L(\mathcal{V}')$, we have $\mathbf{w}_1 = L(\mathbf{v}_1)$ and $\mathbf{w}_2 = L(\mathbf{v}_2)$, for some $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}'$. Then, $\mathbf{w}_1 + \mathbf{w}_2 = L(\mathbf{v}_1) + L(\mathbf{v}_2) = L(\mathbf{v}_1 + \mathbf{v}_2)$ because L is a linear transformation. However, since $\mathcal{V}'$ is a subspace of $\mathcal{V}$, $(\mathbf{v}_1 + \mathbf{v}_2) \in \mathcal{V}'$. Thus, $(\mathbf{w}_1 + \mathbf{w}_2)$ is the image of $(\mathbf{v}_1 + \mathbf{v}_2) \in \mathcal{V}'$, and so $(\mathbf{w}_1 + \mathbf{w}_2) \in L(\mathcal{V}')$. Hence, $L(\mathcal{V}')$ is closed under addition.

Next, suppose that $c \in \mathbb{R}$ and $\mathbf{w} \in L(\mathcal{V}')$. By definition of $L(\mathcal{V}')$, $\mathbf{w} = L(\mathbf{v})$, for some $\mathbf{v} \in \mathcal{V}'$. Then, $c\mathbf{w} = cL(\mathbf{v}) = L(c\mathbf{v})$ since L is a linear transformation. Now, $c\mathbf{v} \in \mathcal{V}'$, because $\mathcal{V}'$ is a subspace of $\mathcal{V}$. Thus, $c\mathbf{w}$ is the image of $c\mathbf{v} \in \mathcal{V}'$, and so $c\mathbf{w} \in L(\mathcal{V}')$. Hence, $L(\mathcal{V}')$ is closed under scalar multiplication. $\square$

Example 12

Let $L: \mathcal{M}_{22} \to \mathbb{R}^3$, where $L\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = [b, 0, c]$. L is a linear transformation (verify!). By Theorem 5.3, the range of any linear transformation is a subspace of the codomain. Hence, the range of $L = \{[b, 0, c] \mid b, c \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^3$.

Also, consider the subspace $\mathcal{U}_2 = \left\{ \begin{bmatrix} a & b \\ 0 & d \end{bmatrix} \middle| a, b, d \in \mathbb{R} \right\}$ of $\mathcal{M}_{22}$. Then the image of $\mathcal{U}_2$ under L is $\{[b, 0, 0] \mid b \in \mathbb{R}\}$. This image is a subspace of $\mathbb{R}^3$, as Theorem 5.3 asserts. Finally, consider the subspace $\mathcal{W} = \{[b, e, 2b] \mid b, e \in \mathbb{R}\}$ of $\mathbb{R}^3$. The pre-image of $\mathcal{W}$ consists of all

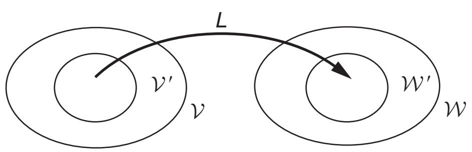


FIGURE 5.5

Subspaces of $\mathcal{V}$ correspond to subspaces of $\mathcal{W}$ under a linear transformation $L: \mathcal{V} \to \mathcal{W}$

matrices in $\mathcal{M}_{22}$ of the form $\begin{bmatrix} a & b \\ 2b & d \end{bmatrix}$. Notice that this pre-image is a subspace of $\mathcal{M}_{22}$, as claimed by Theorem 5.3. ■

New Vocabulary

codomain (of a linear transformation)
composition of linear transformations
contraction (mapping)
dilation (mapping)
domain (of a linear transformation)
identity linear operator
image (of a vector in the domain)
linear operator
linear transformation
pre-image (of a vector in the codomain)
projection (mapping)
range (of a linear transformation)
reflection (mapping)
rotation (mapping)
shear (mapping)
translation (mapping)
zero linear operator

Highlights

- A linear transformation is a function from one vector space to another that preserves the operations of addition and scalar multiplication. That is, under a linear transformation, the image of a linear combination of vectors is the linear combination of the images of the vectors having the same coefficients.
- A linear operator is a linear transformation from a vector space to itself.
- A nontrivial translation of the plane ($\mathbb{R}^2$) or of space ($\mathbb{R}^3$) is never a linear operator, but all of the following are linear operators: contraction (of $\mathbb{R}^n$), dilation (of $\mathbb{R}^n$), reflection of space through the xy-plane (or xz-plane or yz-plane), rotation of the plane about the origin through a given angle θ, projection (of $\mathbb{R}^n$) in which one or more of the coordinates are zeroed out.
- Multiplication of vectors in $\mathbb{R}^n$ on the left by a fixed $m \times n$ matrix $\mathbf{A}$ is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$.
- Multiplying a vector on the left by the matrix $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is equivalent to rotating the vector counterclockwise about the origin through the angle θ.
- Linear transformations always map the zero vector of the domain to the zero vector of the codomain.
- A composition of linear transformations is a linear transformation.
- Under a linear transformation, subspaces of the domain map to subspaces of the codomain, and the pre-image of a subspace of the codomain is a subspace of the domain.

EXERCISES FOR SECTION 5.1

1. Determine which of the following functions are linear transformations. Prove that your answers are correct. Which are linear operators?

   ★(a) $f: \mathbb{R}^2 \to \mathbb{R}^2$ given by $f([x,y]) = [3x - 4y, -x + 2y]$

   ★(b) $h: \mathbb{R}^4 \to \mathbb{R}^4$ given by $h([x_1,x_2,x_3,x_4]) = [x_1 + 2, x_2 - 1, x_3, -3]$

   (c) $k: \mathbb{R}^3 \to \mathbb{R}^3$ given by $k([x_1,x_2,x_3]) = [x_2, x_3, x_1]$

   ★(d) $l: \mathcal{M}_{22} \to \mathcal{M}_{22}$ given by $l\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} a - 2c + d & 3b - c \\ -4a & b + c - 3d \end{bmatrix}$

   (e) $n: \mathcal{M}_{22} \to \mathbb{R}$ given by $n\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc$

   ★(f) $r: \mathcal{P}_3 \to \mathcal{P}_2$ given by $r(ax^3 + bx^2 + cx + d) = (\sqrt[3]{a})x^2 - b^2x + c$

   (g) $s: \mathbb{R}^3 \to \mathbb{R}^3$ given by $s([x_1,x_2,x_3]) = [\cos x_1, \sin x_2, e^{x_3}]$

   ★(h) $t: \mathcal{P}_3 \to \mathbb{R}$ given by $t(a_3x^3 + a_2x^2 + a_1x + a_0) = a_3 + a_2 + a_1 + a_0$

   (i) $u: \mathbb{R}^4 \to \mathbb{R}$ given by $u\left([x_1,x_2,x_3,x_4]\right) = |x_2|$

   ★(j) $v: \mathcal{P}_2 \to \mathbb{R}$ given by $v\left(ax^2 + bx + c\right) = abc$

   ★(k) $g: \mathcal{M}_{32} \to \mathcal{P}_4$ given by $g\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}\right) = a_{11}x^4 - a_{21}x^2 + a_{31}$

   ★(l) $e: \mathbb{R}^2 \to \mathbb{R}$ given by $e([x,y]) = \sqrt{x^2 + y^2}$

2. Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces.

   (a) Show that the identity mapping $i: \mathcal{V} \to \mathcal{V}$ given by $i(\mathbf{v}) = \mathbf{v}$, for all $\mathbf{v} \in \mathcal{V}$, is a linear operator.

   (b) Show that the zero mapping $z: \mathcal{V} \to \mathcal{W}$ given by $z(\mathbf{v}) = \mathbf{0}_{\mathcal{W}}$, for all $\mathbf{v} \in \mathcal{V}$, is a linear transformation.

3. Let k be a fixed scalar in $\mathbb{R}$. Show that the mapping $f: \mathbb{R}^n \to \mathbb{R}^n$ given by $f([x_1,x_2,\ldots,x_n]) = k[x_1,x_2,\ldots,x_n]$ is a linear operator.

4. (a) Show that $f: \mathbb{R}^3 \to \mathbb{R}^3$ given by $f([x,y,z]) = [-x,y,z]$ (reflection of a vector through the yz-plane) is a linear operator.

   (b) What mapping from $\mathbb{R}^3$ to $\mathbb{R}^3$ would reflect a vector through the xz-plane? Is it a linear operator? Why or why not?

   (c) What mapping from $\mathbb{R}^2$ to $\mathbb{R}^2$ would reflect a vector through the y-axis? through the x-axis? Are these linear operators? Why or why not?

5. Show that the projection mappings $h: \mathbb{R}^3 \to \mathbb{R}^3$ given by $h([a_1,a_2,a_3]) = [a_1,a_2,0]$ and $j: \mathbb{R}^4 \to \mathbb{R}^4$ given by $j([a_1,a_2,a_3,a_4]) = [0,a_2,0,a_4]$ are linear operators.

6. The mapping $f:\mathbb{R}^n \to \mathbb{R}$ given by $f([x_1,x_2,\ldots,x_i,\ldots,x_n]) = x_i$ is another type of projection mapping. Show that f is a linear transformation.

7. Let $\mathbf{x}$ be a fixed nonzero vector in $\mathbb{R}^3$. Show that the mapping $g:\mathbb{R}^3 \to \mathbb{R}^3$ given by $g(\mathbf{y}) = \mathbf{proj_x y}$ is a linear operator.

8. Let $\mathbf{x}$ be a fixed vector in $\mathbb{R}^n$. Prove that $L: \mathbb{R}^n \to \mathbb{R}$ given by $L(\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$ is a linear transformation.

9. Let θ be a fixed angle in the xy-plane. Show that the linear operator $L:\mathbb{R}^2 \to \mathbb{R}^2$ given by $L\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}$ rotates the vector $[x,y]$ counterclockwise through the angle θ in the plane. (Hint: Consider the vector $[x',y']$, obtained by rotating $[x,y]$ counterclockwise through the angle θ. Let $r = \sqrt{x^2+y^2}$. Then $x = r\cos\alpha$ and $y = r\sin\alpha$, where α is the angle shown in Figure 5.6. Notice that $x' = r(\cos(\theta+\alpha))$ and $y' = r(\sin(\theta+\alpha))$. Then show that $L([x,y]) = [x',y']$.)

10. **(a)** Explain why the mapping $L: \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$L\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

is a linear operator.

(b) Show that the mapping L in part (a) rotates every vector in $\mathbb{R}^3$ about the z-axis through an angle of θ (as measured relative to the xy-plane).

$\star$**(c)** What matrix should be multiplied times $[x,y,z]$ to create the linear operator that rotates $\mathbb{R}^3$ about the y-axis through an angle θ (relative to the xz-plane)? (Hint: When looking down from the positive y-axis toward

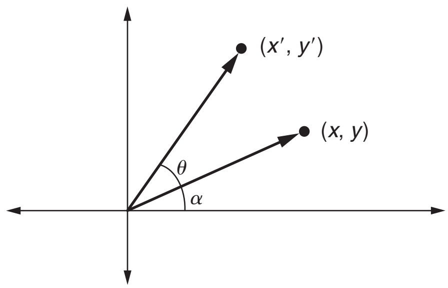


FIGURE 5.6

The vectors $[x,y]$ and $[x',y']$

the xz-plane in a right-handed system, the positive z-axis rotates 90° counterclockwise into the positive x-axis.)

11. **Shears**: Let $f_1, f_2 \colon \mathbb{R}^2 \to \mathbb{R}^2$ be given by

$$f_1\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x + ky \\ y \end{bmatrix}$$

and

$$f_2\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ kx + y \end{bmatrix}.$$

The mapping f_1 is called a **shear in the x-direction with factor k**; f_2 is called a **shear in the y-direction with factor k**. The effect of these functions (for $k > 1$) on the vector $[1, 1]$ is shown in Figure 5.7. Show that f_1 and f_2 are linear operators directly, without using Example 10.

12. Let $f \colon \mathcal{M}_{nn} \to \mathbb{R}$ be given by $f(\mathbf{A}) = \operatorname{trace}(\mathbf{A})$. (The trace is defined in Exercise 14 of Section 1.4.) Prove that f is a linear transformation.

13. Show that the mappings $g, h \colon \mathcal{M}_{nn} \to \mathcal{M}_{nn}$ given by $g(\mathbf{A}) = \mathbf{A} + \mathbf{A}^T$ and $h(\mathbf{A}) = \mathbf{A} - \mathbf{A}^T$ are linear operators on $\mathcal{M}_{nn}$.

14. **(a)** Show that if $\mathbf{p} \in \mathcal{P}_n$, then the (indefinite integral) function $f \colon \mathcal{P}_n \to \mathcal{P}_{n+1}$, where $f(\mathbf{p})$ is the vector $\int \mathbf{p}(x)\,dx$ with zero constant term, is a linear transformation.

(b) Show that if $\mathbf{p} \in \mathcal{P}_n$, then the (definite integral) function $g \colon \mathcal{P}_n \to \mathbb{R}$ given by $g(\mathbf{p}) = \int_a^b \mathbf{p}\,dx$ is a linear transformation, for any fixed $a, b \in \mathbb{R}$.

15. Let $\mathcal{V}$ be the vector space of all functions f from $\mathbb{R}$ to $\mathbb{R}$ that are infinitely differentiable (that is, for which $f^{(n)}$, the nth derivative of f, exists for every

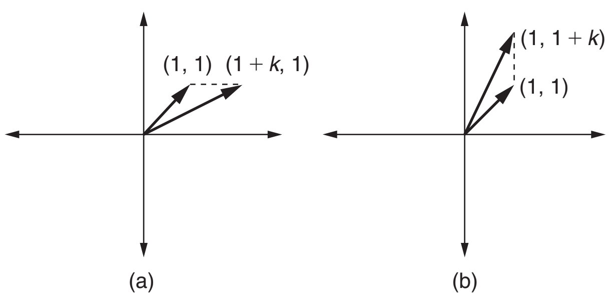


FIGURE 5.7

(a) Shear in the x-direction; (b) shear in the y-direction (both for $k > 0$)

integer $n \geq 1$). Use induction and Theorem 5.2 to show that for any given integer $k \geq 1$, $L: \mathcal{V} \to \mathcal{V}$ given by $L(f) = f^{(k)}$ is a linear operator.

16. Consider the function $f: \mathcal{M}_{nn} \to \mathcal{M}_{nn}$ given by $f(\mathbf{A}) = \mathbf{BA}$, where $\mathbf{B}$ is some fixed $n \times n$ matrix. Show that f is a linear operator.

17. Let $\mathbf{B}$ be a fixed nonsingular matrix in $\mathcal{M}_{nn}$. Show that the mapping $f: \mathcal{M}_{nn} \to \mathcal{M}_{nn}$ given by $f(\mathbf{A}) = \mathbf{B}^{-1}\mathbf{AB}$ is a linear operator.

18. Let a be a fixed real number.

(a) Let $L: \mathcal{P}_n \to \mathbb{R}$ be given by $L(\mathbf{p}(x)) = \mathbf{p}(a)$. (That is, L evaluates polynomials in $\mathcal{P}_n$ at $x = a$.) Show that L is a linear transformation.

(b) Let $L: \mathcal{P}_n \to \mathcal{P}_n$ be given by $L(\mathbf{p}(x)) = \mathbf{p}(x + a)$. (For example, when a is positive, L shifts the graph of $\mathbf{p}(x)$ to the *left* by a units.) Prove that L is a linear operator.

19. Let $\mathbf{A}$ be a fixed matrix in $\mathcal{M}_{nn}$. Define $f: \mathcal{P}_n \to \mathcal{M}_{nn}$ by

$$\begin{aligned} &f(a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0) \\ &\quad = a_n \mathbf{A}^n + a_{n-1}\mathbf{A}^{n-1} + \cdots + a_1 \mathbf{A} + a_0 \mathbf{I}_n. \end{aligned}$$

Show that f is a linear transformation.

20. Let $\mathcal{V}$ be the unusual vector space from Example 7 in Section 4.1. Show that $L: \mathcal{V} \to \mathbb{R}$ given by $L(x) = \ln(x)$ is a linear transformation.

21. Let $\mathcal{V}$ be a vector space, and let $\mathbf{x} \neq \mathbf{0}$ be a fixed vector in $\mathcal{V}$. Prove that the translation function $f: \mathcal{V} \to \mathcal{V}$ given by $f(\mathbf{v}) = \mathbf{v} + \mathbf{x}$ is not a linear transformation.

22. Show that if $\mathbf{A}$ is a fixed matrix in $\mathcal{M}_{mn}$ and $\mathbf{y} \neq \mathbf{0}$ is a fixed vector in $\mathbb{R}^m$, then the mapping $f: \mathbb{R}^n \to \mathbb{R}^m$ given by $f(\mathbf{x}) = \mathbf{Ax} + \mathbf{y}$ is not a linear transformation by showing that part (1) of Theorem 5.1 fails for f.

23. Prove that $f: \mathcal{M}_{33} \to \mathbb{R}$ given by $f(\mathbf{A}) = |\mathbf{A}|$ is not a linear transformation. (A similar result is true for $\mathcal{M}_{nn}$, for $n > 1$.)

24. Suppose $L_1: \mathcal{V} \to \mathcal{W}$ is a linear transformation and $L_2: \mathcal{V} \to \mathcal{W}$ is defined by $L_2(\mathbf{v}) = L_1(2\mathbf{v})$. Show that L_2 is a linear transformation.

25. Suppose $L: \mathbb{R}^3 \to \mathbb{R}^3$ is a linear operator and $L([1,0,0]) = [-2,1,0]$, $L([0,1,0]) = [3,-2,1]$, and $L([0,0,1]) = [0,-1,3]$. Find $L([-3,2,4])$. Give a formula for $L([x,y,z])$, for any $[x,y,z] \in \mathbb{R}^3$.

★26. Suppose $L: \mathbb{R}^2 \to \mathbb{R}^2$ is a linear operator and $L(\mathbf{i} + \mathbf{j}) = \mathbf{i} - 3\mathbf{j}$ and $L(-2\mathbf{i} + 3\mathbf{j}) = -4\mathbf{i} + 2\mathbf{j}$. Express $L(\mathbf{i})$ and $L(\mathbf{j})$ as linear combinations of $\mathbf{i}$ and $\mathbf{j}$.

27. Let $L: \mathcal{V} \to \mathcal{W}$ be a linear transformation. Show that $L(\mathbf{x} - \mathbf{y}) = L(\mathbf{x}) - L(\mathbf{y})$, for all vectors $\mathbf{x}, \mathbf{y} \in \mathcal{V}$.

28. Part (3) of Theorem 5.1 assures us that if $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation, then $L(a\mathbf{v}_1 + b\mathbf{v}_2) = aL(\mathbf{v}_1) + bL(\mathbf{v}_2)$, for all $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}$ and all $a, b \in \mathbb{R}$. Prove that the converse of this statement is true. (Hint: Consider two cases: first $a = b = 1$ and then $b = 0$.)

▶**29.** Finish the proof of part (3) of Theorem 5.1 by doing the Inductive Step.

30. **(a)** Suppose that $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation. Show that if $\{L(\mathbf{v}_1), L(\mathbf{v}_2), \ldots, L(\mathbf{v}_n)\}$ is a linearly independent set of n distinct vectors in $\mathcal{W}$, for some vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathcal{V}$, then $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a linearly independent set in $\mathcal{V}$.

⋆**(b)** Find a counterexample to the converse of part (a).

▶**31.** Finish the proof of Theorem 5.3 by showing that if $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation and $\mathcal{W}'$ is a subspace of $\mathcal{W}$ with pre-image $L^{-1}(\mathcal{W}')$, then $L^{-1}(\mathcal{W}')$ is a subspace of $\mathcal{V}$.

32. Show that every linear operator $L: \mathbb{R} \to \mathbb{R}$ has the form $L(\mathbf{x}) = c\mathbf{x}$, for some $c \in \mathbb{R}$.

33. Finish the proof of Theorem 5.2 by proving property (2) of a linear transformation for $L_2 \circ L_1$.

34. Let $L_1, L_2: \mathcal{V} \to \mathcal{W}$ be linear transformations. Define $(L_1 \oplus L_2): \mathcal{V} \to \mathcal{W}$ by $(L_1 \oplus L_2)(\mathbf{v}) = L_1(\mathbf{v}) + L_2(\mathbf{v})$ (where the latter addition takes place in $\mathcal{W}$). Also define $(c \odot L_1): \mathcal{V} \to \mathcal{W}$ by $(c \odot L_1)(\mathbf{v}) = c\,(L_1(\mathbf{v}))$ (where the latter scalar multiplication takes place in $\mathcal{W}$).

(a) Show that $(L_1 \oplus L_2)$ and $(c \odot L_1)$ are linear transformations.

(b) Use the results in part (a) above and part (b) of Exercise 2 to show that the set of all linear transformations from $\mathcal{V}$ to $\mathcal{W}$ is a vector space under the operations $\oplus$ and $\odot$.

35. Let $L: \mathbb{R}^2 \to \mathbb{R}^2$ be a nonzero linear operator. Show that L maps a line to either a line or a point.

⋆**36.** True or False:

(a) If $L: \mathcal{V} \to \mathcal{W}$ is a function between vector spaces for which $L(c\mathbf{v}) = cL(\mathbf{v})$, then L is a linear transformation.

(b) If $\mathcal{V}$ is an n-dimensional vector space with ordered basis B, then $L: \mathcal{V} \to \mathbb{R}^n$ given by $L(\mathbf{v}) = [\mathbf{v}]_B$ is a linear transformation.

(c) The function $L: \mathbb{R}^3 \to \mathbb{R}^3$ given by $L([x, y, z]) = [x+1, y-2, z+3]$ is a linear operator.

(d) If $\mathbf{A}$ is a 4×3 matrix, then $L(\mathbf{v}) = \mathbf{A}\mathbf{v}$ is a linear transformation from $\mathbb{R}^4$ to $\mathbb{R}^3$.

(e) A linear transformation from $\mathcal{V}$ to $\mathcal{W}$ always maps $\mathbf{0}_{\mathcal{V}}$ to $\mathbf{0}_{\mathcal{W}}$.

(f) If $M_1: \mathcal{V} \to \mathcal{W}$ and $M_2: \mathcal{W} \to \mathcal{X}$ are linear transformations, then $M_1 \circ M_2$ is a well-defined linear transformation.

(g) If $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation, then the image of any subspace of $\mathcal{V}$ is a subspace of $\mathcal{W}$.

(h) If $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation, then the pre-image of $\{\mathbf{0}_{\mathcal{W}}\}$ is a subspace of $\mathcal{V}$.

5.2 THE MATRIX OF A LINEAR TRANSFORMATION

In this section, we show that the behavior of any linear transformation $L: \mathcal{V} \to \mathcal{W}$ is determined by its effect on a basis for $\mathcal{V}$. In particular, when $\mathcal{V}$ and $\mathcal{W}$ are finite dimensional and ordered bases for $\mathcal{V}$ and $\mathcal{W}$ are chosen, we can obtain a matrix corresponding to L that is useful in computing images under L. Finally, we investigate how the matrix for L changes as the bases for $\mathcal{V}$ and $\mathcal{W}$ change.

A Linear Transformation Is Determined by Its Action on a Basis

If the action of a linear transformation $L: \mathcal{V} \to \mathcal{W}$ on a basis for $\mathcal{V}$ is known, then the action of L can be computed for all elements of $\mathcal{V}$, as we see in the next example.

Example 1

You can quickly verify that

$$B = \left([0,4,0,1], [-2,5,0,2], [-3,5,1,1], [-1,2,0,1]\right)$$

is an ordered basis for $\mathbb{R}^4$. Now suppose that $L: \mathbb{R}^4 \to \mathbb{R}^3$ is a linear transformation for which

$$L\left([0,4,0,1]\right) = [3,1,2], \qquad L([-2,5,0,2]) = [2,-1,1],$$

$$L([-3,5,1,1]) = [-4,3,0], \quad \text{and} \quad L([-1,2,0,1]) = [6,1,-1].$$

We can use the values of L on B to compute L for other vectors in $\mathbb{R}^4$. For example, let $\mathbf{v} = [-4,14,1,4]$. By using row reduction, we see that $[\mathbf{v}]_B = [2,-1,1,3]$ (verify!). So,

$$\begin{aligned} L(\mathbf{v}) &= L\left(2[0,4,0,1] - 1[-2,5,0,2] + 1[-3,5,1,1] + 3[-1,2,0,1]\right) \\ &= 2L\left([0,4,0,1]\right) - 1L([-2,5,0,2]) + 1L([-3,5,1,1]) \\ &\quad + 3L([-1,2,0,1]) \\ &= 2[3,1,2] - [2,-1,1] + [-4,3,0] + 3[6,1,-1] \\ &= [18,9,0]. \end{aligned}$$

In general, if $\mathbf{v} \in \mathbb{R}^4$ and $[\mathbf{v}]_B = [k_1, k_2, k_3, k_4]$, then

$$\begin{aligned} L(\mathbf{v}) &= k_1[3,1,2] + k_2[2,-1,1] + k_3[-4,3,0] + k_4[6,1,-1] \\ &= [3k_1 + 2k_2 - 4k_3 + 6k_4,\ k_1 - k_2 + 3k_3 + k_4,\ 2k_1 + k_2 - k_4]. \end{aligned}$$

Thus, we have derived a general formula for L from its effect on the basis B. ■

Example 1 illustrates the next theorem.

Theorem 5.4 Let $B = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ be an ordered basis for a vector space $\mathcal{V}$. Let $\mathcal{W}$ be a vector space, and let $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$ be any n vectors in $\mathcal{W}$. Then there is a unique linear transformation $L\colon \mathcal{V} \to \mathcal{W}$ such that $L(\mathbf{v}_1) = \mathbf{w}_1$, $L(\mathbf{v}_2) = \mathbf{w}_2, \ldots, L(\mathbf{v}_n) = \mathbf{w}_n$.

Proof. (Abridged) Let $B = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ be an ordered basis for $\mathcal{V}$, and let $\mathbf{v} \in \mathcal{V}$. Then $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$, for some unique a_i's in $\mathbb{R}$. Let $\mathbf{w}_1, \ldots, \mathbf{w}_n$ be any vectors in $\mathcal{W}$. Define $L\colon \mathcal{V} \to \mathcal{W}$ by $L(\mathbf{v}) = a_1\mathbf{w}_1 + a_2\mathbf{w}_2 + \cdots + a_n\mathbf{w}_n$. Notice that $L(\mathbf{v})$ is well defined since the a_i's are unique.

To show that L is a linear transformation, we must prove that $L(\mathbf{x}_1 + \mathbf{x}_2) = L(\mathbf{x}_1) + L(\mathbf{x}_2)$ and $L(c\mathbf{x}_1) = cL(\mathbf{x}_1)$, for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{V}$ and all $c \in \mathbb{R}$. Suppose that $\mathbf{x}_1 = d_1\mathbf{v}_1 + \cdots + d_n\mathbf{v}_n$ and $\mathbf{x}_2 = e_1\mathbf{v}_1 + \cdots + e_n\mathbf{v}_n$. Then, by definition of L, $L(\mathbf{x}_1) = d_1\mathbf{w}_1 + \cdots + d_n\mathbf{w}_n$ and $L(\mathbf{x}_2) = e_1\mathbf{w}_n + \cdots + e_n\mathbf{w}_n$. However,

$$\begin{aligned} \mathbf{x}_1 + \mathbf{x}_2 &= (d_1 + e_1)\mathbf{v}_1 + \cdots + (d_n + e_n)\mathbf{v}_n, \\ \text{so,}\quad L(\mathbf{x}_1 + \mathbf{x}_2) &= (d_1 + e_1)\mathbf{w}_1 + \cdots + (d_n + e_n)\mathbf{w}_n, \end{aligned}$$

again by definition of L. Hence, $L(\mathbf{x}_1) + L(\mathbf{x}_2) = L(\mathbf{x}_1 + \mathbf{x}_2)$.

Similarly, suppose $\mathbf{x} \in \mathcal{V}$, and $\mathbf{x} = t_1\mathbf{v}_1 + \cdots + t_n\mathbf{v}_n$. Then, $c\mathbf{x} = ct_1\mathbf{v}_1 + \cdots + ct_n\mathbf{v}_n$, and so $L(c\mathbf{x}) = ct_1\mathbf{w}_1 + \cdots + ct_n\mathbf{w}_n = cL(\mathbf{x})$. Hence, L is a linear transformation.

Finally, the proof of the uniqueness assertion is straightforward and is left as Exercise 25. □

The Matrix of a Linear Transformation

Our next goal is to show that every linear transformation on a finite dimensional vector space can be expressed as a matrix multiplication. This will allow us to solve problems involving linear transformations by performing matrix multiplications, which can easily be done by computer. As we will see, the matrix for a linear transformation is determined by the ordered bases B and C chosen for the domain and codomain, respectively. Our goal is to find a matrix that takes the B-coordinates of a vector in the domain to the C-coordinates of its image vector in the codomain.

Recall the linear transformation $L\colon \mathbb{R}^4 \to \mathbb{R}^3$ with the ordered basis B for $\mathbb{R}^4$ from Example 1. For $\mathbf{v} \in \mathbb{R}^4$, we let $[\mathbf{v}]_B = [k_1, k_2, k_3, k_4]$, and obtained the following formula for L:

$$L(\mathbf{v}) = [3k_1 + 2k_2 - 4k_3 + 6k_4,\ k_1 - k_2 + 3k_3 + k_4,\ 2k_1 + k_2 - k_4].$$

Now, to keep matters simple, we select the standard basis $C = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ for the codomain $\mathbb{R}^3$, so that the C-coordinates of vectors in the codomain are the same as the vectors themselves. (That is, $L(\mathbf{v}) = [L(\mathbf{v})]_C$, since C is the standard basis.) Then this formula for L takes the B-coordinates of each vector in the domain to the C-coordinates of its image vector in the codomain. Now, notice that if

$$\mathbf{A}_{BC} = \begin{bmatrix} 3 & 2 & -4 & 6 \\ 1 & -1 & 3 & 1 \\ 2 & 1 & 0 & -1 \end{bmatrix}, \text{ then } \mathbf{A}_{BC} \begin{bmatrix} k_1 \\ k_2 \\ k_3 \\ k_4 \end{bmatrix} = \begin{bmatrix} 3k_1 + 2k_2 - 4k_3 + 6k_4 \\ k_1 - k_2 + 3k_3 + k_4 \\ 2k_1 + k_2 - k_4 \end{bmatrix}.$$

Hence, the matrix $\mathbf{A}$ contains all of the information needed for carrying out the linear transformation L with respect to the chosen bases B and C.

A similar process can be used for any linear transformation between finite dimensional vector spaces.

Theorem 5.5 Let $\mathcal{V}$ and $\mathcal{W}$ be nontrivial vector spaces, with $\dim(\mathcal{V}) = n$ and $\dim(\mathcal{W}) = m$. Let $B = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ and $C = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m)$ be ordered bases for $\mathcal{V}$ and $\mathcal{W}$, respectively. Let $L: \mathcal{V} \to \mathcal{W}$ be a linear transformation. Then there is a unique $m \times n$ matrix $\mathbf{A}_{BC}$ such that $\mathbf{A}_{BC}[\mathbf{v}]_B = [L(\mathbf{v})]_C$, for all $\mathbf{v} \in \mathcal{V}$. (That is, $\mathbf{A}_{BC}$ times the coordinatization of $\mathbf{v}$ with respect to B gives the coordinatization of $L(\mathbf{v})$ with respect to C.)

Furthermore, for $1 \le i \le n$, the ith column of $\mathbf{A}_{BC} = [L(\mathbf{v}_i)]_C$.

Theorem 5.5 asserts that once ordered bases for $\mathcal{V}$ and $\mathcal{W}$ have been selected, *each linear transformation $L: \mathcal{V} \to \mathcal{W}$ is equivalent to multiplication by a unique corresponding matrix*. The matrix $\mathbf{A}_{BC}$ in this theorem is known as the **matrix of the linear transformation L with respect to the ordered bases** B (for $\mathcal{V}$) **and** C (for $\mathcal{W}$). Theorem 5.5 also says that the matrix $\mathbf{A}_{BC}$ is computed as follows: find the image of each domain basis element $\mathbf{v}_i$ in turn, and then express these images in C-coordinates to get the respective columns of $\mathbf{A}_{BC}$.

The subscripts B and C on $\mathbf{A}$ are sometimes omitted when the bases being used are clear from context. Beware! If different ordered bases are chosen for $\mathcal{V}$ or $\mathcal{W}$, the matrix for the linear transformation will probably change.

Proof. Consider the $m \times n$ matrix $\mathbf{A}_{BC}$ whose ith column equals $[L(\mathbf{v}_i)]_C$, for $1 \le i \le n$. Let $\mathbf{v} \in \mathcal{V}$. We first prove that $\mathbf{A}_{BC}[\mathbf{v}]_B = [L(\mathbf{v})]_C$.

Suppose that $[\mathbf{v}]_B = [k_1, k_2, \ldots, k_n]$. Then $\mathbf{v} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_n\mathbf{v}_n$, and $L(\mathbf{v}) = k_1L(\mathbf{v}_1) + k_2L(\mathbf{v}_2) + \cdots + k_nL(\mathbf{v}_n)$, by Theorem 5.1. Hence,

$$\begin{aligned} [L(\mathbf{v})]_C &= [k_1L(\mathbf{v}_1) + k_2L(\mathbf{v}_2) + \cdots + k_nL(\mathbf{v}_n)]_C \\ &= k_1[L(\mathbf{v}_1)]_C + k_2[L(\mathbf{v}_2)]_C + \cdots + k_n[L(\mathbf{v}_n)]_C \qquad \text{by Theorem 4.19} \end{aligned}$$

$$= k_1(\text{1st column of } \mathbf{A}_{BC}) + k_2(\text{2nd column of } \mathbf{A}_{BC}) + \cdots + k_n(n\text{th column of } \mathbf{A}_{BC})$$

$$= \mathbf{A}_{BC}\begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_n \end{bmatrix} = \mathbf{A}_{BC}[\mathbf{v}]_B.$$

To complete the proof, we need to establish the uniqueness of $\mathbf{A}_{BC}$. Suppose that $\mathbf{H}$ is an $m \times n$ matrix such that $\mathbf{H}[\mathbf{v}]_B = [L(\mathbf{v})]_C$ for all $\mathbf{v} \in \mathcal{V}$. We will show that $\mathbf{H} = \mathbf{A}_{BC}$. It is enough to show that the ith column of $\mathbf{H}$ equals the ith column of $\mathbf{A}_{BC}$, for $1 \le i \le n$. Consider the ith vector, $\mathbf{v}_i$, of the ordered basis B for $\mathcal{V}$. Since $[\mathbf{v}_i]_B = \mathbf{e}_i$, we have ith column of $\mathbf{H} = \mathbf{He}_i = \mathbf{H}[\mathbf{v}_i]_B = [L(\mathbf{v}_i)]_C$, and this is the ith column of $\mathbf{A}_{BC}$. □

Notice that in the special case where the codomain $\mathcal{W}$ is $\mathbb{R}^m$, and the basis C for $\mathcal{W}$ is the standard basis, Theorem 5.5 asserts that the ith column of $\mathbf{A}_{BC}$ is simply $L(\mathbf{v}_i)$ itself (why?).

Example 2

Table 5.1 lists the matrices corresponding to some geometric linear operators on $\mathbb{R}^3$, with respect to the standard basis. The columns of each matrix are quickly calculated using Theorem 5.5, since we simply find the images $L(\mathbf{e}_1)$, $L(\mathbf{e}_2)$, and $L(\mathbf{e}_3)$ of the domain basis elements $\mathbf{e}_1, \mathbf{e}_2$, and $\mathbf{e}_3$. (Each image is equal to its coordinatization in the codomain since we are using the standard basis for the codomain as well.)

Once the matrix for each transformation is calculated, we can easily find the image of any vector using matrix multiplication. For example, to find the effect of the reflection L_1 in Table 5.1 on the vector $[3, -4, 2]$, we simply multiply by the matrix for L_1 to get

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

■

Example 3

We will find the matrix for the linear transformation $L\colon \mathcal{P}_3 \to \mathbb{R}^3$ given by

$$L(a_3x^3 + a_2x^2 + a_1x + a_0) = [a_0 + a_1, 2a_2, a_3 - a_0]$$

with respect to the standard ordered bases $B = (x^3, x^2, x, 1)$ for $\mathcal{P}_3$ and $C = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ for $\mathbb{R}^3$. We first need to find $L(\mathbf{v})$, for each $\mathbf{v} \in B$. By definition of L, we have

$$L(x^3) = [0,0,1], \quad L(x^2) = [0,2,0], \quad L(x) = [1,0,0], \text{ and } \quad L(1) = [1,0,-1].$$

Table 5.1 Matrices for several geometric linear operators on $\mathbb{R}^3$

Transformation	Formula	Matrix
Reflection (through xy-plane)	$L_1\left(\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}\right) = \begin{bmatrix} a_1 \\ a_2 \\ -a_3 \end{bmatrix}$	$\begin{array}{c} \begin{array}{ccc} L_1(\mathbf{e}_1) & L_1(\mathbf{e}_2) & L_1(\mathbf{e}_3) \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \end{array}$
Contraction or dilation	$L_2\left(\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}\right) = \begin{bmatrix} ca_1 \\ ca_2 \\ ca_3 \end{bmatrix}$, for $c \in \mathbb{R}$	$\begin{array}{c} \begin{array}{ccc} L_2(\mathbf{e}_1) & L_2(\mathbf{e}_2) & L_2(\mathbf{e}_3) \end{array} \\ \begin{bmatrix} c & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & c \end{bmatrix} \end{array}$
Projection (onto xy-plane)	$L_3\left(\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}\right) = \begin{bmatrix} a_1 \\ a_2 \\ 0 \end{bmatrix}$	$\begin{array}{c} \begin{array}{ccc} L_3(\mathbf{e}_1) & L_3(\mathbf{e}_2) & L_3(\mathbf{e}_3) \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$
Rotation (about z-axis through angle θ) (relative to the xy-plane)	$L_4\left(\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}\right) = \begin{bmatrix} a_1\cos\theta - a_2\sin\theta \\ a_1\sin\theta + a_2\cos\theta \\ a_3 \end{bmatrix}$	$\begin{array}{c} \begin{array}{ccc} L_4(\mathbf{e}_1) & L_4(\mathbf{e}_2) & L_4(\mathbf{e}_3) \end{array} \\ \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array}$
Shear (in the z-direction with factor k) (analog of Exercise 11 in Section 5.1)	$L_5\left(\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}\right) = \begin{bmatrix} a_1 + ka_3 \\ a_2 + ka_3 \\ a_3 \end{bmatrix}$	$\begin{array}{c} \begin{array}{ccc} L_5(\mathbf{e}_1) & L_5(\mathbf{e}_2) & L_5(\mathbf{e}_3) \end{array} \\ \begin{bmatrix} 1 & 0 & k \\ 0 & 1 & k \\ 0 & 0 & 1 \end{bmatrix} \end{array}$

Since we are using the standard basis C for $\mathbb{R}^3$, each of these images in $\mathbb{R}^3$ is its own C-coordinatization. Then by Theorem 5.5, the matrix $\mathbf{A}_{BC}$ for L is the matrix whose columns are these images; that is,

$$\mathbf{A}_{BC} = \begin{array}{c} \begin{array}{cccc} L(x^3) & L(x^2) & L(x) & L(1) \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 2 & 0 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix} \end{array}.$$

We will compute $L(5x^3 - x^2 + 3x + 2)$ using this matrix. Now, $[5x^3 - x^2 + 3x + 2]_B = [5, -1, 3, 2]$. Hence, multiplication by $\mathbf{A}_{BC}$ gives

$$\left[L(5x^3 - x^2 + 3x + 2)\right]_C = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 2 & 0 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 5 \\ -1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \\ 3 \end{bmatrix}.$$

Since C is the standard basis for $\mathbb{R}^3$, we have $L(5x^3 - x^2 + 3x + 2) = [5, -2, 3]$, which can be quickly verified to be the correct answer. ∎

Example 4

We will find the matrix for the same linear transformation $L: \mathcal{P}_3 \to \mathbb{R}^3$ of Example 3 with respect to the different ordered bases

$$D = (x^3 + x^2,\ x^2 + x,\ x + 1,\ 1)$$
$$\text{and} \quad E = ([-2,1,-3],[1,-3,0],[3,-6,2]).$$

You should verify that D and E are bases for $\mathcal{P}_3$ and $\mathbb{R}^3$, respectively.

We first need to find $L(\mathbf{v})$, for each $\mathbf{v} \in D$. By definition of L, we have $L(x^3 + x^2) = [0,2,1]$, $L(x^2 + x) = [1,2,0]$, $L(x + 1) = [2,0,-1]$, and $L(1) = [1,0,-1]$. Now we must find the coordinatization of each of these images in terms of the basis E for $\mathbb{R}^3$. Since we must solve for the coordinates of many vectors, it is quicker to use the transition matrix $\mathbf{Q}$ from the standard basis C for $\mathbb{R}^3$ to the basis E. From Theorem 4.22, $\mathbf{Q}$ is the inverse of the matrix whose columns are the vectors in E; that is,

$$\mathbf{Q} = \begin{bmatrix} -2 & 1 & 3 \\ 1 & -3 & -6 \\ -3 & 0 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} -6 & -2 & 3 \\ 16 & 5 & -9 \\ -9 & -3 & 5 \end{bmatrix}.$$

Now, multiplying $\mathbf{Q}$ by each of the images, we get

$$\left[L(x^3 + x^2)\right]_E = \mathbf{Q}\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}, \qquad \left[L(x^2 + x)\right]_E = \mathbf{Q}\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} -10 \\ 26 \\ -15 \end{bmatrix},$$

$$[L(x + 1)]_E = \mathbf{Q}\begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} -15 \\ 41 \\ -23 \end{bmatrix}, \quad \text{and} \quad [L(1)]_E = \mathbf{Q}\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} -9 \\ 25 \\ -14 \end{bmatrix}.$$

By Theorem 5.5, the matrix $\mathbf{A}_{DE}$ for L is the matrix whose columns are these products.

$$\mathbf{A}_{DE} = \begin{bmatrix} -1 & -10 & -15 & -9 \\ 1 & 26 & 41 & 25 \\ -1 & -15 & -23 & -14 \end{bmatrix}$$

We will compute $L(5x^3 - x^2 + 3x + 2)$ using this matrix. We must first find the representation for $5x^3 - x^2 + 3x + 2$ in terms of the basis D. Solving $5x^3 - x^2 + 3x + 2 = a(x^3 + x^2) + b(x^2 + x) + c(x + 1) + d(1)$ for a, b, c, and d, we get the unique solution $a = 5$, $b = -6$, $c = 9$, and $d = -7$ (verify!). Hence, $\left[5x^3 - x^2 + 3x + 2\right]_D = [5,-6,9,-7]$. Then

$$\left[L(5x^3 - x^2 + 3x + 2)\right]_E = \begin{bmatrix} -1 & -10 & -15 & -9 \\ 1 & 26 & 41 & 25 \\ -1 & -15 & -23 & -14 \end{bmatrix}\begin{bmatrix} 5 \\ -6 \\ 9 \\ -7 \end{bmatrix} = \begin{bmatrix} -17 \\ 43 \\ -24 \end{bmatrix}.$$

This answer represents a coordinate vector in terms of the basis E, and so

$$L(5x^3 - x^2 + 3x + 2) = -17\begin{bmatrix} -2 \\ 1 \\ -3 \end{bmatrix} + 43\begin{bmatrix} 1 \\ -3 \\ 0 \end{bmatrix} - 24\begin{bmatrix} 3 \\ -6 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ -2 \\ 3 \end{bmatrix},$$

which agrees with the answer in Example 3.

Finding the New Matrix for a Linear Transformation after a Change of Basis

The next theorem indicates precisely how the matrix for a linear transformation changes when we alter the bases for the domain and codomain.

Theorem 5.6 Let $\mathcal{V}$ and $\mathcal{W}$ be two nontrivial finite dimensional vector spaces with ordered bases B and C, respectively. Let $L: \mathcal{V} \to \mathcal{W}$ be a linear transformation with matrix $\mathbf{A}_{BC}$ with respect to bases B and C. Suppose that D and E are other ordered bases for $\mathcal{V}$ and $\mathcal{W}$, respectively. Let $\mathbf{P}$ be the transition matrix from B to D, and let $\mathbf{Q}$ be the transition matrix from C to E. Then the matrix $\mathbf{A}_{DE}$ for L with respect to bases D and E is given by $\mathbf{A}_{DE} = \mathbf{Q}\mathbf{A}_{BC}\mathbf{P}^{-1}$.

The situation in Theorem 5.6 is summarized in Figure 5.8.

Proof. For all $\mathbf{v} \in \mathcal{V}$,

$$\begin{array}{rll} & \mathbf{A}_{BC}[\mathbf{v}]_B = [L(\mathbf{v})]_C & \text{by Theorem 5.5} \\ \Rightarrow & \mathbf{Q}\mathbf{A}_{BC}[\mathbf{v}]_B = \mathbf{Q}[L(\mathbf{v})]_C & \\ \Rightarrow & \mathbf{Q}\mathbf{A}_{BC}[\mathbf{v}]_B = [L(\mathbf{v})]_E & \text{because } \mathbf{Q} \text{ is the transition matrix from } C \text{ to } E \\ \Rightarrow & \mathbf{Q}\mathbf{A}_{BC}\mathbf{P}^{-1}[\mathbf{v}]_D = [L(\mathbf{v})]_E. & \text{because } \mathbf{P}^{-1} \text{ is the transition matrix from } D \text{ to } B \end{array}$$

However, $\mathbf{A}_{DE}$ is the *unique* matrix such that $\mathbf{A}_{DE}[\mathbf{v}]_D = [L(\mathbf{v})]_E$, for all $\mathbf{v} \in \mathcal{V}$. Hence, $\mathbf{A}_{DE} = \mathbf{Q}\mathbf{A}_{BC}\mathbf{P}^{-1}$. □

$$\begin{tikzcd}[column sep=huge, row sep=huge]
{[\mathbf{v}]_B} \arrow[r, "\mathbf{A}_{BC}", "\text{(Matrix for } L \text{ using } B, C)"'] \arrow[d, "\text{Transition matrix } \mathbf{P}"'] & {[L(\mathbf{v})]_C} \arrow[d, "\text{Transition matrix } \mathbf{Q}"] \\
{[\mathbf{v}]_D} \arrow[r, "\mathbf{A}_{DE}", "\text{(Matrix for } L \text{ using } D, E)"'] & {[L(\mathbf{v})]_E}
\end{tikzcd}$$

FIGURE 5.8

Relationship between matrices $\mathbf{A}_{BC}$ and $\mathbf{A}_{DE}$ for a linear transformation under a change of basis

Theorem 5.6 gives us an alternate method for finding the matrix of a linear transformation with respect to one pair of bases when the matrix for another pair of bases is known.

Example 5

Recall the linear transformation $L: \mathcal{P}_3 \to \mathbb{R}^3$ from Examples 3 and 4, given by

$$L(a_3x^3 + a_2x^2 + a_1x + a_0) = [a_0 + a_1, 2a_2, a_3 - a_0].$$

Example 3 shows that the matrix for L using the standard bases B (for $\mathcal{P}_3$) and C (for $\mathbb{R}^3$) is

$$\mathbf{A}_{BC} = \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 2 & 0 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix}.$$

Also, in Example 4, we computed directly to find the matrix $\mathbf{A}_{DE}$ for the ordered bases $D = (x^3 + x^2, x^2 + x, x + 1, 1)$ for $\mathcal{P}_3$ and $E = ([-2, 1, -3], [1, -3, 0], [3, -6, 2])$ for $\mathbb{R}^3$. Instead, we now use Theorem 5.6 to calculate $\mathbf{A}_{DE}$. Recall from Example 4 that the transition matrix $\mathbf{Q}$ from bases C to E is

$$\mathbf{Q} = \begin{bmatrix} -6 & -2 & 3 \\ 16 & 5 & -9 \\ -9 & -3 & 5 \end{bmatrix}.$$

Also, the transition matrix $\mathbf{P}^{-1}$ from bases D to B is

$$\mathbf{P}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}. \quad \text{(Verify!)}$$

Hence,

$$\mathbf{A}_{DE} = \mathbf{Q}\mathbf{A}_{BC}\mathbf{P}^{-1} = \begin{bmatrix} -6 & -2 & 3 \\ 16 & 5 & -9 \\ -9 & -3 & 5 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 2 & 0 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} -1 & -10 & -15 & -9 \\ 1 & 26 & 41 & 25 \\ -1 & -15 & -23 & -14 \end{bmatrix},$$

which agrees with the result obtained for $\mathbf{A}_{DE}$ in Example 4.

Linear Operators and Similarity

Suppose L is a linear operator on a finite dimensional vector space $\mathcal{V}$. If B is a basis for $\mathcal{V}$, then there is some matrix $\mathbf{A}_{BB}$ for L with respect to B. Also, if C is another basis for $\mathcal{V}$, then there is some matrix $\mathbf{A}_{CC}$ for L with respect to C. Let $\mathbf{P}$ be the transition matrix from B to C (see Figure 5.9). Notice that by Theorem 5.6 we have $\mathbf{A}_{BB} = \mathbf{P}^{-1}\mathbf{A}_{CC}\mathbf{P}$, and so, by the definition of similar matrices, $\mathbf{A}_{BB}$ and $\mathbf{A}_{CC}$ are similar. This argument shows that any two matrices for the same linear operator with respect to different bases are similar. In fact, the converse is also true (see Exercise 20).

Example 6

Consider the linear operator $L\colon \mathbb{R}^3 \to \mathbb{R}^3$ whose matrix with respect to the standard basis B for $\mathbb{R}^3$ is

$$\mathbf{A}_{BB} = \frac{1}{7}\begin{bmatrix} 6 & 3 & -2 \\ 3 & -2 & 6 \\ -2 & 6 & 3 \end{bmatrix}.$$

We will use eigenvectors to find another basis D for $\mathbb{R}^3$ so that with respect to D, L has a much simpler matrix representation. Now, $p_{\mathbf{A}_{BB}}(x) = |x\mathbf{I}_3 - \mathbf{A}_{BB}| = x^3 - x^2 - x + 1 = (x-1)^2(x+1)$ (verify!).

By row reducing $(1\mathbf{I}_3 - \mathbf{A}_{BB})$ and $(-1\mathbf{I}_3 - \mathbf{A}_{BB})$ we find the basis $\{[3,1,0], [-2,0,1]\}$ for the eigenspace E_1 for $\mathbf{A}_{BB}$ and the basis $\{[1,-3,2]\}$ for the eigenspace E_{-1} for $\mathbf{A}_{BB}$. (Again, verify!) A quick check verifies that $D = \{[3,1,0], [-2,0,1], [1,-3,2]\}$ is a basis for $\mathbb{R}^3$ consisting of eigenvectors for $\mathbf{A}_{BB}$.

Next, recall that $\mathbf{A}_{DD}$ is similar to $\mathbf{A}_{BB}$. In particular, from the remarks right before this example, $\mathbf{A}_{DD} = \mathbf{P}^{-1}\mathbf{A}_{BB}\mathbf{P}$, where $\mathbf{P}$ is the transition matrix from D to B. Now, the matrix whose columns are the vectors in D is the transition matrix from D to the standard basis B. Thus,

$$\mathbf{P} = \begin{bmatrix} 3 & -2 & 1 \\ 1 & 0 & -3 \\ 0 & 1 & 2 \end{bmatrix}, \quad \text{with} \quad \mathbf{P}^{-1} = \frac{1}{14}\begin{bmatrix} 3 & 5 & 6 \\ -2 & 6 & 10 \\ 1 & -3 & 2 \end{bmatrix}.$$

$$\begin{tikzcd}[column sep=huge, row sep=huge]
{[\mathbf{v}]_B} \arrow[r, "\mathbf{A}_{BB}", "\text{(Matrix for } L \text{ using } B\text{)}"'] & {[L(\mathbf{v})]_B} \arrow[d, "\text{Transition matrix } \mathbf{P}"] \\
{[\mathbf{v}]_C} \arrow[u, "\text{Transition matrix } \mathbf{P}^{-1}"] \arrow[r, "\mathbf{A}_{CC}", "\text{(Matrix for } L \text{ using } C\text{)}"'] & {[L(\mathbf{v})]_C}
\end{tikzcd}$$

FIGURE 5.9

Relationship between matrices $\mathbf{A}_{BB}$ and $\mathbf{A}_{CC}$ for a linear operator under a change of basis as the transition matrix from B to D. Then,

$$\mathbf{A}_{DD} = \mathbf{P}^{-1}\mathbf{A}_{BB}\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix},$$

a diagonal matrix with the eigenvalues 1 and -1 on the main diagonal.

Written in this form, the operator L is more comprehensible. Compare $\mathbf{A}_{DD}$ to the matrix for a reflection through the xy-plane given in Table 5.1. Now, because D is not the standard basis for $\mathbb{R}^3$, L is *not* a reflection through the xy-plane. But we can show that L is a reflection of all vectors in $\mathbb{R}^3$ through the plane formed by the two basis vectors for E_1 (that is, the plane is the eigenspace E_1 itself). By the uniqueness assertion in Theorem 5.4, it is enough to show that L acts as a reflection through the plane E_1 for each of the three basis vectors of D.

Since $[3,1,0]$ and $[-2,0,1]$ are in the plane E_1, we need to show that L "reflects" these vectors to themselves. But this is true since $L([3,1,0]) = 1[3,1,0] = [3,1,0]$, and similarly for $[-2,0,1]$. Finally, notice that $[1,-3,2]$ is orthogonal to the plane E_1 (since it is orthogonal to both $[3,1,0]$ and $[-2,0,1]$). Therefore, we need to show that L "reflects" this vector to its opposite. But, $L([1,-3,2]) = -1[1,-3,2] = -[1,-3,2]$, and we are done. Hence, L is a reflection through the plane E_1.

Because the matrix $\mathbf{A}_{DD}$ in Example 6 is diagonal, it is easy to see that $p_{\mathbf{A}_{DD}}(x) = (x-1)^2(x+1)$. In Exercise 6 of Section 3.4, you were asked to prove that similar matrices have the same characteristic polynomial. Therefore, $p_{\mathbf{A}_{BB}}(x)$ also equals $(x-1)^2(x+1)$.

Matrix for the Composition of Linear Transformations

Our final theorem for this section shows how to find the corresponding matrix for the composition of linear transformations. The proof is left as Exercise 15.

Theorem 5.7 Let $\mathcal{V}_1, \mathcal{V}_2$, and $\mathcal{V}_3$ be nontrivial finite dimensional vector spaces with ordered bases B, C, and D, respectively. Let $L_1\colon \mathcal{V}_1 \to \mathcal{V}_2$ be a linear transformation with matrix $\mathbf{A}_{BC}$ with respect to bases B and C, and let $L_2\colon \mathcal{V}_2 \to \mathcal{V}_3$ be a linear transformation with matrix $\mathbf{A}_{CD}$ with respect to bases C and D. Then the matrix $\mathbf{A}_{BD}$ for the composite linear transformation $L_2 \circ L_1\colon \mathcal{V}_1 \to \mathcal{V}_3$ with respect to bases B and D is the product $\mathbf{A}_{CD}\mathbf{A}_{BC}$.

Theorem 5.7 can be generalized to compositions of several linear transformations, as in the next example.

Example 7

Let $L_1, L_2, \ldots, L_5$ be the geometric linear operators on $\mathbb{R}^3$ given in Table 5.1. Let $\mathbf{A}_1, \ldots, \mathbf{A}_5$ be the matrices for these operators using the standard basis for $\mathbb{R}^3$. Then, the matrix for the composition $L_4 \circ L_5$ is

$$\mathbf{A}_4\mathbf{A}_5 = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & k \\ 0 & 1 & k \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & k\cos\theta - k\sin\theta \\ \sin\theta & \cos\theta & k\sin\theta + k\cos\theta \\ 0 & 0 & 1 \end{bmatrix}.$$

Similarly, the matrix for the composition $L_2 \circ L_3 \circ L_1 \circ L_5$ is

$$\mathbf{A}_2\mathbf{A}_3\mathbf{A}_1\mathbf{A}_5 = \begin{bmatrix} c & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & c \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & k \\ 0 & 1 & k \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} c & 0 & kc \\ 0 & c & kc \\ 0 & 0 & 0 \end{bmatrix}.$$

■

♦ **Supplemental Material:** You have now covered the prerequisites for Section 7.3, "Complex Vector Spaces."

♦ **Application:** You have now covered the prerequisites for Section 8.8, "Computer Graphics."

New Vocabulary

matrix for a linear transformation

Highlights

- A linear transformation between finite dimensional vector spaces is uniquely determined once the images of an ordered basis for the domain are specified. (More specifically, let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces, with $\dim(\mathcal{V}) = n$. Let $B = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ be an ordered basis for $\mathcal{V}$, and let $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$ be any n (not necessarily distinct) vectors in $\mathcal{W}$. Then there is a unique linear transformation $L: \mathcal{V} \to \mathcal{W}$ such that $L(\mathbf{v}_i) = \mathbf{w}_i$, for $1 \le i \le n$.)
- Every linear transformation between (nontrivial) finite dimensional vector spaces has a unique matrix $\mathbf{A}_{BC}$ with respect to the ordered bases B and C chosen for the domain and codomain, respectively. (More specifically, let $L: \mathcal{V} \to \mathcal{W}$ be a linear transformation, with $\dim(\mathcal{V}) = n, \dim(\mathcal{W}) = m$. Let $B = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ and $C = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m)$ be ordered bases for $\mathcal{V}$ and $\mathcal{W}$, respectively. Then there is a unique $m \times n$ matrix $\mathbf{A}_{BC}$ such that $\mathbf{A}_{BC}[\mathbf{v}]_B = [L(\mathbf{v})]_C$, for all $\mathbf{v} \in \mathcal{V}$.)
- If $\mathbf{A}_{BC}$ is the matrix for a linear transformation with respect to the ordered bases B and C chosen for the domain and codomain, respectively, then the ith column of $\mathbf{A}_{BC}$ is the C-coordinatization of the image of the ith vector in B. That is, the ith column of $\mathbf{A}_{BC}$ equals $[L(\mathbf{v}_i)]_C$.
- After a change of bases for the domain and codomain, the new matrix for a given linear transformation can be found using the original matrix and the transition

matrices between bases. (More specifically, let $L: \mathcal{V} \to \mathcal{W}$ be a linear transformation between (nontrivial) finite dimensional vector spaces with ordered bases B and C, respectively, and with matrix $\mathbf{A}_{BC}$ in terms of bases B and C. If D and E are other ordered bases for $\mathcal{V}$ and $\mathcal{W}$, respectively, and $\mathbf{P}$ is the transition matrix from B to D, and $\mathbf{Q}$ is the transition matrix from C to E, then the matrix $\mathbf{A}_{DE}$ for L in terms of bases D and E is $\mathbf{A}_{DE} = \mathbf{Q}\mathbf{A}_{BC}\mathbf{P}^{-1}$.)

- Matrices for several useful geometric operators on $\mathbb{R}^3$ are given in Table 5.1.
- The matrix for a linear operator (on a finite dimensional vector space) after a change of basis is similar to the original matrix.
- The matrix for the composition of linear transformations (using the same ordered bases) is the product of the matrices for the individual linear transformations in reverse order. (More specifically, if $L_1: \mathcal{V}_1 \to \mathcal{V}_2$ is a linear transformation with matrix $\mathbf{A}_{BC}$ with respect to ordered bases B and C, and $L_2: \mathcal{V}_2 \to \mathcal{V}_3$ is a linear transformation with matrix $\mathbf{A}_{CD}$ with respect to bases C and D, then the matrix $\mathbf{A}_{BD}$ for $L_2 \circ L_1: \mathcal{V}_1 \to \mathcal{V}_3$ with respect to bases B and D is given by $\mathbf{A}_{BD} = \mathbf{A}_{CD}\mathbf{A}_{BC}$.)

EXERCISES FOR SECTION 5.2

1. Verify that the correct matrix is given for each of the geometric linear operators in Table 5.1.

2. For each of the following linear transformations $L: \mathcal{V} \to \mathcal{W}$, find the matrix for L with respect to the standard bases for $\mathcal{V}$ and $\mathcal{W}$.

★**(a)** $L: \mathbb{R}^3 \to \mathbb{R}^3$ given by $L([x,y,z]) = [-6x + 4y - z, -2x + 3y - 5z, 3x - y + 7z]$

(b) $L: \mathbb{R}^4 \to \mathbb{R}^2$ given by $L([x,y,z,w]) = [3x - 5y + z - 2w, 5x + y - 2z + 8w]$

★**(c)** $L: \mathcal{P}_3 \to \mathbb{R}^3$ given by $L(ax^3 + bx^2 + cx + d) = [4a - b + 3c + 3d, a + 3b - c + 5d, -2a - 7b + 5c - d]$

(d) $L: \mathcal{P}_3 \to \mathcal{M}_{22}$ given by

$$L(ax^3 + bx^2 + cx + d) = \begin{bmatrix} -3a - 2c & -b + 4d \\ 4b - c + 3d & -6a - b + 2d \end{bmatrix}$$

3. For each of the following linear transformations $L: \mathcal{V} \to \mathcal{W}$, find the matrix $\mathbf{A}_{BC}$ for L with respect to the given bases B for $\mathcal{V}$ and C for $\mathcal{W}$ using the method of Theorem 5.5:

★**(a)** $L: \mathbb{R}^3 \to \mathbb{R}^2$ given by $L([x,y,z]) = [-2x + 3z, x + 2y - z]$ with $B = ([1,-3,2],[-4,13,-3],[2,-3,20])$ and $C = ([-2,-1],[5,3])$

(b) $L: \mathbb{R}^2 \to \mathbb{R}^3$ given by $L([x,y]) = [13x - 9y, -x - 2y, -11x + 6y]$ with $B = ([2,3],[-3,-4])$ and $C = ([-1,2,2],[-4,1,3],[1,-1,-1])$

★(c) $L: \mathbb{R}^2 \to \mathcal{P}_2$ given by $L([a,b]) = (-a+5b)x^2 + (3a-b)x + 2b$ with $B = ([5,3],[3,2])$ and $C = (3x^2 - 2x, -2x^2 + 2x - 1, x^2 - x + 1)$

(d) $L: \mathcal{M}_{22} \to \mathbb{R}^3$ given by $L\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = [a - c + 2d, 2a + b - d, -2c + d]$

with $B = \left(\begin{bmatrix} 2 & 5 \\ 2 & -1 \end{bmatrix}, \begin{bmatrix} -2 & -2 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -3 & -4 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} -1 & -3 \\ 0 & 1 \end{bmatrix}\right)$ and

$C = ([7,0,-3],[2,-1,-2],[-2,0,1])$

★(e) $L: \mathcal{P}_2 \to \mathcal{M}_{23}$ given by

$$L(ax^2 + bx + c) = \begin{bmatrix} -a & 2b + c & 3a - c \\ a + b & c & -2a + b - c \end{bmatrix}$$

with $B = (-5x^2 - x - 1, -6x^2 + 3x + 1, 2x + 1)$ and $C =$
$\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right)$

4. In each case, find the matrix $\mathbf{A}_{DE}$ for the given linear transformation $L: \mathcal{V} \to \mathcal{W}$ with respect to the given bases D and E by first finding the matrix for L with respect to the standard bases B and C for $\mathcal{V}$ and $\mathcal{W}$, respectively, and then using the method of Theorem 5.6.

★(a) $L: \mathbb{R}^3 \to \mathbb{R}^3$ given by $L([a,b,c]) = [-2a + b, -b - c, a + 3c]$ with $D = ([15,-6,4],[2,0,1],[3,-1,1])$ and $E = ([1,-3,1],[0,3,-1],[2,-2,1])$

★(b) $L: \mathcal{M}_{22} \to \mathbb{R}^2$ given by

$$L\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = [6a - b + 3c - 2d, -2a + 3b - c + 4d]$$

with

$$D = \left(\begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\right) \quad \text{and}$$

$$E = ([-2,5],[-1,2])$$

(c) $L: \mathcal{M}_{22} \to \mathcal{P}_2$ given by

$$L\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = (b - c)x^2 + (3a - d)x + (4a - 2c + d)$$

with

$$D = \left(\begin{bmatrix} 3 & -4 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} -2 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 2 & -2 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} -2 & 1 \\ 0 & 1 \end{bmatrix} \right) \quad \text{and}$$

$$E = (2x - 1, -5x^2 + 3x - 1, x^2 - 2x + 1)$$

5. Verify that the same matrix is obtained for L in Exercise 3(d) by first finding the matrix for L with respect to the standard bases and then using the method of Theorem 5.6.

6. In each case, find the matrix $\mathbf{A}_{BB}$ for each of the given linear operators $L: \mathcal{V} \to \mathcal{V}$ with respect to the given basis B by using the method of Theorem 5.5. Then, check your answer by calculating the matrix for L using the standard basis and applying the method of Theorem 5.6.

★**(a)** L: $\mathbb{R}^2 \to \mathbb{R}^2$ given by $L([x,y]) = [2x - y,\ x - 3y]$ with $B = ([4,-1], [-7,2])$

★**(b)** L: $\mathcal{P}_2 \to \mathcal{P}_2$ given by $L(ax^2 + bx + c) = (b - 2c)x^2 + (2a + c)x + (a - b - c)$ with $B = (2x^2 + 2x - 1, x, -3x^2 - 2x + 1)$

(c) L: $\mathcal{M}_{22} \to \mathcal{M}_{22}$ given by

$$L\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) = \begin{bmatrix} 2a - c + d & a - b \\ -3b - 2d & -a - 2c + 3d \end{bmatrix}$$

with

$$B = \left(\begin{bmatrix} -2 & -1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 3 & 1 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} -2 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} \right)$$

7. ★**(a)** Let L: $\mathcal{P}_3 \to \mathcal{P}_2$ be given by $L(\mathbf{p}) = \mathbf{p}'$, for $\mathbf{p} \in \mathcal{P}_3$. Find the matrix for L with respect to the standard bases for $\mathcal{P}_3$ and $\mathcal{P}_2$. Use this matrix to calculate $L(4x^3 - 5x^2 + 6x - 7)$ by matrix multiplication.

(b) Let L: $\mathcal{P}_2 \to \mathcal{P}_3$ be the indefinite integral linear transformation; that is, $L(\mathbf{p})$ is the vector $\int \mathbf{p}(x)\, dx$ with zero constant term. Find the matrix for L with respect to the standard bases for $\mathcal{P}_2$ and $\mathcal{P}_3$. Use this matrix to calculate $L(2x^2 - x + 5)$ by matrix multiplication.

8. Let $L: \mathbb{R}^2 \to \mathbb{R}^2$ be the linear operator that performs a counterclockwise rotation through an angle of $\frac{\pi}{6}$ radians (30°).

★**(a)** Find the matrix for L with respect to the standard basis for $\mathbb{R}^2$.

(b) Find the matrix for L with respect to the basis $B = ([4,-3],[3,-2])$.

9. Let $L: \mathcal{M}_{23} \to \mathcal{M}_{32}$ be given by $L(\mathbf{A}) = \mathbf{A}^T$.

(a) Find the matrix for L with respect to the standard bases.

★(b) Find the matrix for L with respect to the bases

$$B = \left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)$$ for $\mathcal{M}_{23}$, and

$$C = \left(\begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & -1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & -1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & -1 \end{bmatrix} \right)$$

for $\mathcal{M}_{32}$.

★**10.** Let B be a basis for $\mathcal{V}_1$, C be a basis for $\mathcal{V}_2$, and D be a basis for $\mathcal{V}_3$. Suppose $L_1\colon \mathcal{V}_1 \to \mathcal{V}_2$ and $L_2\colon \mathcal{V}_2 \to \mathcal{V}_3$ are represented, respectively, by the matrices

$$\mathbf{A}_{BC} = \begin{bmatrix} -2 & 3 & -1 \\ 4 & 0 & -2 \end{bmatrix} \quad \text{and} \quad \mathbf{A}_{CD} = \begin{bmatrix} 4 & -1 \\ 2 & 0 \\ -1 & -3 \end{bmatrix}.$$

Find the matrix $\mathbf{A}_{BD}$ representing the composition $L_2 \circ L_1\colon \mathcal{V}_1 \to \mathcal{V}_3$.

11. Let $L_1\colon \mathbb{R}^3 \to \mathbb{R}^4$ be given by $L_1([x,y,z]) = [x-y-z, 2y+3z, x+3y, -2x+z]$, and let $L_2\colon \mathbb{R}^4 \to \mathbb{R}^2$ be given by $L_2([x,y,z,w]) = [2y-2z+3w, x-z+w]$.

(a) Find the matrices for L_1 and L_2 with respect to the standard bases in each case.

(b) Find the matrix for $L_2 \circ L_1$ with respect to the standard bases for $\mathbb{R}^3$ and $\mathbb{R}^2$ using Theorem 5.7.

(c) Check your answer to part (b) by computing $(L_2 \circ L_1)([x,y,z])$ and finding the matrix for $L_2 \circ L_1$ directly from this result.

12. Let $\mathbf{A} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$, the matrix representing the counterclockwise rotation of $\mathbb{R}^2$ about the origin through an angle θ.

(a) Use Theorem 5.7 to show that

$$\mathbf{A}^2 = \begin{bmatrix} \cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{bmatrix}.$$

(b) Generalize the result of part (a) to show that for any integer $n \geq 1$,

$$\mathbf{A}^n = \begin{bmatrix} \cos n\theta & -\sin n\theta \\ \sin n\theta & \cos n\theta \end{bmatrix}.$$

13. Let $B = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ be an ordered basis for a vector space $\mathcal{V}$. Find the matrix with respect to B for each of the following linear operators $L\colon \mathcal{V} \to \mathcal{V}$:

★**(a)** $L(\mathbf{v}) = \mathbf{v}$, for all $\mathbf{v} \in \mathcal{V}$ (identity linear operator)

(b) $L(\mathbf{v}) = \mathbf{0}$, for all $\mathbf{v} \in \mathcal{V}$ (zero linear operator)

⋆(c) $L(\mathbf{v}) = c\mathbf{v}$, for all $\mathbf{v} \in \mathcal{V}$, and for some fixed $c \in \mathbb{R}$ (scalar linear operator)

(d) $L: \mathcal{V} \to \mathcal{V}$ given by $L(\mathbf{v}_1) = \mathbf{v}_2$, $L(\mathbf{v}_2) = \mathbf{v}_3, \ldots, L(\mathbf{v}_{n-1}) = \mathbf{v}_n$, $L(\mathbf{v}_n) = \mathbf{v}_1$ (forward replacement of basis vectors)

⋆(e) $L: \mathcal{V} \to \mathcal{V}$ given by $L(\mathbf{v}_1) = \mathbf{v}_n$, $L(\mathbf{v}_2) = \mathbf{v}_1, \ldots, L(\mathbf{v}_{n-1}) = \mathbf{v}_{n-2}$, $L(\mathbf{v}_n) = \mathbf{v}_{n-1}$ (reverse replacement of basis vectors)

14. Let $L: \mathbb{R}^n \to \mathbb{R}$ be a linear transformation. Prove that there is a vector $\mathbf{x}$ in $\mathbb{R}^n$ such that $L(\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$ for all $\mathbf{y} \in \mathbb{R}^n$.

►15. Prove Theorem 5.7.

16. Let $L: \mathbb{R}^3 \to \mathbb{R}^3$ be given by $L([x,y,z]) = [-4y - 13z, -6x + 5y + 6z, 2x - 2y - 3z]$.

(a) What is the matrix for L with respect to the standard basis for $\mathbb{R}^3$?

(b) What is the matrix for L with respect to the basis

$$B = ([-1,-6,2],[3,4,-1],[-1,-3,1])?$$

(c) What does your answer to part (b) tell you about the vectors in B? Explain.

17. In Example 6, verify that $p_{\mathbf{A}_{BB}}(x) = (x-1)^2(x+1)$, $\{[3,1,0],[-2,0,1]\}$ is a basis for the eigenspace E_1, $\{[1,-3,2]\}$ is a basis for the eigenspace E_{-1}, the transition matrices $\mathbf{P}$ and $\mathbf{P}^{-1}$ are as indicated, and, finally, $\mathbf{A}_{DD} = \mathbf{P}^{-1}\mathbf{A}_{BB}\mathbf{P}$ is a diagonal matrix with entries $1, 1$, and -1, respectively, on the main diagonal.

18. Let $L: \mathbb{R}^3 \to \mathbb{R}^3$ be the linear operator whose matrix with respect to the standard basis B for $\mathbb{R}^3$ is

$$\mathbf{A}_{BB} = \frac{1}{9}\begin{bmatrix} 8 & 2 & 2 \\ 2 & 5 & -4 \\ 2 & -4 & 5 \end{bmatrix}.$$

⋆(a) Calculate and factor $p_{\mathbf{A}_{BB}}(x)$. (Be sure to incorporate $\frac{1}{9}$ correctly into your calculations.)

⋆(b) Solve for a basis for each eigenspace for L. Combine these to form a basis C for $\mathbb{R}^3$.

⋆(c) Find the transition matrix $\mathbf{P}$ from C to B.

(d) Calculate $\mathbf{A}_{CC}$ using $\mathbf{A}_{BB}$, $\mathbf{P}$, and $\mathbf{P}^{-1}$.

(e) Use $\mathbf{A}_{CC}$ to give a geometric description of the operator L, as was done in Example 6.

19. Let L be a linear operator on a vector space $\mathcal{V}$ with ordered basis $B = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$. Suppose that k is a nonzero real number, and let C be the ordered basis $(k\mathbf{v}_1, \ldots, k\mathbf{v}_n)$ for $\mathcal{V}$. Show that $\mathbf{A}_{BB} = \mathbf{A}_{CC}$.

20. Let $\mathcal{V}$ be an n-dimensional vector space, and let $\mathbf{X}$ and $\mathbf{Y}$ be similar $n \times n$ matrices. Prove that there is a linear operator $L: \mathcal{V} \to \mathcal{V}$ and bases B and C such that $\mathbf{X}$ is the matrix for L with respect to B and $\mathbf{Y}$ is the matrix for L with respect to C. (Hint: Suppose that $\mathbf{Y} = \mathbf{P}^{-1}\mathbf{X}\mathbf{P}$. Choose any basis B for $\mathcal{V}$. Then create the linear operator $L: \mathcal{V} \to \mathcal{V}$ whose matrix with respect to B is $\mathbf{X}$. Let $\mathbf{v}_i$ be the vector so that $[\mathbf{v}_i]_B = i$th column of $\mathbf{P}$. Define C to be $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$. Prove that C is a basis for $\mathcal{V}$. Then show that $\mathbf{P}^{-1}$ is the transition matrix from B to C and that $\mathbf{Y}$ is the matrix for L with respect to C.)

21. Let $B = ([a,b],[c,d])$ be a basis for $\mathbb{R}^2$. Then $ad - bc \neq 0$ (why?). Let $L: \mathbb{R}^2 \to \mathbb{R}^2$ be a linear operator such that $L([a,b]) = [c,d]$ and $L([c,d]) = [a,b]$. Show that the matrix for L with respect to the standard basis for $\mathbb{R}^2$ is

$$\frac{1}{ad-bc}\begin{bmatrix} cd-ab & a^2-c^2 \\ d^2-b^2 & ab-cd \end{bmatrix}.$$

22. Let $L: \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation where $L(\mathbf{v})$ is the reflection of $\mathbf{v}$ through the line $y = mx$. (Assume that the initial point of $\mathbf{v}$ is the origin.) Show that the matrix for L with respect to the standard basis for $\mathbb{R}^2$ is

$$\frac{1}{1+m^2}\begin{bmatrix} 1-m^2 & 2m \\ 2m & m^2-1 \end{bmatrix}.$$

(Hint: Use Exercise 19 in Section 1.2.)

23. Find the set of all matrices with respect to the standard basis for $\mathbb{R}^2$ for all linear operators that

(a) Take all vectors of the form $[0,y]$ to vectors of the form $[0,y']$

(b) Take all vectors of the form $[x,0]$ to vectors of the form $[x',0]$

(c) Satisfy both parts (a) and (b) simultaneously

24. Let $\mathcal{V}$ and $\mathcal{W}$ be finite dimensional vector spaces, and let $\mathcal{Y}$ be a subspace of $\mathcal{V}$. Suppose that $L: \mathcal{Y} \to \mathcal{W}$ is a linear transformation. Prove that there is a linear transformation $L': \mathcal{V} \to \mathcal{W}$ such that $L'(\mathbf{y}) = L(\mathbf{y})$ for every $\mathbf{y} \in \mathcal{Y}$. ($L'$ is called an **extension** of L to $\mathcal{V}$.)

►**25.** Prove the uniqueness assertion in Theorem 5.4. (Hint: Let $\mathbf{v}$ be any vector in $\mathcal{V}$. Show that there is only one possible answer for $L(\mathbf{v})$ by expressing $L(\mathbf{v})$ as a linear combination of the $\mathbf{w}_i$'s.)

★**26.** True or False:

(a) If $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation, and $B = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ is an ordered basis for $\mathcal{V}$, then for any $\mathbf{v} \in \mathcal{V}$, $L(\mathbf{v})$ can be computed if $L(\mathbf{v}_1), L(\mathbf{v}_2), \ldots, L(\mathbf{v}_n)$ are known.

(b) There is a unique linear transformation $L: \mathbb{R}^3 \to \mathcal{P}_3$ such that $L([1,0,0]) = x^3 - x^2$, $L([0,1,0]) = x^3 - x^2$, and $L([0,0,1]) = x^3 - x^2$.

(c) If $\mathcal{V}, \mathcal{W}$ are nontrivial finite dimensional vector spaces and $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation, then there is a unique matrix $\mathbf{A}$ corresponding to L.

(d) If $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation and B is a (finite nonempty) ordered basis for $\mathcal{V}$, and C is a (finite nonempty) ordered basis for $\mathcal{W}$, then $[\mathbf{v}]_B = \mathbf{A}_{BC}[L(\mathbf{v})]_C$.

(e) If $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation and $B = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ is an ordered basis for $\mathcal{V}$, and C is a (finite nonempty) ordered basis for $\mathcal{W}$, then the ith column of $\mathbf{A}_{BC}$ is $[L(\mathbf{v}_i)]_C$.

(f) The matrix for the projection of $\mathbb{R}^3$ onto the xz-plane (with respect to the standard basis) is $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

(g) If $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation, and B and D are (finite nonempty) ordered bases for $\mathcal{V}$, and C and E are (finite nonempty) ordered bases for $\mathcal{W}$, then $\mathbf{A}_{DE}\mathbf{P} = \mathbf{Q}\mathbf{A}_{BC}$, where $\mathbf{P}$ is the transition matrix from B to D, and $\mathbf{Q}$ is the transition matrix from C to E.

(h) If $L: \mathcal{V} \to \mathcal{V}$ is a linear operator on a nontrivial finite dimensional vector space, and B and D are ordered bases for $\mathcal{V}$, then $\mathbf{A}_{BB}$ is similar to $\mathbf{A}_{DD}$.

(i) Similar square matrices have identical characteristic polynomials.

(j) If $L_1, L_2: \mathbb{R}^2 \to \mathbb{R}^2$ are linear transformations with matrices $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, respectively, with respect to the standard basis, then the matrix for $L_2 \circ L_1$ with respect to the standard basis equals $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

5.3 THE DIMENSION THEOREM

In this section, we introduce two special subspaces associated with a linear transformation $L: \mathcal{V} \to \mathcal{W}$: the kernel of L (a subspace of $\mathcal{V}$) and the range of L (a subspace of $\mathcal{W}$). We illustrate techniques for calculating bases for both the kernel and range and show their dimensions are related to the rank of any matrix for the linear transformation. We then use this to show that any matrix and its transpose have the same rank.

Kernel and Range

Definition Let $L: \mathcal{V} \to \mathcal{W}$ be a linear transformation. The **kernel** of L, denoted by $\ker(L)$, is the subset of all vectors in $\mathcal{V}$ that map to $\mathbf{0}_{\mathcal{W}}$. That is, $\ker(L) = \{\mathbf{v} \in \mathcal{V} \mid L(\mathbf{v}) = \mathbf{0}_{\mathcal{W}}\}$. The **range** of L, or, range(L), is the subset of all vectors in $\mathcal{W}$ that are the image of some vector in $\mathcal{V}$. That is, $\text{range}(L) = \{L(\mathbf{v}) \mid \mathbf{v} \in \mathcal{V}\}$.

Remember that the kernel[1] is a subset of the *domain* and that the range is a subset of the *codomain*. Since the kernel of $L: \mathcal{V} \to \mathcal{W}$ is the pre-image of the subspace $\{\mathbf{0}_{\mathcal{W}}\}$ of $\mathcal{W}$, it must be a subspace of $\mathcal{V}$ by Theorem 5.3. That theorem also assures us that the range of L is a subspace of $\mathcal{W}$. Hence, we have

Theorem 5.8 If $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation, then the kernel of L is a subspace of $\mathcal{V}$ and the range of L is a subspace of $\mathcal{W}$.

Example 1

Projection: For $n \geq 3$, consider the linear operator $L: \mathbb{R}^n \to \mathbb{R}^n$ given by $L([a_1, a_2, \ldots, a_n]) = [a_1, a_2, 0, \ldots, 0]$. Now, $\ker(L)$ consists of those elements of the domain that map to $[0, 0, \ldots, 0]$, the zero vector of the codomain. Hence, for vectors in the kernel, $a_1 = a_2 = 0$, but $a_3, \ldots, a_n$ can have any values. Thus,

$$\ker(L) = \left\{ [0, 0, a_3, \ldots, a_n] \middle| a_3, \ldots, a_n \in \mathbb{R} \right\}.$$

Notice that $\ker(L)$ is a subspace of the domain and that $\dim(\ker(L)) = n - 2$, because the standard basis vectors $\mathbf{e}_3, \ldots, \mathbf{e}_n$ of $\mathbb{R}^n$ span $\ker(L)$.

Also, range(L) consists of those elements of the codomain $\mathbb{P}^2$ that are images of domain elements. Hence, $\text{range}(L) = \{[a_1, a_2, 0, \ldots, 0] \mid a_1, a_2 \in \mathbb{R}\}$. Notice that range($L$) is a subspace of the codomain and that $\dim(\text{range}(L)) = 2$, since the standard basis vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ span range(L).

Example 2

Differentiation: Consider the linear transformation $L: \mathcal{P}_3 \to \mathcal{P}_2$ given by $L(ax^3 + bx^2 + cx + d) = 3ax^2 + 2bx + c$. Now, $\ker(L)$ consists of the polynomials in $\mathcal{P}_3$ that map to the zero polynomial in $\mathcal{P}_2$. However, if $3ax^2 + 2bx + c = 0$, we must have $a = b = c = 0$. Hence, $\ker(L) = \left\{ 0x^3 + 0x^2 + 0x + d \middle| d \in \mathbb{R} \right\}$; that is, $\ker(L)$ is just the subset of $\mathcal{P}_3$ of all constant polynomials. Notice that $\ker(L)$ is a subspace of $\mathcal{P}_3$ and that $\dim(\ker(L)) = 1$ because the single polynomial "1" spans $\ker(L)$.

[1] Some textbooks refer to the kernel of L as the **nullspace** of L.

Also, range(L) consists of all polynomials in the codomain $\mathcal{P}_2$ of the form $3ax^2 + 2bx + c$. Since every polynomial $Ax^2 + Bx + C$ of degree 2 or less can be expressed in this form (take $a = A/3$, $b = B/2$, $c = C$), range(L) is all of $\mathcal{P}_2$. Therefore, range(L) is a subspace of $\mathcal{P}_2$, and $\dim(\text{range}(L)) = 3$. ■

Example 3

Rotation: Recall that the linear transformation $L\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$L\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix},$$

for some (fixed) angle θ, represents the counterclockwise rotation of any vector $[x, y]$ with initial point at the origin through the angle θ.

Now, $\ker(L)$ consists of all vectors in the domain $\mathbb{R}^2$ that map to $[0,0]$ in the codomain $\mathbb{R}^2$. However, only $[0,0]$ itself is rotated by L to the zero vector. Hence, $\ker(L) = \{[0,0]\}$. Notice that $\ker(L)$ is a subspace of $\mathbb{R}^2$, and $\dim(\ker(L)) = 0$.

Also, range(L) is all of the codomain $\mathbb{R}^2$ because every nonzero vector $\mathbf{v}$ in $\mathbb{R}^2$ is the image of the vector of the same length at the angle θ *clockwise* from $\mathbf{v}$. Thus, range(L) $= \mathbb{R}^2$, and so, range(L) is a subspace of $\mathbb{R}^2$ with $\dim(\text{range}(L)) = 2$. ■

Finding the Kernel from the Matrix of a Linear Transformation

Consider the linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$ given by $L(\mathbf{X}) = \mathbf{AX}$, where $\mathbf{A}$ is a (fixed) $m \times n$ matrix and $\mathbf{X} \in \mathbb{R}^n$. Now, $\ker(L)$ is the subspace of all vectors $\mathbf{X}$ in the domain $\mathbb{R}^n$ that are solutions of the homogeneous system $\mathbf{AX} = \mathbf{O}$. If $\mathbf{B}$ is the reduced row echelon form matrix for $\mathbf{A}$, we find a basis for $\ker(L)$ by solving for particular solutions to the system $\mathbf{BX} = \mathbf{O}$ by systematically setting each independent variable equal to 1 in turn, while setting the others equal to 0. (You should be familiar with this process from the Diagonalization Method for finding fundamental eigenvectors in Section 3.4.) Thus, $\dim(\ker(L))$ equals the number of independent variables in the system $\mathbf{BX} = \mathbf{O}$.

We present an example of this technique.

Example 4

Let $L\colon \mathbb{R}^5 \longrightarrow \mathbb{R}^4$ be given by $L(\mathbf{X}) = \mathbf{AX}$, where

$$\mathbf{A} = \begin{bmatrix} 8 & 4 & 16 & 32 & 0 \\ 4 & 2 & 10 & 22 & -4 \\ -2 & -1 & -5 & -11 & 7 \\ 6 & 3 & 15 & 33 & -7 \end{bmatrix}.$$

To solve for $\ker(L)$, we first row reduce $\mathbf{A}$ to

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

The homogeneous system $\mathbf{BX} = \mathbf{O}$ has independent variables x_2 and x_4, and

$$\begin{cases} x_1 &= -\frac{1}{2}x_2 + 2x_4 \\ x_3 &= \qquad\quad - 3x_4 \\ x_5 &= \quad 0 \end{cases}.$$

We construct two particular solutions, first by setting $x_2 = 1$ and $x_4 = 0$ to obtain $\mathbf{v}_1 = [-\frac{1}{2}, 1, 0, 0, 0]$, and then setting $x_2 = 0$ and $x_4 = 1$, yielding $\mathbf{v}_2 = [2, 0, -3, 1, 0]$. The set $\{\mathbf{v}_1, \mathbf{v}_2\}$ forms a basis for $\ker(L)$, and thus, $\dim(\ker(L)) = 2$, the number of independent variables. The entire subspace $\ker(L)$ consists of all linear combinations of the basis vectors; that is,

$$\ker(L) = \{a\mathbf{v}_1 + b\mathbf{v}_2 \,|\, a, b \in \mathbb{R}\} = \left\{ \left[-\frac{1}{2}a + 2b, a, -3b, b, 0 \right] \middle|\, a, b \in \mathbb{R} \right\}.$$

Finally, note that we could have eliminated fractions in this basis, just as we did with fundamental eigenvectors in Section 3.4, by replacing $\mathbf{v}_1$ with $2\mathbf{v}_1 = [-1, 2, 0, 0, 0]$. ∎

Example 4 illustrates the following general technique:

Method for Finding a Basis for the Kernel of a Linear Transformation (Kernel Method)
Let $L: \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation given by $\mathbf{L}(\mathbf{X}) = \mathbf{AX}$ for some $m \times n$ matrix $\mathbf{A}$. To find a basis for $\ker(L)$, perform the following steps:

Step 1: Find $\mathbf{B}$, the reduced row echelon form of $\mathbf{A}$.

Step 2: Solve for one particular solution for each independent variable in the homogeneous system $\mathbf{BX} = \mathbf{O}$. The ith such solution, $\mathbf{v}_i$, is found by setting the ith independent variable equal to 1 and setting all other independent variables equal to 0.

Step 3: The set $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a basis for $\ker(L)$. (We can replace any $\mathbf{v}_i$ with $c\mathbf{v}_i$, where $c \neq 0$, to eliminate fractions.)

The method for finding a basis for $\ker(L)$ is practically identical to Step 3 of the Diagonalization Method of Section 3.4, in which we create a basis of fundamental eigenvectors for the eigenspace E_λ for a matrix $\mathbf{A}$. This is to be expected, since E_λ is really the kernel of the linear transformation L whose matrix is $(\lambda\mathbf{I}_n - \mathbf{A})$.

Finding the Range from the Matrix of a Linear Transformation

Next, we determine a method for finding a basis for the range of $L: \mathbb{R}^n \to \mathbb{R}^m$ given by $L(\mathbf{X}) = \mathbf{AX}$. In Section 1.5, we saw that $\mathbf{AX}$ can be expressed as a linear combination of the columns of $\mathbf{A}$. In particular, if $\mathbf{X} = [x_1, \ldots x_n]$, then $\mathbf{AX} = x_1$ (1st column of $\mathbf{A}$) $+ \cdots + x_n$ (nth column of $\mathbf{A}$). Thus, range(L) is spanned by the set of columns of $\mathbf{A}$; that is, range(L) = span({columns of $\mathbf{A}$}). Note that $L(\mathbf{e}_i)$ equals the ith column of $\mathbf{A}$. Thus, we can also say that $\{L(\mathbf{e}_1), \ldots, L(\mathbf{e}_n)\}$ spans range(L).

The fact that the columns of $\mathbf{A}$ span range(L) combined with the Independence Test Method yields the following general technique for finding a basis for the range:

Method for Finding a Basis for the Range of a Linear Transformation (Range Method)

Let $L: \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation given by $\mathbf{L}(\mathbf{X}) = \mathbf{AX}$, for some $m \times n$ matrix $\mathbf{A}$. To find a basis for **range**(L), perform the following steps:

Step 1: Find $\mathbf{B}$, the reduced row echelon form of $\mathbf{A}$.

Step 2: Form the set of those columns of $\mathbf{A}$ whose corresponding columns in $\mathbf{B}$ have nonzero pivots. This set is a basis for **range**(L).

Example 5

Consider the linear transformation $L: \mathbb{R}^5 \to \mathbb{R}^4$ given in Example 4. After row reducing the matrix $\mathbf{A}$ for L, we obtained a matrix $\mathbf{B}$ in reduced row echelon form having nonzero pivots in columns $1, 3$, and 5. Hence, columns 1, 3, and 5 of $\mathbf{A}$ form a basis for **range**(L). In particular, we get the basis $\{[8,4,-2,6],\ [16,10,-5,15],\ [0,-4,7,-7]\}$, and so $\dim(\mathbf{range}(L)) = 3$. ■

From Examples 4 and 5, we see that $\dim(\ker(L)) + \dim(\text{range}(L)) = 2 + 3 = 5 = \dim(\mathbb{R}^5) = \dim(\text{domain}(L))$, for the given linear transformation L. We can understand why this works by examining our methods for calculating bases for the kernel and range. For $\ker(L)$, we get one basis vector for each independent variable, which corresponds to a nonpivot column of $\mathbf{A}$ after row reducing. For range(L), we get one basis vector for each pivot column of $\mathbf{A}$. Together, these account for the total number of columns of $\mathbf{A}$, which is the dimension of the domain.

The fact that the number of nonzero pivots of $\mathbf{A}$ equals the number of nonzero rows in the reduced row echelon form matrix for $\mathbf{A}$ shows that $\dim(\text{range}(L)) = \text{rank}(\mathbf{A})$. This result is stated in the following theorem, which also holds when bases other than the standard bases are used (see Exercise 17).

Theorem 5.9 If $L\colon \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation with matrix $\mathbf{A}$ with respect to any bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, then

(1) $\dim(\text{range}(L)) = \text{rank}(\mathbf{A})$
(2) $\dim(\ker(L)) = n - \text{rank}(\mathbf{A})$
(3) $\dim(\ker(L)) + \dim(\text{range}(L)) = \dim(\text{domain}(L)) = n$.

The Dimension Theorem

The result in part (3) of Theorem 5.9 generalizes to linear transformations between any vector spaces $\mathcal{V}$ and $\mathcal{W}$, as long as the dimension of the domain is finite. We state this important theorem here, but postpone its proof until after a discussion of isomorphism in Section 5.5. An alternate proof of the Dimension Theorem that does not involve the matrix of the linear transformation is outlined in Exercise 18 of this section.

Theorem 5.10 (Dimension Theorem) If $L\colon \mathcal{V} \to \mathcal{W}$ is a linear transformation and $\mathcal{V}$ is finite dimensional, then $\text{range}(L)$ is finite dimensional, and

$$\dim(\ker(L)) + \dim(\text{range}(L)) = \dim(\mathcal{V}).$$

We have already seen that for the linear transformation in Examples 4 and 5, the dimensions of the kernel and the range add up to the dimension of the domain, as the Dimension Theorem asserts. Notice the Dimension Theorem holds for the linear transformations in Examples 1 through 3 as well.

Example 6

Consider $L\colon \mathcal{M}_{nn} \to \mathcal{M}_{nn}$ given by $L(\mathbf{A}) = \mathbf{A} + \mathbf{A}^T$. Now, $\ker(L) = \{\mathbf{A} \in \mathcal{M}_{nn} \mid \mathbf{A} + \mathbf{A}^T = \mathbf{O}_n\}$. However, $\mathbf{A} + \mathbf{A}^T = \mathbf{O}_n$ implies that $\mathbf{A} = -\mathbf{A}^T$. Hence, $\ker(L)$ is precisely the set of all skew-symmetric $n \times n$ matrices.

The range of L is the set of all matrices $\mathbf{B}$ of the form $\mathbf{A} + \mathbf{A}^T$ for some $n \times n$ matrix $\mathbf{A}$. However, if $\mathbf{B} = \mathbf{A} + \mathbf{A}^T$, then $\mathbf{B}^T = \left(\mathbf{A} + \mathbf{A}^T\right)^T = \mathbf{A}^T + \mathbf{A} = \mathbf{B}$, so $\mathbf{B}$ is symmetric. Thus, $\text{range}(L) \subseteq \{\text{symmetric } n \times n \text{ matrices}\}$.

Next, if $\mathbf{B}$ is a symmetric $n \times n$ matrix, then $L(\frac{1}{2}\mathbf{B}) = \frac{1}{2}L(\mathbf{B}) = \frac{1}{2}(\mathbf{B} + \mathbf{B}^T) = \frac{1}{2}(\mathbf{B} + \mathbf{B}) = \mathbf{B}$, and so $\mathbf{B} \in \text{range}(L)$, thus proving $\{\text{symmetric } n \times n \text{ matrices}\} \subseteq \text{range}(L)$. Hence, $\text{range}(L)$ is the set of all symmetric $n \times n$ matrices.

In Exercise 12 of Section 4.6, we found that $\dim(\{\text{skew-symmetric } n \times n \text{ matrices}\}) = (n^2 - n)/2$ and that $\dim(\{\text{symmetric } n \times n \text{ matrices}\}) = \left(n^2 + n\right)/2$. Notice that the Dimension Theorem holds here, since $\dim(\ker(L)) + \dim(\text{range}(L)) = (n^2 - n)/2 + \left(n^2 + n\right)/2 = n^2 = \dim(\mathcal{M}_{nn})$. ∎

Rank of the Transpose

We can use the Range Method to prove the following result:[2]

Corollary 5.11 If $\mathbf{A}$ is any matrix, then $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^T)$.

Proof. Let $\mathbf{A}$ be an $m \times n$ matrix. Consider the linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^m$ with associated matrix $\mathbf{A}$ (using the standard bases). By the Range Method, $\text{range}(L)$ is the span of the column vectors of $\mathbf{A}$. Hence, $\text{range}(L)$ is the span of the row vectors of $\mathbf{A}^T$; that is, $\text{range}(L)$ is the row space of $\mathbf{A}^T$. Thus, $\dim(\text{range}(L)) = \text{rank}(\mathbf{A}^T)$, by the Simplified Span Method. But by Theorem 5.9, $\dim(\text{range}(L)) = \text{rank}(\mathbf{A})$. Hence, $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^T)$. □

Example 7

Let $\mathbf{A}$ be the matrix from Examples 4 and 5. We calculated its reduced row echelon form $\mathbf{B}$ in Example 4 and found it has three nonzero rows. Hence, $\text{rank}(\mathbf{A}) = 3$. Now,

$$\mathbf{A}^T = \begin{bmatrix} 8 & 4 & -2 & 6 \\ 4 & 2 & -1 & 3 \\ 16 & 10 & -5 & 15 \\ 32 & 22 & -11 & 33 \\ 0 & -4 & 7 & -7 \end{bmatrix} \quad \text{row reduces to} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \frac{7}{5} \\ 0 & 0 & 1 & -\frac{1}{5} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

showing that $\text{rank}(\mathbf{A}^T) = 3$ as well. ■

In some textbooks, rank(**A**) is called the **row rank** of **A** and rank($\mathbf{A}^T$) is called the **column rank** of **A**. Thus, Corollary 5.11 asserts that the row rank of **A** equals the column rank of **A**.

Recall that rank(**A**) = dim(row space of **A**). Analogous to the concept of row space, we define the **column space** of a matrix **A** as the span of the columns of **A**. In Corollary 5.11, we observed that if $L\colon \mathbb{R}^n \to \mathbb{R}^m$ with $L(\mathbf{X}) = \mathbf{AX}$ (using the standard bases), then range(L) = span({columns of **A**}) = column space of **A**, and so dim(range(L)) = dim(column space of **A**) = rank($\mathbf{A}^T$). With this new terminology, Corollary 5.11 asserts that dim(row space of **A**) = dim(column space of **A**). Be careful! This statement does not imply that these *spaces* are equal, only that their *dimensions* are equal. In fact, unless **A** is square, they contain vectors of different sizes. Notice that for the matrix **A** in Example 7, the row space of **A** is a subspace of $\mathbb{R}^5$, but the column space of **A** is a subspace of $\mathbb{R}^4$.

[2] In Exercise 18 of Section 4.6, you were asked to prove Corollary 5.11 by essentially the same method given here, only using different notation.

New Vocabulary

column rank (of a matrix)
column space (of a matrix)
Dimension Theorem
kernel (of a linear transformation)
Kernel Method
range (of a linear transformation)
Range Method
row rank (of a matrix)

Highlights

- The kernel of a linear transformation consists of all vectors of the domain that map to the zero vector of the codomain. The kernel is always a subspace of the domain.
- The range of a linear transformation consists of all vectors of the codomain that are images of vectors in the domain. The range is always a subspace of the codomain.
- If **A** is the matrix (with respect to any bases) for a linear transformation $L: \mathbb{R}^n \to \mathbb{R}^m$, then $\dim(\ker(L)) = n - \operatorname{rank}(\mathbf{A})$ and $\dim(\operatorname{range}(L)) = \operatorname{rank}(\mathbf{A})$.
- Kernel Method: A basis for the kernel of a linear transformation $L(\mathbf{X}) = \mathbf{AX}$ is obtained from the solution set of $\mathbf{BX} = \mathbf{O}$ by letting each independent variable in turn equal 1 and all other independent variables equal 0, where **B** is the reduced row echelon form of **A**.
- Range Method: A basis for the range of a linear transformation $L(\mathbf{X}) = \mathbf{AX}$ is obtained by selecting the columns of **A** corresponding to pivot columns in the reduced row echelon form matrix **B** for **A**.
- Dimension Theorem: If $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation and $\mathcal{V}$ is finite dimensional, then $\dim(\ker(L)) + \dim(\operatorname{range}(L)) = \dim(\mathcal{V})$.
- The rank of any matrix (= row rank) is equal to the rank of its transpose (= column rank).

EXERCISES FOR SECTION 5.3

1. Let $L: \mathbb{R}^3 \to \mathbb{R}^3$ be given by

$$L\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 5 & 1 & -1 \\ -3 & 0 & 1 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

★**(a)** Is $[1, -2, 3]$ in $\ker(L)$? Why or why not?

(b) Is $[2, -1, 4]$ in $\ker(L)$? Why or why not?

★**(c)** Is $[2, -1, 4]$ in $\operatorname{range}(L)$? Why or why not?

(d) Is $[-16, 12, -8]$ in $\operatorname{range}(L)$? Why or why not?

2. Let L: $\mathcal{P}_3 \to \mathcal{P}_3$ be given by $L(ax^3 + bx^2 + cx + d) = 2cx^3 + (a + b)x + (d + c)$.

★**(a)** Is $x^3 - 5x^2 + 3x - 6$ in $\ker(L)$? Why or why not?

(b) Is $4x^3 - 4x^2$ in $\ker(L)$? Why or why not?

★**(c)** Is $8x^3 - x - 1$ in range(L)? Why or why not?

(d) Is $4x^3 - 3x^2 + 7$ in range(L)? Why or why not?

3. For each of the following linear transformations $L: \mathcal{V} \to \mathcal{W}$, find a basis for $\ker(L)$ and a basis for range(L). Verify that $\dim(\ker(L)) + \dim(\text{range}(L)) = \dim(\mathcal{V})$.

★**(a)** $L: \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$L\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 1 & -1 & 5 \\ -2 & 3 & -13 \\ 3 & -3 & 15 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

(b) $L: \mathbb{R}^3 \to \mathbb{R}^4$ given by

$$L\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 4 & -2 & 8 \\ 7 & 1 & 5 \\ -2 & -1 & 0 \\ 3 & -2 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

(c) $L: \mathbb{R}^3 \to \mathbb{R}^2$ given by

$$L\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 3 & 2 & 11 \\ 2 & 1 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

★**(d)** $L: \mathbb{R}^4 \to \mathbb{R}^5$ given by

$$L\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}\right) = \begin{bmatrix} -14 & -8 & -10 & 2 \\ -4 & -1 & 1 & -2 \\ -6 & 2 & 12 & -10 \\ 3 & -7 & -24 & 17 \\ 4 & 2 & 2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

4. For each of the following linear transformations $L: \mathcal{V} \to \mathcal{W}$, find a basis for $\ker(L)$ and a basis for range(L), and verify that $\dim(\ker(L)) + \dim(\text{range}(L)) = \dim(\mathcal{V})$:

★**(a)** $L: \mathbb{R}^3 \to \mathbb{R}^2$ given by $L([x_1, x_2, x_3]) = [0, x_2]$

(b) $L: \mathbb{R}^2 \to \mathbb{R}^3$ given by $L([x_1, x_2]) = [x_1, x_1 + x_2, x_2]$

(c) $L: \mathcal{M}_{22} \to \mathcal{M}_{32}$ given by $L\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = \begin{bmatrix} 0 & -a_{12} \\ -a_{21} & 0 \\ 0 & 0 \end{bmatrix}$

⋆(d) $L: \mathcal{P}_4 \to \mathcal{P}_2$ given by $L(ax^4 + bx^3 + cx^2 + dx + e) = cx^2 + dx + e$

(e) $L: \mathcal{P}_2 \to \mathcal{P}_3$ given by $L(ax^2 + bx + c) = cx^3 + bx^2 + ax$

⋆(f) $L: \mathbb{R}^3 \to \mathbb{R}^3$ given by $L([x_1, x_2, x_3]) = [x_1, 0, x_1 - x_2 + x_3]$

⋆(g) $L: \mathcal{M}_{22} \to \mathcal{M}_{22}$ given by $L(\mathbf{A}) = \mathbf{A}^T$

(h) $L: \mathcal{M}_{33} \to \mathcal{M}_{33}$ given by $L(\mathbf{A}) = \mathbf{A} - \mathbf{A}^T$

⋆(i) $L: \mathcal{P}_2 \to \mathbb{R}^2$ given by $L(\mathbf{p}) = [\mathbf{p}(1), \mathbf{p}'(1)]$

(j) $L: \mathcal{P}_4 \to \mathbb{R}^3$ given by $L(\mathbf{p}) = [\mathbf{p}(-1), \mathbf{p}(0), \mathbf{p}(1)]$

5. (a) Suppose that $L: \mathcal{V} \to \mathcal{W}$ is the linear transformation given by $L(\mathbf{v}) = \mathbf{0}_{\mathcal{W}}$, for all $\mathbf{v} \in \mathcal{V}$. What is $\ker(L)$? What is $\text{range}(L)$?

(b) Suppose that $L: \mathcal{V} \to \mathcal{V}$ is the linear transformation given by $L(\mathbf{v}) = \mathbf{v}$, for all $\mathbf{v} \in \mathcal{V}$. What is $\ker(L)$? What is $\text{range}(L)$?

⋆6. Consider the mapping $L: \mathcal{M}_{33} \to \mathbb{R}$ given by $L(\mathbf{A}) = \text{trace}(\mathbf{A})$ (see Exercise 14 in Section 1.4). Show that L is a linear transformation. What is $\ker(L)$? What is $\text{range}(L)$? Calculate $\dim(\ker(L))$ and $\dim(\text{range}(L))$.

7. Let $\mathcal{V}$ be a vector space with fixed basis $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. Define $L: \mathcal{V} \to \mathcal{V}$ by $L(\mathbf{v}_1) = \mathbf{v}_2$, $L(\mathbf{v}_2) = \mathbf{v}_3, \ldots,$ $L(\mathbf{v}_{n-1}) = \mathbf{v}_n$, $L(\mathbf{v}_n) = \mathbf{v}_1$. Find $\text{range}(L)$. What is $\ker(L)$?

⋆8. Consider $L: \mathcal{P}_2 \to \mathcal{P}_4$ given by $L(\mathbf{p}) = x^2\mathbf{p}$. What is $\ker(L)$? What is $\text{range}(L)$? Verify that $\dim(\ker(L)) + \dim(\text{range}(L)) = \dim(\mathcal{P}_2)$.

9. Consider $L: \mathcal{P}_4 \to \mathcal{P}_2$ given by $L(\mathbf{p}) = \mathbf{p}''$. What is $\ker(L)$? What is $\text{range}(L)$? Verify that $\dim(\ker(L)) + \dim(\text{range}(L)) = \dim(\mathcal{P}_4)$.

⋆10. Consider $L: \mathcal{P}_n \to \mathcal{P}_n$ given by $L(\mathbf{p}) = \mathbf{p}^{(k)}$ (the kth derivative of $\mathbf{p}$), where $k \le n$. What is $\dim(\ker(L))$? What is $\dim(\text{range}(L))$? What happens when $k > n$?

11. Let a be a fixed real number. Consider $L: \mathcal{P}_n \to \mathbb{R}$ given by $L(\mathbf{p}(x)) = \mathbf{p}(a)$ (that is, the evaluation of $\mathbf{p}$ at $x = a$). (Recall from Exercise 18 in Section 5.1 that L is a linear transformation.) Show that $\{x - a, x^2 - a^2, \ldots, x^n - a^n\}$ is a basis for $\ker(L)$. (Hint: What is $\text{range}(L)$?)

⋆12. Suppose that $L: \mathbb{R}^n \to \mathbb{R}^n$ is a linear operator given by $L(\mathbf{X}) = \mathbf{AX}$, where $|\mathbf{A}| \neq 0$. What is $\ker(L)$? What is $\text{range}(L)$?

13. Let $\mathcal{V}$ be a finite dimensional vector space, and let $L: \mathcal{V} \to \mathcal{V}$ be a linear operator. Show that $\ker(L) = \{\mathbf{0}_{\mathcal{V}}\}$ if and only if $\text{range}(L) = \mathcal{V}$.

14. Let $L: \mathcal{V} \to \mathcal{W}$ be a linear transformation. Prove directly that $\ker(L)$ is a subspace of $\mathcal{V}$ and that $\text{range}(L)$ is a subspace of $\mathcal{W}$ using Theorem 4.2, that is, without invoking Theorem 5.8.

15. Let $L_1: \mathcal{V} \to \mathcal{W}$ and $L_2: \mathcal{W} \to \mathcal{X}$ be linear transformations.

(a) Show that $\ker(L_1) \subseteq \ker(L_2 \circ L_1)$.

(b) Show that $\text{range}(L_2 \circ L_1) \subseteq \text{range}(L_2)$.

(c) If $\mathcal{V}$ is finite dimensional, prove that $\dim(\text{range}(L_2 \circ L_1)) \leq \dim(\text{range}(L_1))$.

★16. Give an example of a linear operator $L: \mathbb{R}^2 \to \mathbb{R}^2$ such that $\ker(L) = \text{range}(L)$.

17. Let $L: \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with $m \times n$ matrix $\mathbf{A}$ for L with respect to the standard bases and $m \times n$ matrix $\mathbf{B}$ for L with respect to bases B and C.

(a) Prove that $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{B})$. (Hint: Use Exercise 16 in the Review Exercises of Chapter 2.)

(b) Use part (a) to finish the proof of Theorem 5.9. (Hint: Notice that Theorem 5.9 allows *any* bases to be used for $\mathbb{R}^n$ and $\mathbb{R}^m$. You can assume, from the remarks before Theorem 5.9, that the theorem is true when the standard bases are used for $\mathbb{R}^n$ and $\mathbb{R}^m$.)

18. This exercise outlines an alternate proof of the Dimension Theorem. Let $L: \mathcal{V} \to \mathcal{W}$ be a linear transformation with $\mathcal{V}$ finite dimensional. Figure 5.10 illustrates the relationships among the vectors referenced throughout this exercise.

(a) Let $\{\mathbf{k}_1, \ldots, \mathbf{k}_s\}$ be a basis for $\ker(L)$. Show that there exist vectors $\mathbf{q}_1, \ldots, \mathbf{q}_t$ such that $\{\mathbf{k}_1, \ldots, \mathbf{k}_s, \mathbf{q}_1, \ldots, \mathbf{q}_t\}$ is a basis for $\mathcal{V}$. Express $\dim(\mathcal{V})$ in terms of s and t.

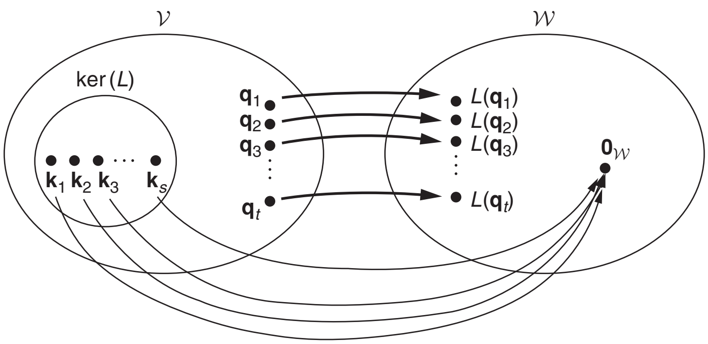


FIGURE 5.10

Images of basis elements in Exercise 18

(b) Use part (a) to show that for every $\mathbf{v} \in \mathcal{V}$, there exist scalars $b_1, \ldots, b_t$ such that $L(\mathbf{v}) = b_1 L(\mathbf{q}_1) + \cdots + b_t L(\mathbf{q}_t)$.

(c) Use part (b) to show that $\{L(\mathbf{q}_1), \ldots, L(\mathbf{q}_t)\}$ spans range(L). Conclude that $\dim(\text{range}(L)) \leq t$, and, hence, is finite.

(d) Suppose that $c_1 L(\mathbf{q}_1) + \cdots + c_t L(\mathbf{q}_t) = \mathbf{0}_{\mathcal{W}}$. Prove that $c_1\mathbf{q}_1 + \cdots + c_t\mathbf{q}_t \in \ker(L)$.

(e) Use part (d) to show that there are scalars $d_1, \ldots, d_s$ such that $c_1\mathbf{q}_1 + \cdots + c_t\mathbf{q}_t = d_1\mathbf{k}_1 + \cdots + d_s\mathbf{k}_s$.

(f) Use part (e) and the fact that $\{\mathbf{k}_1, \ldots, \mathbf{k}_s, \mathbf{q}_1, \ldots, \mathbf{q}_t\}$ is a basis for $\mathcal{V}$ to prove that $c_1 = c_2 = \cdots = c_t = d_1 = \cdots = d_s = 0$.

(g) Use parts (d) and (f) to conclude that $\{L(\mathbf{q}_1), \ldots, L(\mathbf{q}_t)\}$ is linearly independent.

(h) Use parts (c) and (g) to prove that $\{L(\mathbf{q}_1), \ldots, L(\mathbf{q}_t)\}$ is a basis for range(L).

(i) Conclude that $\dim(\ker(L)) + \dim(\text{range}(L)) = \dim(\mathcal{V})$.

19. Prove the following corollary of the Dimension Theorem: Let $L: \mathcal{V} \to \mathcal{W}$ be a linear transformation with $\mathcal{V}$ finite dimensional. Then $\dim(\ker(L)) \leq \dim(\mathcal{V})$ and $\dim(\text{range}(L)) \leq \dim(\mathcal{V})$.

★20. True or False:

(a) If $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation, then $\ker(L) = \{L(\mathbf{v}) \mid \mathbf{v} \in \mathcal{V}\}$.

(b) If $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation, then range(L) is a subspace of $\mathcal{V}$.

(c) If $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation and $\dim(\mathcal{V}) = n$, then $\dim(\ker(L)) = n - \dim(\text{range}(L))$.

(d) If $L: \mathcal{V} \to \mathcal{W}$ is a linear transformation and $\dim(\mathcal{V}) = 5$ and $\dim(\mathcal{W}) = 3$, then the Dimension Theorem implies that $\dim(\ker(L)) = 2$.

(e) If $L: \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation and $L(\mathbf{X}) = \mathbf{AX}$, then $\dim(\ker(L))$ equals the number of nonpivot columns in the reduced row echelon form matrix for $\mathbf{A}$.

(f) If $L: \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation and $L(\mathbf{X}) = \mathbf{AX}$, then $\dim(\text{range}(L)) = n - \text{rank}(\mathbf{A})$.

(g) If $\mathbf{A}$ is a 5×5 matrix, and $\text{rank}(\mathbf{A}) = 2$, then $\text{rank}\left(\mathbf{A}^T\right) = 3$.

(h) If $\mathbf{A}$ is any matrix, then the row space of $\mathbf{A}$ equals the column space of $\mathbf{A}$.

5.4 ONE-TO-ONE AND ONTO LINEAR TRANSFORMATIONS

The kernel and the range of a linear transformation are related to the function properties one-to-one and onto. Consequently, in this section we study linear transformations that are one-to-one or onto.

One-to-One and Onto Linear Transformations

One-to-one functions and onto functions are defined and discussed in Appendix B. In particular, Appendix B contains the usual methods for proving that a given function is, or is not, one-to-one or onto. Now, we are interested primarily in linear transformations, so we restate the definitions of *one-to-one* and *onto* specifically as they apply to this type of function.

Definition Let $L\colon \mathcal{V} \to \mathcal{W}$ be a linear transformation.

(1) L is **one-to-one** if and only if distinct vectors in $\mathcal{V}$ have different images in $\mathcal{W}$. That is, L is **one-to-one** if and only if, for all $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}$, $L(\mathbf{v}_1) = L(\mathbf{v}_2)$ implies $\mathbf{v}_1 = \mathbf{v}_2$.

(2) L is **onto** if and only if every vector in the codomain $\mathcal{W}$ is the image of some vector in the domain $\mathcal{V}$. That is, L is **onto** if and only if, for every $\mathbf{w} \in \mathcal{W}$, there is some $\mathbf{v} \in \mathcal{V}$ such that $L(\mathbf{v}) = \mathbf{w}$.

Notice that the two descriptions of a one-to-one linear transformation given in this definition are really contrapositives of each other.

Example 1

Rotation: Recall the rotation linear operator $L\colon \mathbb{R}^2 \to \mathbb{R}^2$ from Example 9 in Section 5.1 given by $L(\mathbf{v}) = \mathbf{A}\mathbf{v}$, where $\mathbf{A} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. We will show that L is *both one-to-one and onto.*

To show that L is one-to-one, we take any two arbitrary vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ in the domain $\mathbb{R}^2$, assume that $L(\mathbf{v}_1) = L(\mathbf{v}_2)$, and prove that $\mathbf{v}_1 = \mathbf{v}_2$. Now, if $L(\mathbf{v}_1) = L(\mathbf{v}_2)$, then $\mathbf{A}\mathbf{v}_1 = \mathbf{A}\mathbf{v}_2$. Because $\mathbf{A}$ is nonsingular, we can multiply both sides on the left by $\mathbf{A}^{-1}$ to obtain $\mathbf{v}_1 = \mathbf{v}_2$. Hence, L is one-to-one.

To show that L is onto, we must take any arbitrary vector $\mathbf{w}$ in the codomain $\mathbb{R}^2$ and show that there is some vector $\mathbf{v}$ in the domain $\mathbb{R}^2$ that maps to $\mathbf{w}$. Recall that multiplication by $\mathbf{A}^{-1}$ undoes the action of multiplication by $\mathbf{A}$, and so it must represent a *clockwise* rotation through the angle θ. Hence, we can find a pre-image for $\mathbf{w}$ by rotating it *clockwise* through the angle θ; that is, consider $\mathbf{v} = \mathbf{A}^{-1}\mathbf{w} \in \mathbb{R}^2$. When we apply L to $\mathbf{v}$, we rotate it *counterclockwise* through the same angle θ: $L(\mathbf{v}) = \mathbf{A}(\mathbf{A}^{-1}\mathbf{w}) = \mathbf{w}$, thus obtaining the original vector $\mathbf{w}$. Since $\mathbf{v}$ is in the domain and $\mathbf{v}$ maps to $\mathbf{w}$ under L, L is onto.

■

Example 2

Differentiation: Consider the linear transformation $L: \mathcal{P}_3 \to \mathcal{P}_2$ given by $L(\mathbf{p}) = \mathbf{p}'$. We will show that L is *onto but not one-to-one*.

To show that L is not one-to-one, we must find two different vectors $\mathbf{p}_1$ and $\mathbf{p}_2$ in the domain $\mathcal{P}_3$ that have the same image. Consider $\mathbf{p}_1 = x + 1$ and $\mathbf{p}_2 = x + 2$. Since $L(\mathbf{p}_1) = L(\mathbf{p}_2) = 1$, L is not one-to-one.

To show that L is onto, we must take an arbitrary vector $\mathbf{q}$ in $\mathcal{P}_2$ and find some vector $\mathbf{p}$ in $\mathcal{P}_3$ such that $L(\mathbf{p}) = \mathbf{q}$. Consider the vector $\mathbf{p} = \int \mathbf{q}(x)\,dx$ with zero constant term. Because $L(\mathbf{p}) = \mathbf{q}$, we see that L is onto.

If in Example 2 we had used $\mathcal{P}_3$ for the codomain instead of $\mathcal{P}_2$, the linear transformation would not have been onto because x^3 would have no pre-image (why?). This provides an example of a linear transformation that is neither one-to-one nor onto. Also, Exercise 6 illustrates a linear transformation that is one-to-one but not onto. These examples, together with Examples 1 and 2, show that the concepts of one-to-one and onto are independent of each other; that is, there are linear transformations that have either property with or without the other.

Theorem B.1 in Appendix B shows that the composition of one-to-one linear transformations is one-to-one, and similarly, the composition of onto linear transformations is onto.

Kernel and Range

The next theorem gives an alternate way of characterizing one-to-one linear transformations and onto linear transformations.

Theorem 5.12 Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces, and let $L: \mathcal{V} \to \mathcal{W}$ be a linear transformation. Then:

(1) L is one-to-one if and only if $\ker(L) = \{\mathbf{0}_{\mathcal{V}}\}$ (or, equivalently, if and only if $\dim(\ker(L)) = 0$), and

(2) If $\mathcal{W}$ is finite dimensional, then L is onto if and only if $\dim(\text{range}(L)) = \dim(\mathcal{W})$.

Thus, a linear transformation whose kernel contains a nonzero vector cannot be one-to-one.

Proof. First suppose that L is one-to-one, and let $\mathbf{v} \in \ker(L)$. We must show that $\mathbf{v} = \mathbf{0}_{\mathcal{V}}$. Now, $L(\mathbf{v}) = \mathbf{0}_{\mathcal{W}}$. However, by Theorem 5.1, $L(\mathbf{0}_{\mathcal{V}}) = \mathbf{0}_{\mathcal{W}}$. Because $L(\mathbf{v}) = L(\mathbf{0}_{\mathcal{V}})$ and L is one-to-one, we must have $\mathbf{v} = \mathbf{0}_{\mathcal{V}}$.

Conversely, suppose that $\ker(L) = \{\mathbf{0}_{\mathcal{V}}\}$. We must show that L is one-to-one. Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}$, with $L(\mathbf{v}_1) = L(\mathbf{v}_2)$. We must show that $\mathbf{v}_1 = \mathbf{v}_2$. Now, $L(\mathbf{v}_1) - L(\mathbf{v}_2) = \mathbf{0}_{\mathcal{W}}$, implying that $L(\mathbf{v}_1 - \mathbf{v}_2) = \mathbf{0}_{\mathcal{W}}$. Hence, $\mathbf{v}_1 - \mathbf{v}_2 \in \ker(L)$, by definition of the kernel. Since $\ker(L) = \{\mathbf{0}_{\mathcal{V}}\}$, $\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{0}_{\mathcal{V}}$ and so $\mathbf{v}_1 = \mathbf{v}_2$.

Finally, note that, by definition, L is onto if and only if $\mathbf{range}(L) = \mathcal{W}$, and therefore part (2) of the theorem follows immediately from Theorem 4.16. □

Example 3

Consider the linear transformation $L: \mathcal{M}_{22} \to \mathcal{M}_{23}$ given by $L\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} a-b & 0 & c-d \\ c+d & a+b & 0 \end{bmatrix}$. If $\begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \ker(L)$, then $a-b=c-d=c+d=a+b=0$. Solving these equations yields $a=b=c=d=0$, and so $\ker(L)$ contains only the zero matrix $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$; that is, $\dim(\ker(L)) = 0$. Thus, by part (1) of Theorem 5.12, L is one-to-one. However, by the Dimension Theorem, $\dim(\mathbf{range}(L)) = \dim(\mathcal{M}_{22}) - \dim(\ker(L)) = \dim(\mathcal{M}_{22}) = 4$. Hence, by part (2) of Theorem 5.12, L is not onto. In particular, $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \notin \mathbf{range}(L)$.

On the other hand, consider $M: \mathcal{M}_{23} \to \mathcal{M}_{22}$ given by $M\left(\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}\right) = \begin{bmatrix} a+b & a+c \\ d+e & d+f \end{bmatrix}$. It is easy to see that M is onto, since $M\left(\begin{bmatrix} 0 & b & c \\ 0 & e & f \end{bmatrix}\right) = \begin{bmatrix} b & c \\ e & f \end{bmatrix}$, and thus every 2×2 matrix is in $\mathbf{range}(M)$. Thus, by part (2) of Theorem 5.12, $\dim(\mathbf{range}(M)) = \dim(\mathcal{M}_{22}) = 4$. Then, by the Dimension Theorem, $\ker(M) = \dim(\mathcal{M}_{23}) - \dim(\mathbf{range}(M)) = 6 - 4 = 2$. Hence, by part (1) of Theorem 5.12, M is not one-to-one. In particular, $\begin{bmatrix} 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \in \ker(L)$.

■

Spanning and Linear Independence

The next theorem shows that the one-to-one property is related to linear independence, while the onto property is related to spanning.

Theorem 5.13 Let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces, and let $L: \mathcal{V} \to \mathcal{W}$ be a linear transformation. Then:

(1) If L is one-to-one, and T is a linearly independent subset of $\mathcal{V}$, then $L(T)$ is linearly independent in $\mathcal{W}$.

(2) If L is onto, and S spans $\mathcal{V}$, then $L(S)$ spans $\mathcal{W}$.

Proof. Suppose that L is one-to-one, and T is a linearly independent subset of $\mathcal{V}$. To prove that $L(T)$ is linearly independent in $\mathcal{W}$, it is enough to show that any finite subset of $L(T)$ is linearly independent. Suppose $\{L(\mathbf{x}_1), \ldots, L(\mathbf{x}_n)\}$ is a finite subset of $L(T)$, for vectors $\mathbf{x}_1,\ldots,\mathbf{x}_n \in T$, and suppose $b_1L(\mathbf{x}_1)+\cdots+b_nL(\mathbf{x}_n)=\mathbf{0}_{\mathcal{W}}$. Then, $L(b_1\mathbf{x}_1+\cdots+b_n\mathbf{x}_n)=\mathbf{0}_{\mathcal{W}}$, implying that $b_1\mathbf{x}_1+\cdots+b_n\mathbf{x}_n \in \ker(L)$. But since L is one-to-one, Theorem 5.12 tells us that $\ker(L)=\{\mathbf{0}_{\mathcal{V}}\}$. Hence, $b_1\mathbf{x}_1+\cdots+b_n\mathbf{x}_n=\mathbf{0}_{\mathcal{V}}$. Then, because the vectors in T are linearly independent, $b_1=b_2=\cdots=b_n=0$. Therefore, $\{L(\mathbf{x}_1),\ldots,L(\mathbf{x}_n)\}$ is linearly independent. Hence, $L(T)$ is linearly independent.

Now suppose that L is onto, and S spans $\mathcal{V}$. To prove that $L(S)$ spans $\mathcal{W}$, we must show that any vector $\mathbf{w}\in\mathcal{W}$ can be expressed as a linear combination of vectors in $L(S)$. Since L is onto, there is a $\mathbf{v}\in\mathcal{V}$ such that $L(\mathbf{v})=\mathbf{w}$. Since S spans $\mathcal{V}$, there are scalars $a_1,\ldots,a_n$ and vectors $\mathbf{v}_1,\ldots,\mathbf{v}_n\in S$ such that $\mathbf{v}=a_1\mathbf{v}_1+\cdots+a_n\mathbf{v}_n$. Thus, $\mathbf{w}=L(\mathbf{v})=L(a_1\mathbf{v}_1+\cdots+a_n\mathbf{v}_n)=a_1L(\mathbf{v}_1)+\cdots+a_nL(\mathbf{v}_n)$. Hence, $L(S)$ spans $\mathcal{W}$. □

An almost identical proof gives the following useful generalization of part (2) of Theorem 5.13: For any linear transformation $L:\mathcal{V}\to\mathcal{W}$, and any subset S of $\mathcal{V}$, $L(S)$ spans the subspace $L(\mathrm{span}(S))$ of $\mathcal{W}$. In particular, if S spans $\mathcal{V}$, then $L(S)$ spans $\mathrm{range}(L)$. (See Exercise 8.)

Example 4

Consider the linear transformation $L: P_2\to P_3$ given by $L(ax^2+bx+c)=bx^3+cx^2+ax$. It is easy to see that $\ker(L)=\{\mathbf{0}\}$ since $L(ax^2+bx+c)=0x^3+0x^2+0x+0$ only if $a=b=c=0$, and so L is one-to-one by Theorem 5.12. Consider the linearly independent set $T=\{x^2+x, x+1\}$ in P_2. Notice that $L(T)=\{x^3+x,\ x^3+x^2\}$, and that $L(T)$ is linearly independent, as predicted by part (1) of Theorem 5.13.

Next, let $\mathcal{W}=\{[x,0,z]\}$ be the xz-plane in $\mathbb{R}^3$. Clearly, $\dim(\mathcal{W})=2$. Consider $L:\mathbb{R}^3\to\mathcal{W}$, where L is the projection of $\mathbb{R}^3$ onto the xz-plane; that is, $L([x,y,z])=[x,0,z]$. It is easy to check that $S=\{[2,-1,3],[1,-2,0],[4,3,-1]\}$ spans $\mathbb{R}^3$ using the Simplified Span Method. Part (2) of Theorem 5.13 then asserts that $L(S)=\{[2,0,3],[1,0,0],[4,0,-1]\}$ spans $\mathcal{W}$. In fact, $\{[2,0,3],[1,0,0]\}$ alone spans $\mathcal{W}$, since $\dim(\mathrm{span}(\{[2,0,3],[1,0,0]\}))=2=\dim(\mathcal{W})$. ■

In Section 5.5, we will consider isomorphisms, which are linear transformations that are simultaneously one-to-one and onto. We will see that such functions faithfully carry vector space properties from the domain to the codomain.

New Vocabulary

one-to-one linear transformation | onto linear transformation

Highlights

- A linear transformation is one-to-one if no two distinct vectors of the domain map to the same image in the codomain.
- A linear transformation $L:\ \mathcal{V}\to\mathcal{W}$ is one-to-one if and only if $\ker(L)=\{\mathbf{0}_{\mathcal{V}}\}$ (or, equivalently, if and only if $\dim(\ker(L))=0$).
- If a linear transformation is one-to-one, then the image of every linearly independent subset of the domain is linearly independent.

- A linear transformation is onto if every vector in the codomain is the image of some vector from the domain.
- A linear transformation $L: \mathcal{V} \to \mathcal{W}$ is onto if and only if range$(L) = \mathcal{W}$ (or, equivalently, if and only if dim(range(L)) $=$ dim$(\mathcal{W})$ when $\mathcal{W}$ is finite dimensional).
- If a linear transformation is onto, then the image of every spanning set for the domain spans the codomain.

EXERCISES FOR SECTION 5.4

1. Which of the following linear transformations are one-to-one? Which are onto? Justify your answers without using row reduction.

★**(a)** $L: \mathbb{R}^3 \to \mathbb{R}^4$ given by $L([x,y,z]) = [y,z,-y,0]$

(b) $L: \mathbb{R}^3 \to \mathbb{R}^2$ given by $L([x,y,z]) = [x+y, y+z]$

★**(c)** $L: \mathbb{R}^3 \to \mathbb{R}^3$ given by $L([x,y,z]) = [2x, x+y+z, -y]$

(d) $L: \mathcal{P}_3 \to \mathcal{P}_2$ given by $L(ax^3 + bx^2 + cx + d) = ax^2 + bx + c$

★**(e)** $L: \mathcal{P}_2 \to \mathcal{P}_2$ given by $L(ax^2 + bx + c) = (a+b)x^2 + (b+c)x + (a+c)$

(f) $L: \mathcal{M}_{22} \to \mathcal{M}_{22}$ given by $L\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} d & b+c \\ b-c & a \end{bmatrix}$

★**(g)** $L: \mathcal{M}_{23} \to \mathcal{M}_{22}$ given by $L\left(\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}\right) = \begin{bmatrix} a & -c \\ 2e & d+f \end{bmatrix}$

★**(h)** $L: \mathcal{P}_2 \to \mathcal{M}_{22}$ given by $L(ax^2 + bx + c) = \begin{bmatrix} a+c & 0 \\ b-c & -3a \end{bmatrix}$

2. Which of the following linear transformations are one-to-one? Which are onto? Justify your answers by using row reduction to determine the dimensions of the kernel and range.

★**(a)** $L: \mathbb{R}^2 \to \mathbb{R}^2$ given by $L\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} -4 & -3 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

★**(b)** $L: \mathbb{R}^2 \to \mathbb{R}^3$ given by $L\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} -3 & 4 \\ -6 & 9 \\ 7 & -8 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

★**(c)** $L: \mathbb{R}^3 \to \mathbb{R}^3$ given by $L\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} -7 & 4 & -2 \\ 16 & -7 & 2 \\ 4 & -3 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

(d) $L\colon \mathbb{R}^4 \to \mathbb{R}^3$ given by $L\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}\right) = \begin{bmatrix} -5 & 3 & 1 & 18 \\ -2 & 1 & 1 & 6 \\ -7 & 3 & 4 & 19 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$

3. In each of the following cases, the matrix for a linear transformation with respect to some ordered bases for the domain and codomain is given. Which of these linear transformations are one-to-one? Which are onto? Justify your answers by using row reduction to determine the dimensions of the kernel and range.

★**(a)** $L\colon \mathcal{P}_2 \to \mathcal{P}_2$ having matrix $\begin{bmatrix} 1 & -3 & 0 \\ -4 & 13 & -1 \\ 8 & -25 & 2 \end{bmatrix}$

(b) $L\colon \mathcal{M}_{22} \to \mathcal{M}_{22}$ having matrix $\begin{bmatrix} 6 & -9 & 2 & 8 \\ 10 & -6 & 12 & 4 \\ -3 & 3 & -4 & -4 \\ 8 & -9 & 9 & 11 \end{bmatrix}$

★**(c)** $L\colon \mathcal{M}_{22} \to \mathcal{P}_3$ having matrix $\begin{bmatrix} 2 & 3 & -1 & 1 \\ 5 & 2 & -4 & 7 \\ 1 & 7 & 1 & -4 \\ -2 & 19 & 7 & -19 \end{bmatrix}$

4. Suppose that $m > n$.

(a) Show there is no onto linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$.

(b) Show there is no one-to-one linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$.

5. Let $\mathbf{A}$ be a fixed $n \times n$ matrix, and consider $L\colon \mathcal{M}_{nn} \to \mathcal{M}_{nn}$ given by $L(\mathbf{B}) = \mathbf{AB} - \mathbf{BA}$.

(a) Show that L is not one-to-one. (Hint: Consider $L(\mathbf{I}_n)$.)

(b) Use part (a) to show that L is not onto.

6. Define $L\colon \mathcal{U}_3 \to \mathcal{M}_{33}$ by $L(\mathbf{A}) = \frac{1}{2}(\mathbf{A} + \mathbf{A}^T)$. Prove that L is one-to-one but is *not* onto.

7. Let $L\colon \mathcal{V} \to \mathcal{W}$ be a linear transformation between vector spaces. Suppose that for every linearly independent set T in $\mathcal{V}$, $L(T)$ is linearly independent in $\mathcal{W}$. Prove that L is one-to-one. (Hint: Prove $\ker(L) = \{\mathbf{0}_\mathcal{V}\}$ using a proof by contradiction.)

8. Let $L\colon \mathcal{V} \to \mathcal{W}$ be a linear transformation between vector spaces, and let S be a subset of $\mathcal{V}$.

(a) Prove that $L(S)$ spans the subspace $L(\text{span}(S))$.

(b) Show that if S spans $\mathcal{V}$, then $L(S)$ spans range(L).

(c) Show that if $L(S)$ spans $\mathcal{W}$, then L is onto.

★9. True or False:

(a) A linear transformation $L: \mathcal{V} \to \mathcal{W}$ is one-to-one if for all $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}, \mathbf{v}_1 = \mathbf{v}_2$ implies $L(\mathbf{v}_1) = L(\mathbf{v}_2)$.

(b) A linear transformation $L: \mathcal{V} \to \mathcal{W}$ is onto if for all $\mathbf{v} \in \mathcal{V}$, there is some $\mathbf{w} \in \mathcal{W}$ such that $L(\mathbf{v}) = \mathbf{w}$.

(c) A linear transformation $L: \mathcal{V} \to \mathcal{W}$ is one-to-one if $\ker(L)$ contains no vectors other than $\mathbf{0}_{\mathcal{V}}$.

(d) If L is a linear transformation and S spans the domain of L, then $L(S)$ spans the range of L.

(e) Suppose $\mathcal{V}$ is a finite dimensional vector space. A linear transformation $L: \mathcal{V} \to \mathcal{W}$ is not one-to-one if $\dim(\ker(L)) \neq 0$.

(f) Suppose $\mathcal{W}$ is a finite dimensional vector space. A linear transformation $L: \mathcal{V} \to \mathcal{W}$ is not onto if $\dim(\text{range}(L)) < \dim(\mathcal{W})$.

(f) If L is a linear transformation and T is a linearly independent subset of the domain of L, then $L(T)$ is linearly independent.

(g) If L is a linear transformation $L: \mathcal{V} \to \mathcal{W}$, and S is a subset of $\mathcal{V}$ such that $L(S)$ spans $\mathcal{W}$, then S spans $\mathcal{V}$.

5.5 ISOMORPHISM

In this section, we examine methods for determining whether two vector spaces are equivalent, or *isomorphic*. Isomorphism is important because if certain algebraic results are true in one of two isomorphic vector spaces, corresponding results hold true in the other as well. It is the concept of isomorphism that has allowed us to apply our techniques and formal methods to vector spaces other than $\mathbb{R}^n$.

Isomorphisms: Invertible Linear Transformations

We restate here the definition from Appendix B for the inverse of a function as it applies to linear transformations.

> **Definition** Let $L: \mathcal{V} \to \mathcal{W}$ be a linear transformation. Then L is an **invertible linear transformation** if and only if there is a function $M: \mathcal{W} \to \mathcal{V}$ such that $(M \circ L)(\mathbf{v}) = \mathbf{v}$, for all $\mathbf{v} \in \mathcal{V}$, and $(L \circ M)(\mathbf{w}) = \mathbf{w}$, for all $\mathbf{w} \in \mathcal{W}$. Such a function M is called an **inverse** of L.

7.5 INNER PRODUCT SPACES

Prerequisite: Section 6.3, Orthogonal Diagonalization

In $\mathbb{R}^n$ and $\mathbb{C}^n$, we have the dot product along with the operations of vector addition and scalar multiplication. In other vector spaces, we can often create a similar type of product, known as an inner product.

Inner Products

Definition Let $\mathcal{V}$ be a real [complex] vector space with operations $+$ and $\cdot$, and let $\langle\,,\rangle$ be an operation that assigns to each pair of vectors $\mathbf{x}, \mathbf{y} \in \mathcal{V}$ a real [complex] number, denoted $\langle \mathbf{x}, \mathbf{y}\rangle$. Then $\langle\,,\rangle$ is a **real [complex] inner product** for $\mathcal{V}$ if and only if the following properties hold for all $\mathbf{x}, \mathbf{y} \in \mathcal{V}$ and all $k \in \mathbb{R}$ [$k \in \mathbb{C}$]:

(1) $\langle \mathbf{x}, \mathbf{x}\rangle$ is always real, and $\langle \mathbf{x}, \mathbf{x}\rangle \geq 0$

(2) $\langle \mathbf{x}, \mathbf{x}\rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$

(3) $\langle \mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{y}, \mathbf{x}\rangle$ $\left[\langle \mathbf{x}, \mathbf{y}\rangle = \overline{\langle \mathbf{y}, \mathbf{x}\rangle}\right]$

(4) $\langle \mathbf{x}+\mathbf{y}, \mathbf{z}\rangle = \langle \mathbf{x}, \mathbf{z}\rangle + \langle \mathbf{y}, \mathbf{z}\rangle$

(5) $\langle k\mathbf{x}, \mathbf{y}\rangle = k\langle \mathbf{x}, \mathbf{y}\rangle$.

A vector space together with a real [complex] inner product operation is known as a **real [complex] inner product space**.

Example 1

Consider the real vector space $\mathbb{R}^n$. Let $\mathbf{x} = [x_1, \ldots, x_n]$ and $\mathbf{y} = [y_1, \ldots, y_n]$ be vectors in $\mathbb{R}^n$. By Theorem 1.5, the operation $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}\cdot\mathbf{y} = x_1y_1 + \cdots + x_ny_n$ (usual real dot product) is a real inner product (verify!). Hence, $\mathbb{R}^n$ together with the dot product is a real inner product space.

Similarly, let $\mathbf{x} = [x_1, \ldots, x_n]$ and $\mathbf{y} = [y_1, \ldots, y_n]$ be vectors in the complex vector space $\mathbb{C}^n$. By Theorem 7.1, the operation $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}\cdot\mathbf{y} = x_1\overline{y_1} + \cdots + x_n\overline{y_n}$ (usual complex dot product) is an inner product on $\mathbb{C}^n$. Thus, $\mathbb{C}^n$ together with the complex dot product is a complex inner product space. ■

Example 2

Consider the real vector space $\mathbb{R}^2$. For $\mathbf{x} = [x_1, x_2]$ and $\mathbf{y} = [y_1, y_2]$ in $\mathbb{R}^2$, define $\langle \mathbf{x}, \mathbf{y}\rangle = x_1y_1 - x_1y_2 - x_2y_1 + 2x_2y_2$. We verify the five properties in the definition of an inner product space.

Property (1): $\langle \mathbf{x}, \mathbf{x}\rangle = x_1x_1 - x_1x_2 - x_2x_1 + 2x_2x_2 = x_1^2 - 2x_1x_2 + x_2^2 + x_2^2 = (x_1 - x_2)^2 + x_2^2 \geq 0$.

Property (2): $\langle \mathbf{x}, \mathbf{x}\rangle = 0$ exactly when $x_1 = x_2 = 0$ (that is, when $\mathbf{x} = \mathbf{0}$).

Property (3): $\langle \mathbf{y}, \mathbf{x} \rangle = y_1x_1 - y_1x_2 - y_2x_1 + 2y_2x_2 = x_1y_1 - x_1y_2 - x_2y_1 + 2x_2y_2 = \langle \mathbf{x}, \mathbf{y} \rangle$.

Property (4): Let $\mathbf{z} = [z_1, z_2]$. Then

$$\begin{aligned}\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle &= (x_1 + y_1)z_1 - (x_1 + y_1)z_2 - (x_2 + y_2)z_1 + 2(x_2 + y_2)z_2 \\ &= x_1z_1 + y_1z_1 - x_1z_2 - y_1z_2 - x_2z_1 - y_2z_1 + 2x_2z_2 + 2y_2z_2 \\ &= (x_1z_1 - x_1z_2 - x_2z_1 + 2x_2z_2) + (y_1z_1 - y_1z_2 - y_2z_1 + 2y_2z_2) \\ &= \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle .\end{aligned}$$

Property (5): $\langle k\mathbf{x}, \mathbf{y} \rangle = (kx_1)y_1 - (kx_1)y_2 - (kx_2)y_1 + 2(kx_2)y_2 = k(x_1y_1 - x_1y_2 - x_2y_1 + 2x_2y_2) = k\langle \mathbf{x}, \mathbf{y} \rangle$.

Hence, $\langle\, , \rangle$ is a real inner product on $\mathbb{R}^2$, and $\mathbb{R}^2$ together with this operation $\langle\, , \rangle$ is a real inner product space. ∎

Example 3

Consider the real vector space $\mathbb{R}^n$. Let $\mathbf{A}$ be a nonsingular $n \times n$ real matrix. Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and define $\langle \mathbf{x}, \mathbf{y} \rangle = (\mathbf{Ax}) \cdot (\mathbf{Ay})$ (the usual dot product of $\mathbf{Ax}$ and $\mathbf{Ay}$). It can be shown (see Exercise 1) that $\langle\, , \rangle$ is a real inner product on $\mathbb{R}^n$, and so $\mathbb{R}^n$ together with this operation $\langle\, , \rangle$ is a real inner product space. ∎

Example 4

Consider the real vector space $\mathcal{P}_n$. Let $\mathbf{p}_1 = a_nx^n + \cdots + a_1x + a_0$ and $\mathbf{p}_2 = b_nx^n + \cdots + b_1x + b_0$ be in $\mathcal{P}_n$. Define $\langle \mathbf{p}_1, \mathbf{p}_2 \rangle = a_nb_n + \cdots + a_1b_1 + a_0b_0$. It can be shown (see Exercise 2) that $\langle\, , \rangle$ is a real inner product on $\mathcal{P}_n$, and so $\mathcal{P}_n$ together with this operation $\langle\, , \rangle$ is a real inner product space. ∎

Example 5

Let $a, b \in \mathbb{R}$, with $a < b$, and consider the real vector space $\mathcal{V}$ of all real-valued continuous functions defined on the interval $[a, b]$ (for example, polynomials, $\sin x$, e^x). Let $\mathbf{f}, \mathbf{g} \in \mathcal{V}$. Define $\langle \mathbf{f}, \mathbf{g} \rangle = \int_a^b \mathbf{f}(t)\mathbf{g}(t)\,dt$. It can be shown (see Exercise 3) that $\langle\, , \rangle$ is a real inner product on $\mathcal{V}$, and so $\mathcal{V}$ together with this operation $\langle\, , \rangle$ is a real inner product space.

Analogously, the operation $\langle \mathbf{f}, \mathbf{g} \rangle = \int_a^b \mathbf{f}(t)\overline{\mathbf{g}(t)}\,dt$ makes the complex vector space of all complex-valued continuous functions on $[a, b]$ into a complex inner product space. ∎

Of course, not every operation is an inner product. For example, for the vectors $\mathbf{x} = [x_1, x_2]$ and $\mathbf{y} = [y_1, y_2]$ in $\mathbb{R}^2$, consider the operation $\langle \mathbf{x}, \mathbf{y} \rangle = x_1^2 + y_1^2$. Now, with $\mathbf{x} = \mathbf{y} = [1, 0]$, we have $\langle 2\mathbf{x}, \mathbf{y} \rangle = 2^2 + 1^2 = 5$, but $2\langle \mathbf{x}, \mathbf{y} \rangle = 2(1^2 + 1^2) = 4$. Thus, property (5) fails to hold.

The next theorem lists some useful results for inner product spaces.

Theorem 7.12 Let $\mathcal{V}$ be a real [complex] inner product space with inner product $\langle\,,\,\rangle$. Then for all $\mathbf{x},\mathbf{y} \in \mathcal{V}$ and all $k \in \mathbb{R}\,[k \in \mathbb{C}]$, we have

(1) $\langle \mathbf{0},\mathbf{x}\rangle = \langle \mathbf{x},\mathbf{0}\rangle = 0$
(2) $\langle \mathbf{x},\mathbf{y}+\mathbf{z}\rangle = \langle \mathbf{x},\mathbf{y}\rangle + \langle \mathbf{x},\mathbf{z}\rangle$
(3) $\langle \mathbf{x},k\mathbf{y}\rangle = k\,\langle \mathbf{x},\mathbf{y}\rangle \quad [\langle \mathbf{x},k\mathbf{y}\rangle = \overline{k}\,\langle \mathbf{x},\mathbf{y}\rangle]$.

Note the use of $\overline{k}$ in part (3) for complex vector spaces. The proof of this theorem is straightforward, and parts are left for you to do in Exercise 5.

Length, Distance, and Angles in Inner Product Spaces

The next definition extends the concept of the length of a vector to any inner product space.

Definition If $\mathbf{x}$ is a vector in an inner product space, then the **norm (length)** of $\mathbf{x}$ is $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x},\mathbf{x}\rangle}$.

This definition yields a nonnegative real number for $\|\mathbf{x}\|$, since by definition, $\langle \mathbf{x},\mathbf{x}\rangle$ is always real and nonnegative for any vector $\mathbf{x}$. Also note that this definition agrees with the earlier definition of length in $\mathbb{R}^n$ based on the usual dot product in $\mathbb{R}^n$. We also have the following result:

Theorem 7.13 Let $\mathcal{V}$ be a real [complex] inner product space, with $\mathbf{x} \in \mathcal{V}$. Let $k \in \mathbb{R}$ $[k \in \mathbb{C}]$. Then, $\|k\mathbf{x}\| = |k|\,\|\mathbf{x}\|$.

The proof of this theorem is left for you to do in Exercise 6.

As before, we say that a vector of length 1 in an inner product space is a **unit vector**. For instance, in the inner product space of Example 4, the polynomial $\mathbf{p} = \frac{\sqrt{2}}{2}x + \frac{\sqrt{2}}{2}$ is a unit vector since $\|\mathbf{p}\| = \sqrt{\langle \mathbf{p},\mathbf{p}\rangle} = \sqrt{\left(\frac{\sqrt{2}}{2}\right)^2 + \left(\frac{\sqrt{2}}{2}\right)^2} = 1$.

We define the distance between two vectors in the general inner product space setting as we did for $\mathbb{R}^n$.

Definition Let $\mathbf{x},\mathbf{y} \in \mathcal{V}$, an inner product space. Then the **distance between $\mathbf{x}$ and $\mathbf{y}$** is $\|\mathbf{x}-\mathbf{y}\|$.

Example 6

Consider the real vector space $\mathcal{V}$ of real continuous functions from Example 5, with $a = 0$ and $b = \pi$. That is, $\langle \mathbf{f},\mathbf{g}\rangle = \int_0^{\pi} \mathbf{f}(t)\mathbf{g}(t)\,dt$ for all $\mathbf{f},\mathbf{g} \in \mathcal{V}$. Let $\mathbf{f} = \cos t$ and $\mathbf{g} = \sin t$. Then the distance between $\mathbf{f}$ and $\mathbf{g}$ is

$$\|\mathbf{f}-\mathbf{g}\| = \sqrt{\langle \cos t - \sin t,\ \cos t - \sin t\rangle} = \sqrt{\int_0^{\pi} (\cos t - \sin t)^2\, dt}$$
$$= \sqrt{\int_0^{\pi} \left(\cos^2 t - 2\cos t \sin t + \sin^2 t\right)\, dt}$$
$$= \sqrt{\int_0^{\pi} (1-\sin 2t)\, dt} = \sqrt{\left(t + \frac{1}{2}\cos 2t\right)\Big|_0^{\pi}} = \sqrt{\pi}.$$

Hence, the distance between $\cos t$ and $\sin t$ is $\sqrt{\pi}$ under this inner product. ■

The next theorem shows that some other familiar results from the ordinary dot product carry over to the general inner product.

Theorem 7.14 Let $\mathbf{x}, \mathbf{y} \in \mathcal{V}$, an inner product space, with inner product $\langle\, ,\rangle$. Then

(1) $|\langle \mathbf{x}, \mathbf{y}\rangle| \le \|\mathbf{x}\|\,\|\mathbf{y}\|$ — Cauchy-Schwarz Inequality

(2) $\|\mathbf{x}+\mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|$. — Triangle Inequality

The proofs of these statements are analogous to the proofs for the ordinary dot product and are left for you to do in Exercise 11.

From the Cauchy-Schwarz Inequality, we have $-1 \le \langle \mathbf{x}, \mathbf{y}\rangle / \left(\|\mathbf{x}\|\,\|\mathbf{y}\|\right) \le 1$, for any nonzero vectors $\mathbf{x}$ and $\mathbf{y}$ in a *real* inner product space. Hence, we can make the following definition:

Definition Let $\mathbf{x}, \mathbf{y} \in \mathcal{V}$, a *real* inner product space. Then the **angle between x and y** is the angle θ from 0 to π such that $\cos\theta = \langle \mathbf{x}, \mathbf{y}\rangle / \left(\|\mathbf{x}\|\,\|\mathbf{y}\|\right)$.

Example 7

Consider again the inner product space of Example 6, where $\langle \mathbf{f}, \mathbf{g}\rangle = \int_0^{\pi} \mathbf{f}(t)\mathbf{g}(t)\,dt$. Let $\mathbf{f} = t$ and $\mathbf{g} = \sin t$. Then $\langle \mathbf{f}, \mathbf{g}\rangle = \int_0^{\pi} t \sin t\, dt$. Using integration by parts, we get $(-t\cos t)|_0^{\pi} + \int_0^{\pi} \cos t\, dt = \pi + (\sin t)|_0^{\pi} = \pi$. Also, $\|\mathbf{f}\|^2 = \langle \mathbf{f}, \mathbf{f}\rangle = \int_0^{\pi} (\mathbf{f}(t))^2\, dt = \int_0^{\pi} t^2\, dt = \left(t^3/3\right)|_0^{\pi} = \pi^3/3$, and so $\|\mathbf{f}\| = \sqrt{\pi^3/3}$. Similarly, $\|\mathbf{g}\|^2 = \langle \mathbf{g}, \mathbf{g}\rangle = \int_0^{\pi} \left(\mathbf{g}(t)\right)^2 dt = \int_0^{\pi} \sin^2 t\, dt = \int_0^{\pi} \frac{1}{2}(1-\cos 2t)\, dt = \left(\frac{1}{2}t - \frac{1}{4}\sin 2t\right)\Big|_0^{\pi} = \pi/2$, and so $\|\mathbf{g}\| = \sqrt{\pi/2}$. Hence, the cosine of the angle θ between t and $\sin t$ equals $\langle \mathbf{f}, \mathbf{g}\rangle / \left(\|\mathbf{f}\|\,\|\mathbf{g}\|\right) = \pi / \left(\sqrt{\pi^3/3}\sqrt{\pi/2}\right) = \sqrt{6}/\pi \approx 0.78$. Hence, $\theta \approx 0.68$ radians (38.8°). ■

Orthogonality in Inner Product Spaces

We next define orthogonal vectors in a general inner product space setting and show that nonzero orthogonal vectors are linearly independent.

Definition A subset $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ of vectors in an inner product space $\mathcal{V}$ with inner product $\langle\,,\rangle$ is **orthogonal** if and only if $\langle \mathbf{x}_i, \mathbf{x}_j \rangle = 0$ for $1 \le i, j \le n$, with $i \ne j$. Also, an orthogonal set of vectors in $\mathcal{V}$ is **orthonormal** if and only if each vector in the set is a unit vector.

The next theorem is the analog of Theorem 6.1, and its proof is left for you to do in Exercise 15.

Theorem 7.15 If $\mathcal{V}$ is an inner product space and T is an orthogonal set of nonzero vectors in $\mathcal{V}$, then T is a linearly independent set.

Example 8

Consider again the inner product space $\mathcal{V}$ of Example 5 of real continuous functions with inner product $\langle \mathbf{f}, \mathbf{g} \rangle = \int_a^b \mathbf{f}(t)\mathbf{g}(t)\,dt$, with $a = -\pi$ and $b = \pi$. The set $\{1, \cos t, \sin t\}$ is an orthogonal set in $\mathcal{V}$, since each of the following definite integrals equals zero (verify!):

$$\int_{-\pi}^{\pi} (1)\cos t\,dt, \quad \int_{-\pi}^{\pi} (1)\sin t\,dt, \quad \int_{-\pi}^{\pi} (\cos t)(\sin t)\,dt.$$

Also, note that $\|1\|^2 = \langle 1, 1 \rangle = \int_{-\pi}^{\pi} (1)(1)\,dt = 2\pi$, $\|\cos t\|^2 = \langle \cos t, \cos t \rangle = \int_{-\pi}^{\pi} \cos^2 t\,dt = \pi$ (why?), and $\|\sin t\|^2 = \langle \sin t, \sin t \rangle = \int_{-\pi}^{\pi} \sin^2 t\,dt = \pi$ (why?). Therefore, the set

$$\left\{ \frac{1}{\sqrt{2\pi}}, \frac{\cos t}{\sqrt{\pi}}, \frac{\sin t}{\sqrt{\pi}} \right\}$$

is an orthonormal set in $\mathcal{V}$.

Example 8 can be generalized. The set $\{1, \cos t, \sin t, \cos 2t, \sin 2t, \cos 3t, \sin 3t, \ldots\}$ is an orthogonal set (see Exercise 16) and therefore linearly independent by Theorem 7.15. The functions in this set are important in the theory of partial differential equations. It can be shown that every continuously differentiable function on the interval $[-\pi, \pi]$ can be represented as the (infinite) sum of constant multiples of these functions. Such a sum is known as the **Fourier series** of the function.

A basis for an inner product space $\mathcal{V}$ is an **orthogonal** [**orthonormal**] **basis** if the vectors in the basis form an orthogonal [orthonormal] set.

Example 9

Consider again the inner product space $\mathcal{P}_n$ with the inner product of Example 4; that is, if $\mathbf{p}_1 = a_n x^n + \cdots + a_1 x + a_0$ and $\mathbf{p}_2 = b_n x^n + \cdots + b_1 x + b_0$ are in $\mathcal{P}_n$, then $\langle \mathbf{p}_1, \mathbf{p}_2 \rangle = a_n b_n + \cdots + a_1 b_1 + a_0 b_0$. Now, $\{x^n, x^{n-1}, \ldots, x, 1\}$ is an orthogonal basis for $\mathcal{P}_n$ with this inner product, since $\langle x^k, x^l\rangle = 0$, for $0 \le k, l \le n$, with $k \ne l$ (why?). Since $\|x^k\| = \sqrt{\langle x^k, x^k\rangle} = 1$, for all k, $0 \le k \le n$ (why?), the set $\{x^n, x^{n-1}, \ldots, x, 1\}$ is also an orthonormal basis for this inner product space. ■

A proof analogous to that of Theorem 6.3 gives us the next theorem (see Exercise 17).

Theorem 7.16 If $B = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$ is an orthogonal ordered basis for a subspace $\mathcal{W}$ of an inner product space $\mathcal{V}$, and if $\mathbf{v}$ is any vector in $\mathcal{W}$, then

$$[\mathbf{v}]_B = \left[\frac{\langle \mathbf{v}, \mathbf{v}_1\rangle}{\langle \mathbf{v}_1, \mathbf{v}_1\rangle}, \frac{\langle \mathbf{v}, \mathbf{v}_2\rangle}{\langle \mathbf{v}_2, \mathbf{v}_2\rangle}, \ldots, \frac{\langle \mathbf{v}, \mathbf{v}_k\rangle}{\langle \mathbf{v}_k, \mathbf{v}_k\rangle}\right].$$

In particular, if B is an orthonormal ordered basis for $\mathcal{W}$, then $[\mathbf{v}]_B = [\langle \mathbf{v}, \mathbf{v}_1\rangle, \langle \mathbf{v}, \mathbf{v}_2\rangle, \ldots, \langle \mathbf{v}, \mathbf{v}_k\rangle]$.

Example 10

Recall the inner product space $\mathbb{R}^2$ in Example 2, with inner product given as follows: if $\mathbf{x} = [x_1, x_2]$ and $\mathbf{y} = [y_1, y_2]$, then $\langle \mathbf{x}, \mathbf{y}\rangle = x_1y_1 - x_1y_2 - x_2y_1 + 2x_2y_2$. An ordered orthogonal basis for this space is $B = (\mathbf{v}_1, \mathbf{v}_2) = ([2,1],[0,1])$ (verify!). Recall from Example 2 that $\langle \mathbf{x}, \mathbf{x}\rangle = (x_1 - x_2)^2 + x_2^2$. Thus, $\langle \mathbf{v}_1, \mathbf{v}_1\rangle = (2-1)^2 + 1^2 = 2$, and $\langle \mathbf{v}_2, \mathbf{v}_2\rangle = (0-1)^2 + 1^2 = 2$.

Next, suppose that $\mathbf{v} = [a, b]$ is any vector in $\mathbb{R}^2$. Now, $\langle \mathbf{v}, \mathbf{v}_1\rangle = \langle [a,b],[2,1]\rangle = (a)(2) - (a)(1) - (b)(2) + 2(b)(1) = a$. Also, $\langle \mathbf{v}, \mathbf{v}_2\rangle = \langle [a,b],[0,1]\rangle = (a)(0) - (a)(1) - (b)(0) + 2(b)(1) = -a + 2b$. Then,

$$[\mathbf{v}]_B = \left[\frac{\langle \mathbf{v}, \mathbf{v}_1\rangle}{\langle \mathbf{v}_1, \mathbf{v}_1\rangle}, \frac{\langle \mathbf{v}, \mathbf{v}_2\rangle}{\langle \mathbf{v}_2, \mathbf{v}_2\rangle}\right] = \left[\frac{a}{2}, \frac{-a+2b}{2}\right].$$

Notice that $\frac{a}{2}[2,1] + \left(\frac{-a+2b}{2}\right)[0,1]$ does equal $[a,b] = \mathbf{v}$. ■

The Generalized Gram-Schmidt Process

We can generalize the Gram-Schmidt Process of Section 6.1 to any inner product space. That is, we can replace any linearly independent set of k vectors with an orthogonal set of k vectors that spans the same subspace.

Generalized Gram-Schmidt Process

Let $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ be a linearly independent subset of an inner product space $\mathcal{V}$, with inner product $\langle\,,\rangle$. We create a new set $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ of vectors as follows:

Let $\mathbf{v}_1 = \mathbf{w}_1$.

Let $\mathbf{v}_2 = \mathbf{w}_2 - \left(\frac{\langle \mathbf{w}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\right) \mathbf{v}_1$.

Let $\mathbf{v}_3 = \mathbf{w}_3 - \left(\frac{\langle \mathbf{w}_3, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\right) \mathbf{v}_1 - \left(\frac{\langle \mathbf{w}_3, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\right) \mathbf{v}_2$.

$\vdots$

Let $\mathbf{v}_k = \mathbf{w}_k - \left(\frac{\langle \mathbf{w}_k, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\right) \mathbf{v}_1 - \left(\frac{\langle \mathbf{w}_k, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\right) \mathbf{v}_2 - \cdots - \left(\frac{\langle \mathbf{w}_k, \mathbf{v}_{k-1} \rangle}{\langle \mathbf{v}_{k-1}, \mathbf{v}_{k-1} \rangle}\right) \mathbf{v}_{k-1}$.

A proof similar to that of Theorem 6.4 (see Exercise 21) gives

Theorem 7.17 Let $B = \{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ be a basis for a finite dimensional inner product space $\mathcal{V}$. Then the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ obtained by applying the Generalized Gram-Schmidt Process to B is an orthogonal basis for $\mathcal{V}$.

Hence, every nontrivial finite dimensional inner product space has an orthogonal basis.

Example 11

Recall the inner product space $\mathcal{V}$ from Example 5 of real continuous functions using $a = -1$ and $b = 1$, that is, with inner product $\langle \mathbf{f}, \mathbf{g} \rangle = \int_{-1}^{1} \mathbf{f}(t)\mathbf{g}(t)\,dt$. Now, $\{1, t, t^2, t^3\}$ is a linearly independent set in $\mathcal{V}$. We use this set to find four orthogonal vectors in $\mathcal{V}$.

Let $\mathbf{w}_1 = 1$, $\mathbf{w}_2 = t$, $\mathbf{w}_3 = t^2$, and $\mathbf{w}_4 = t^3$. Using the Generalized Gram-Schmidt Process, we start with $\mathbf{v}_1 = \mathbf{w}_1 = 1$ and obtain

$$\mathbf{v}_2 = \mathbf{w}_2 - \left(\frac{\langle \mathbf{w}_2, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\right) \mathbf{v}_1 = t - \left(\frac{\langle t, 1 \rangle}{\langle 1, 1 \rangle}\right) 1.$$

Now, $\langle t, 1 \rangle = \int_{-1}^{1} (t)(1)\,dt = \left(t^2/2\right)\Big|_{-1}^{1} = 0$. Hence, $\mathbf{v}_2 = t$. Next,

$$\mathbf{v}_3 = \mathbf{w}_3 - \left(\frac{\langle \mathbf{w}_3, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\right) \mathbf{v}_1 - \left(\frac{\langle \mathbf{w}_3, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\right) \mathbf{v}_2 = t^2 - \left(\frac{\langle t^2, 1 \rangle}{\langle 1, 1 \rangle}\right) 1 - \left(\frac{\langle t^2, t \rangle}{\langle t, t \rangle}\right) t.$$

After a little calculation, we obtain $\langle t^2, 1 \rangle = \frac{2}{3}$, $\langle 1, 1 \rangle = 2$, and $\langle t^2, t \rangle = 0$. Hence, $\mathbf{v}_3 = t^2 - \left(\left(\frac{2}{3}\right)/2\right) 1 = t^2 - \frac{1}{3}$. Finally,

$$\begin{aligned} \mathbf{v}_4 &= \mathbf{w}_4 - \left(\frac{\langle \mathbf{w}_4, \mathbf{v}_1 \rangle}{\langle \mathbf{v}_1, \mathbf{v}_1 \rangle}\right) \mathbf{v}_1 - \left(\frac{\langle \mathbf{w}_4, \mathbf{v}_2 \rangle}{\langle \mathbf{v}_2, \mathbf{v}_2 \rangle}\right) \mathbf{v}_2 - \left(\frac{\langle \mathbf{w}_4, \mathbf{v}_3 \rangle}{\langle \mathbf{v}_3, \mathbf{v}_3 \rangle}\right) \mathbf{v}_3 \\ &= t^3 - \left(\frac{\langle t^3, 1 \rangle}{\langle 1, 1 \rangle}\right) 1 - \left(\frac{\langle t^3, t \rangle}{\langle t, t \rangle}\right) t - \left(\frac{\langle t^3, t^2 \rangle}{\langle t^2, t^2 \rangle}\right) t^2. \end{aligned}$$

Now, $\langle t^3, 1\rangle = 0$, $\langle t^3, t\rangle = \frac{2}{5}$, $\langle t, t\rangle = \frac{2}{3}$, and $\langle t^3, t^2\rangle = 0$. Hence, $\mathbf{v}_4 = t^3 - \left(\left(\frac{2}{5}\right) / \left(\frac{2}{3}\right)\right) t = t^3 - \frac{3}{5}t$.

Thus, the set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\} = \left\{1, t, t^2 - \frac{1}{3}, t^3 - \frac{3}{5}t\right\}$ is an orthogonal set of vectors in this inner product space.[3]

We saw in Theorem 6.8 that the transition matrix between orthonormal bases of $\mathbb{R}^n$ is an orthogonal matrix. This result generalizes to inner product spaces as follows:

Theorem 7.18 Let $\mathcal{V}$ be a finite dimensional real [complex] inner product space, and let B and C be ordered orthonormal bases for $\mathcal{V}$. Then the transition matrix from B to C is an orthogonal [unitary] matrix.

Orthogonal Complements and Orthogonal Projections in Inner Product Spaces

We can generalize the notion of an orthogonal complement of a subspace to inner product spaces as follows:

Definition Let $\mathcal{W}$ be a subspace of a real (or complex) inner product space $\mathcal{V}$. Then the **orthogonal complement** $\mathcal{W}^\perp$ of $\mathcal{W}$ in $\mathcal{V}$ is the set of all vectors $\mathbf{x} \in \mathcal{V}$ with the property that $\langle \mathbf{x}, \mathbf{w}\rangle = 0$, for all $\mathbf{w} \in \mathcal{W}$.

Example 12

Consider again the real vector space $\mathcal{P}_n$, with the inner product of Example 4 — for $\mathbf{p}_1 = a_n x^n + \cdots + a_1 x + a_0$ and $\mathbf{p}_2 = b_n x^n + \cdots + b_1 x + b_0$, $\langle \mathbf{p}_1, \mathbf{p}_2\rangle = a_n b_n + \cdots + a_1 b_1 + a_0 b_0$. Example 9 showed that $\left\{x^n, x^{n-1}, \ldots, x, 1\right\}$ is an orthogonal basis for $\mathcal{P}_n$ under this inner product. Now, consider the subspace $\mathcal{W}$ spanned by $\{x, 1\}$. A little thought will convince you that $\mathcal{W}^\perp = \text{span}\left\{x^n, x^{n-1}, \ldots, x^2\right\}$ and so, $\dim(\mathcal{W}) + \dim\left(\mathcal{W}^\perp\right) = 2 + (n-1) = n + 1 = \dim(\mathcal{P}_n)$.

The following properties of orthogonal complements are the analogs to Theorems 6.11 and 6.12 and Corollaries 6.13 and 6.14 and are proved in a similar manner (see Exercise 22):

[3] The polynomials 1, t, $t^2 - \frac{1}{3}$, and $t^3 - \frac{3}{5}t$ from Example 11 are multiples of the first four **Legendre polynomials**: $1, t, \frac{3}{2}t^2 - \frac{1}{2}, \frac{5}{2}t^3 - \frac{3}{2}t$. All Legendre polynomials equal 1 when $t = 1$. To find the complete set of Legendre polynomials, we can continue the Generalized Gram-Schmidt Process with t^4, t^5, t^6, and so on, and take appropriate multiples so that the resulting polynomials equal 1 when $t = 1$. These polynomials form an (infinite) orthogonal set for the inner product space of Example 11.

Theorem 7.19 Let $\mathcal{W}$ be a subspace of a real (or complex) inner product space $\mathcal{V}$. Then

(1) $\mathcal{W}^{\perp}$ is a subspace of $\mathcal{V}$.

(2) $\mathcal{W} \cap \mathcal{W}^{\perp} = \{\mathbf{0}\}$.

(3) $\mathcal{W} \subseteq \left(\mathcal{W}^{\perp}\right)^{\perp}$.

Furthermore, if $\mathcal{V}$ is finite dimensional, then

(4) If $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthogonal basis for $\mathcal{W}$ contained in an orthogonal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$ for $\mathcal{V}$, then $\{\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$ is an orthogonal basis for $\mathcal{W}^{\perp}$.

(5) $\dim(\mathcal{W}) + \dim\left(\mathcal{W}^{\perp}\right) = \dim(\mathcal{V})$.

(6) $\left(\mathcal{W}^{\perp}\right)^{\perp} = \mathcal{W}$.

Note that if $\mathcal{V}$ is not finite dimensional, $\left(\mathcal{W}^{\perp}\right)^{\perp}$ is not necessarily equal to $\mathcal{W}$, although it is always true that $\mathcal{W} \subseteq \left(\mathcal{W}^{\perp}\right)^{\perp}$.[4]

The next theorem is the analog of Theorem 6.15. It holds for any inner product space $\mathcal{V}$ where the subspace $\mathcal{W}$ is finite dimensional. The proof is left for you to do in Exercise 25.

Theorem 7.20 (Projection Theorem) Let $\mathcal{W}$ be a finite dimensional subspace of an inner product space $\mathcal{V}$. Then every vector $\mathbf{v} \in \mathcal{V}$ can be expressed in a unique way as $\mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{w}_1 \in \mathcal{W}$ and $\mathbf{w}_2 \in \mathcal{W}^{\perp}$.

As before, we define the **orthogonal projection** of a vector $\mathbf{v}$ onto a subspace $\mathcal{W}$ as follows:

Definition If $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is an orthonormal basis for $\mathcal{W}$, a subspace of an inner product space $\mathcal{V}$, then the vector $\mathbf{proj}_{\mathcal{W}}\mathbf{v} = \langle \mathbf{v}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \cdots + \langle \mathbf{v}, \mathbf{v}_k \rangle \mathbf{v}_k$ is called the **orthogonal projection of v onto** $\mathcal{W}$. If $\mathcal{W}$ is the trivial subspace of $\mathcal{V}$, then $\mathbf{proj}_{\mathcal{W}}\mathbf{v} = \mathbf{0}$.

It can be shown that the formula for $\mathbf{proj}_{\mathcal{W}}\mathbf{v}$ yields the unique vector $\mathbf{w}_1$ in the Projection Theorem. Therefore, the choice of orthonormal basis in the definition

[4] The following is an example of a subspace $\mathcal{W}$ of an infinite dimensional inner product space such that $\mathcal{W} \neq (\mathcal{W}^{\perp})^{\perp}$. Let $\mathcal{V}$ be the inner product space of Example 5 with $a = 0$, $b = 1$, and let $\mathbf{f}_n(x) = \begin{cases} 1, & \text{if } x > \frac{1}{n} \\ nx, & \text{if } 0 \le x \le \frac{1}{n} \end{cases}$. Let $\mathcal{W}$ be the subspace of $\mathcal{V}$ spanned by $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \ldots\}$. It can be shown that $\mathbf{f}(x) = 1$ is not in $\mathcal{W}$, but $\mathbf{f}(x) \in (\mathcal{W}^{\perp})^{\perp}$. Hence, $\mathcal{W} \neq (\mathcal{W}^{\perp})^{\perp}$.

does not matter because any choice leads to the same vector for $\mathbf{proj}_{\mathcal{W}}\mathbf{v}$. Hence, the Projection Theorem can be restated as follows:

If $\mathcal{W}$ is a finite dimensional subspace of an inner product space $\mathcal{V}$, and if $\mathbf{v} \in \mathcal{V}$, then $\mathbf{v}$ can be expressed as $\mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{w}_1 = \mathbf{proj}_{\mathcal{W}}\mathbf{v} \in \mathcal{W}$ and $\mathbf{w}_2 = \mathbf{v} - \mathbf{proj}_{\mathcal{W}}\mathbf{v} \in \mathcal{W}^{\perp}$.

Example 13

Consider again the real vector space $\mathcal{V}$ of real continuous functions in Example 8, where $\langle \mathbf{f}, \mathbf{g} \rangle = \int_{-\pi}^{\pi} \mathbf{f}(t)\mathbf{g}(t)\,dt$. Notice from that example that the set $\{1/\sqrt{2\pi}, (\sin t)/\sqrt{\pi}\}$ is an orthonormal (and hence, linearly independent) set of vectors in $\mathcal{V}$. Let $\mathcal{W} = \operatorname{span}(\{1/\sqrt{2\pi}, (\sin t)/\sqrt{\pi}\})$ in $\mathcal{V}$. Then any continuous function $\mathbf{f}$ in $\mathcal{V}$ can be expressed uniquely as $\mathbf{f}_1 + \mathbf{f}_2$, where $\mathbf{f}_1 \in \mathcal{W}$ and $\mathbf{f}_2 \in \mathcal{W}^{\perp}$.

We illustrate this decomposition for the function $\mathbf{f} = t + 1$. Now,

$$\mathbf{f}_1 = \mathbf{proj}_{\mathcal{W}}\mathbf{f} = c_1\left(\frac{1}{\sqrt{2\pi}}\right) + c_2\left(\frac{\sin t}{\sqrt{\pi}}\right),$$

where $c_1 = \langle (t+1), 1/\sqrt{2\pi} \rangle$ and $c_2 = \langle (t+1), (\sin t)/\sqrt{\pi} \rangle$. Then

$$\begin{aligned} c_1 &= \int_{-\pi}^{\pi} (t+1)\left(\frac{1}{\sqrt{2\pi}}\right) dt = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} (t+1)\,dt \\ &= \frac{1}{\sqrt{2\pi}}\left(\frac{t^2}{2} + t\right)\Bigg|_{-\pi}^{\pi} = \frac{2\pi}{\sqrt{2\pi}} = \sqrt{2\pi}. \end{aligned}$$

Also,

$$\begin{aligned} c_2 &= \int_{-\pi}^{\pi} (t+1)\left(\frac{\sin t}{\sqrt{\pi}}\right) dt = \frac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} (t+1)\sin t\,dt \\ &= \frac{1}{\sqrt{\pi}}\left(\int_{-\pi}^{\pi} t\sin t\,dt + \int_{-\pi}^{\pi} \sin t\,dt\right). \end{aligned}$$

The very last integral equals zero. Using integration by parts on the other integral, we obtain

$$c_2 = \frac{1}{\sqrt{\pi}}\left((-t\cos t)\big|_{-\pi}^{\pi} + \int_{-\pi}^{\pi} \cos t\,dt\right) = \left(\frac{1}{\sqrt{\pi}}\right)2\pi = 2\sqrt{\pi}.$$

Hence,

$$\mathbf{f}_1 = c_1\left(\frac{1}{\sqrt{2\pi}}\right) + c_2\left(\frac{\sin t}{\sqrt{\pi}}\right) = \sqrt{2\pi}\left(\frac{1}{\sqrt{2\pi}}\right) + 2\sqrt{\pi}\left(\frac{\sin t}{\sqrt{\pi}}\right) = 1 + 2\sin t.$$

Then by the Projection Theorem, $\mathbf{f}_2 = \mathbf{f} - \mathbf{f}_1 = (t+1) - (1 + 2\sin t) = t - 2\sin t$ is orthogonal to $\mathcal{W}$. We check that $\mathbf{f}_2 \in \mathcal{W}^{\perp}$ by showing that $\mathbf{f}_2$ is orthogonal to both $1/\sqrt{2\pi}$ and $(\sin t)/\sqrt{\pi}$.

$$\left\langle \mathbf{f}_2, \frac{1}{\sqrt{2\pi}} \right\rangle = \int_{-\pi}^{\pi} (t - 2\sin t)\left(\frac{1}{\sqrt{2\pi}}\right) dt = \left(\frac{1}{\sqrt{2\pi}}\right)\left(\frac{t^2}{2} + 2\cos t\right)\Bigg|_{-\pi}^{\pi} = 0.$$

Also,

$$\left\langle \mathbf{f}_2, \frac{\sin t}{\sqrt{\pi}} \right\rangle = \int_{-\pi}^{\pi} (t - 2\sin t)\left(\frac{\sin t}{\sqrt{\pi}}\right) dt = \frac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} t \sin t\, dt - \frac{2}{\sqrt{\pi}} \int_{-\pi}^{\pi} \sin^2 t\, dt,$$

which equals $2\sqrt{\pi} - 2\sqrt{\pi} = 0$. ■

New Vocabulary

angle between vectors (in an inner product space)
Cauchy-Schwarz Inequality (in an inner product space)
complex inner product (on a complex vector space)
complex inner product space
distance between vectors (in an inner product space)
Fourier series
Generalized Gram-Schmidt Process (in an inner product space)
Legendre polynomials
norm (length) of a vector (in an inner product space)
orthogonal basis (in an inner product space)
orthogonal complement (of a subspace in an inner product space)
orthogonal projection (of a vector onto a subspace of an inner product space)
orthogonal set of vectors (in an inner product space)
orthonormal basis (in an inner product space)
orthonormal set of vectors (in an inner product space)
real inner product (on a real vector space)
real inner product space
Triangle Inequality (in an inner product space)
unit vector (in an inner product space)

Highlights

- Real and complex inner products are generalizations of the real and complex dot products, respectively.
- An inner product space is a vector space that possesses three operations: vector addition, scalar multiplication, and inner product.
- For vectors $\mathbf{x}, \mathbf{y}$ and scalar k in a real inner product space, $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$, and $\langle \mathbf{x}, k\mathbf{y} \rangle = k\langle \mathbf{x}, \mathbf{y} \rangle$.
- For vectors $\mathbf{x}$, $\mathbf{y}$ and scalar k in a real or complex inner product space, $\langle k\mathbf{x}, \mathbf{y} \rangle = k\langle \mathbf{x}, \mathbf{y} \rangle$.
- For vectors $\mathbf{x}, \mathbf{y}$ and scalar k in a complex inner product space, $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$, $\langle \mathbf{x}, k\mathbf{y} \rangle = \overline{k}\langle \mathbf{x}, \mathbf{y} \rangle$, and $\|k\mathbf{x}\| = |k|\,\|\mathbf{x}\|$.
- The length of a vector $\mathbf{x}$ in an inner product space is $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$, and the distance between vectors $\mathbf{x}$ and $\mathbf{y}$ in an inner product space is $\|\mathbf{x} - \mathbf{y}\|$.

- The angle θ between two vectors in a real inner product space is defined as the angle between 0 and π such that $\cos\theta = \langle \mathbf{x},\mathbf{y}\rangle / \left(\|\mathbf{x}\|\,\|\mathbf{y}\|\right)$.
- Orthogonal and orthonormal sets of vectors, and orthogonal complements of subspaces, are defined for inner product spaces analogously as for real vector spaces.
- An orthogonal set of nonzero vectors in an inner product space is a linearly independent set.
- If $B = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k)$ is an orthogonal ordered basis for a subspace $\mathcal{W}$ of an inner product space $\mathcal{V}$, and if $\mathbf{v}$ is any vector in $\mathcal{W}$, then $[\mathbf{v}]_B = \left[\frac{\langle \mathbf{v},\mathbf{v}_1\rangle}{\langle \mathbf{v}_1,\mathbf{v}_1\rangle}, \frac{\langle \mathbf{v},\mathbf{v}_2\rangle}{\langle \mathbf{v}_2,\mathbf{v}_2\rangle}, \ldots, \frac{\langle \mathbf{v},\mathbf{v}_k\rangle}{\langle \mathbf{v}_k,\mathbf{v}_k\rangle}\right]$.
- The Generalized Gram-Schmidt Process can be used to find an orthogonal basis for any subspace spanned by a finite linearly independent subset.
- If $\mathcal{W}$ is a finite dimensional subspace of an inner product space $\mathcal{V}$, then every vector $\mathbf{v}$ in $\mathcal{V}$ can be expressed uniquely as the sum of vectors $\mathbf{w}_1 = \mathbf{proj}_{\mathcal{W}}\mathbf{v} \in \mathcal{W}$ and $\mathbf{w}_2 = \mathbf{v} - \mathbf{proj}_{\mathcal{W}}\mathbf{v} \in \mathcal{W}^{\perp}$.
- The transition matrix from one ordered orthonormal basis to another in a real [complex] inner product space is an orthogonal [unitary] matrix.

EXERCISES FOR SECTION 7.5

1. **(a)** Let $\mathbf{A}$ be a nonsingular $n \times n$ real matrix. For $\mathbf{x},\mathbf{y} \in \mathbb{R}^n$, define an operation $\langle \mathbf{x},\mathbf{y}\rangle = (\mathbf{A}\mathbf{x}) \cdot (\mathbf{A}\mathbf{y})$ (dot product). Prove that this operation is a real inner product on $\mathbb{R}^n$.

$\star$**(b)** For the inner product in part (a) with $\mathbf{A} = \begin{bmatrix} 5 & 4 & 2 \\ -2 & 3 & 1 \\ 1 & -1 & 0 \end{bmatrix}$, find $\langle \mathbf{x},\mathbf{y}\rangle$ and $\|\mathbf{x}\|$, for $\mathbf{x} = [3,-2,4]$ and $\mathbf{y} = [-2,1,-1]$.

2. Define an operation $\langle\, ,\rangle$ on $\mathcal{P}_n$ as follows: if $\mathbf{p}_1 = a_n x^n + \cdots + a_1 x + a_0$ and $\mathbf{p}_2 = b_n x^n + \cdots + b_1 x + b_0$, let $\langle \mathbf{p}_1, \mathbf{p}_2\rangle = a_n b_n + \cdots + a_1 b_1 + a_0 b_0$. Prove that this operation is a real inner product on $\mathcal{P}_n$.

3. **(a)** Let a and b be fixed real numbers with $a < b$, and let $\mathcal{V}$ be the set of all real continuous functions on $[a,b]$. Define $\langle\, ,\rangle$ on $\mathcal{V}$ by $\langle \mathbf{f},\mathbf{g}\rangle = \int_a^b \mathbf{f}(t)\mathbf{g}(t)\,dt$. Prove that this operation is a real inner product on $\mathcal{V}$.

$\star$**(b)** For the inner product of part (a) with $a = 0$ and $b = \pi$, find $\langle \mathbf{f},\mathbf{g}\rangle$ and $\|\mathbf{f}\|$, for $\mathbf{f} = e^t$ and $\mathbf{g} = \sin t$.

4. Define $\langle\, ,\rangle$ on the real vector space $\mathcal{M}_{mn}$ by $\langle \mathbf{A},\mathbf{B}\rangle = \text{trace}(\mathbf{A}^T\mathbf{B})$. Prove that this operation is a real inner product on $\mathcal{M}_{mn}$. (Hint: Refer to Exercise 14 in Section 1.4 and Exercise 26 in Section 1.5.)

5. **(a)** Prove part (1) of Theorem 7.12. (Hint: $\mathbf{0} = \mathbf{0} + \mathbf{0}$. Use property (4) in the definition of an inner product space.)

(b) Prove part (3) of Theorem 7.12. (Be sure to give a proof for both real and complex inner product spaces.)

▶**6.** Prove Theorem 7.13.

7. Let $\mathbf{x}, \mathbf{y} \in \mathcal{V}$, a real inner product space.

(a) Prove that $\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2\langle \mathbf{x}, \mathbf{y}\rangle + \|\mathbf{y}\|^2$.

(b) Show that $\mathbf{x}$ and $\mathbf{y}$ are orthogonal in $\mathcal{V}$ if and only if $\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$.

(c) Show that $\frac{1}{2}\left(\|\mathbf{x} + \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}\|^2\right) = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$.

8. The following formulas show how the value of the inner product can be derived from the norm (length):

(a) Let $\mathbf{x}, \mathbf{y} \in \mathcal{V}$, a real inner product space. Prove the following (real) **Polarization Identity:**

$$\langle \mathbf{x}, \mathbf{y}\rangle = \frac{1}{4}\left(\|\mathbf{x} + \mathbf{y}\|^2 - \|\mathbf{x} - \mathbf{y}\|^2\right).$$

(b) Let $\mathbf{x}, \mathbf{y} \in \mathcal{V}$, a complex inner product space. Prove the following **Complex Polarization Identity:**

$$\langle \mathbf{x}, \mathbf{y}\rangle = \frac{1}{4}\left(\left(\|\mathbf{x} + \mathbf{y}\|^2 - \|\mathbf{x} - \mathbf{y}\|^2\right) + i\left(\|\mathbf{x} + i\mathbf{y}\|^2 - \|\mathbf{x} - i\mathbf{y}\|^2\right)\right).$$

9. Consider the inner product space $\mathcal{V}$ of Example 5, with $a = 0$ and $b = \pi$.

⋆**(a)** Find the distance between $\mathbf{f} = t$ and $\mathbf{g} = \sin t$ in $\mathcal{V}$.

(b) Find the angle between $\mathbf{f} = e^t$ and $\mathbf{g} = \sin t$ in $\mathcal{V}$.

10. Consider the inner product space $\mathcal{V}$ of Example 3, using

$$\mathbf{A} = \begin{bmatrix} -2 & 0 & 1 \\ 1 & -1 & 2 \\ 3 & -1 & -1 \end{bmatrix}.$$

(a) Find the distance between $\mathbf{x} = [2, -1, 3]$ and $\mathbf{y} = [5, -2, 2]$ in $\mathcal{V}$.

⋆**(b)** Find the angle between $\mathbf{x} = [2, -1, 3]$ and $\mathbf{y} = [5, -2, 2]$ in $\mathcal{V}$.

11. Let $\mathcal{V}$ be an inner product space.

(a) Prove part (1) of Theorem 7.14. (Hint: Modify the proof of Theorem 1.6.)

(b) Prove part (2) of Theorem 7.14. (Hint: Modify the proof of Theorem 1.7.)

12. Let f and g be continuous real-valued functions defined on a closed interval $[a,b]$. Show that

$$\left(\int_a^b f(t)g(t)\,dt\right)^2 \le \int_a^b \left(f(t)\right)^2\,dt \int_a^b \left(g(t)\right)^2\,dt.$$

(Hint: Use the Cauchy-Schwarz Inequality in an appropriate inner product space.)

13. A **metric space** is a set in which every pair of elements x, y has been assigned a real number distance d with the following properties:

(i) $d(x,y) = d(y,x)$.

(ii) $d(x,y) \ge 0$, with $d(x,y) = 0$ if and only if $x = y$.

(iii) $d(x,y) \le d(x,z) + d(z,y)$, for all z in the set.

Prove that every inner product space is a metric space with $d(\mathbf{x},\mathbf{y})$ taken to be $\|\mathbf{x}-\mathbf{y}\|$ for all vectors $\mathbf{x}$ and $\mathbf{y}$ in the space.

14. Determine whether the following sets of vectors are orthogonal:

⋆**(a)** $\{t^2,\ t+1,\ t-1\}$ in $\mathcal{P}_3$, under the inner product of Example 4

(b) $\{[15,9,19],[-2,-1,-2],[-12,-9,-14]\}$ in $\mathbb{R}^3$, under the inner product of Example 3, with

$$\mathbf{A} = \begin{bmatrix} -3 & 1 & 2 \\ 0 & -2 & 1 \\ 2 & -1 & -1 \end{bmatrix}$$

⋆**(c)** $\{[5,-2],[3,4]\}$ in $\mathbb{R}^2$, under the inner product of Example 2

(d) $\{3t^2-1,\ 4t,\ 5t^3-3t\}$ in $\mathcal{P}_3$, under the inner product of Example 11

15. Prove Theorem 7.15. (Hint: Modify the proof of Result 7 in Section 1.3.)

16. **(a)** Show that $\int_{-\pi}^{\pi} \cos mt\,dt = 0$ and $\int_{-\pi}^{\pi} \sin nt\,dt = 0$, for all integers $m,n \ge 1$.

(b) Show that $\int_{-\pi}^{\pi} \cos mt \cos nt\,dt = 0$ and $\int_{-\pi}^{\pi} \sin mt \sin nt\,dt = 0$, for any *distinct* integers $m,n \ge 1$. (Hint: Use trigonometric identities.)

(c) Show that $\int_{-\pi}^{\pi} \cos mt \sin nt\,dt = 0$, for any integers $m,n \ge 1$.

(d) Conclude from parts (a), (b), and (c) that $\{1, \cos t, \sin t, \cos 2t, \sin 2t, \cos 3t, \sin 3t, \ldots\}$ is an orthogonal set of real continuous functions on $[-\pi,\pi]$, as claimed after Example 8.

17. Prove Theorem 7.16. (Hint: Modify the proof of Theorem 6.3.)

18. Let $\{\mathbf{v}_1,\ldots,\mathbf{v}_k\}$ be an orthonormal basis for a complex inner product space $\mathcal{V}$. Prove that for all $\mathbf{v},\mathbf{w} \in \mathcal{V}$,

$$\langle \mathbf{v},\mathbf{w}\rangle = \langle \mathbf{v},\mathbf{v}_1\rangle\,\overline{\langle \mathbf{w},\mathbf{v}_1\rangle} + \langle \mathbf{v},\mathbf{v}_2\rangle\,\overline{\langle \mathbf{w},\mathbf{v}_2\rangle} + \cdots + \langle \mathbf{v},\mathbf{v}_k\rangle\,\overline{\langle \mathbf{w},\mathbf{v}_k\rangle}.$$

(Compare this with Exercise 9(a) in Section 6.1.)

★19. Use the Generalized Gram-Schmidt Process to find an orthogonal basis for $\mathcal{P}_2$ containing $t^2 - t + 1$ under the inner product of Example 11.

20. Use the Generalized Gram-Schmidt Process to find an orthogonal basis for $\mathbb{R}^3$ containing $[-9, -4, 8]$ under the inner product of Example 3 with the matrix

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

21. Prove Theorem 7.17. (Hint: Modify the proof of Theorem 6.4.)

22. (a) Prove parts (1) and (2) of Theorem 7.19. (Hint: Modify the proof of Theorem 6.11.)

▶(b) Prove parts (4) and (5) of Theorem 7.19. (Hint: Modify the proofs of Theorem 6.12 and Corollary 6.13.)

(c) Prove part (3) of Theorem 7.19.

▶(d) Prove part (6) of Theorem 7.19. (Hint: Use part (5) of Theorem 7.19 to show that $\dim(\mathcal{W}) = \dim\left(\left(\mathcal{W}^{\perp}\right)^{\perp}\right)$. Then use part (c) and apply Theorem 4.16, or its complex analog.)

★23. Find $\mathcal{W}^{\perp}$ if $\mathcal{W} = \text{span}\left(\left\{t^3 + t^2,\ t - 1\right\}\right)$ in $\mathcal{P}_3$ with the inner product of Example 4.

24. Find an orthogonal basis for $\mathcal{W}^{\perp}$ if $\mathcal{W} = \text{span}(\{(t-1)^2\})$ in $\mathcal{P}_2$, with the inner product $\langle \mathbf{f}, \mathbf{g} \rangle = \int_0^1 \mathbf{f}(t)\mathbf{g}(t)\,dt$, for all $\mathbf{f}, \mathbf{g} \in \mathcal{P}_2$.

▶25. Prove Theorem 7.20. (Hint: Choose an orthonormal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ for $\mathcal{W}$. Then define $\mathbf{w}_1 = \mathbf{proj}_{\mathcal{W}}\mathbf{v} = \langle \mathbf{v}, \mathbf{v}_1 \rangle \mathbf{v}_1 + \cdots + \langle \mathbf{v}, \mathbf{v}_k \rangle \mathbf{v}_k$. Let $\mathbf{w}_2 = \mathbf{v} - \mathbf{w}_1$, and prove $\mathbf{w}_2 \in \mathcal{W}^{\perp}$. Finally, see the proof of Theorem 6.15 for uniqueness.)

★26. In the inner product space of Example 8, decompose $\mathbf{f} = \frac{1}{k}e^t$, where $k = e^{\pi} - e^{-\pi}$, as $\mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{w}_1 \in \mathcal{W} = \text{span}(\{\cos t, \sin t\})$ and $\mathbf{w}_2 \in \mathcal{W}^{\perp}$. Check that $\langle \mathbf{w}_1, \mathbf{w}_2 \rangle = 0$. (Hint: First find an orthonormal basis for $\mathcal{W}$.)

27. Decompose $\mathbf{v} = 4t^2 - t + 3$ in $\mathcal{P}_2$ as $\mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{w}_1 \in \mathcal{W} = \text{span}(\{2t^2 - 1, t + 1\})$ and $\mathbf{w}_2 \in \mathcal{W}^{\perp}$, under the inner product of Example 11. Check that $\langle \mathbf{w}_1, \mathbf{w}_2 \rangle = 0$. (Hint: First find an orthonormal basis for $\mathcal{W}$.)

28. **Bessel's Inequality:** Let $\mathcal{V}$ be a real inner product space, and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be an orthonormal set in $\mathcal{V}$. Prove that for any vector $\mathbf{v} \in \mathcal{V}$, $\sum_{i=1}^{k} \langle \mathbf{v}, \mathbf{v}_i \rangle^2 \leq \|\mathbf{v}\|^2$. (Hint: Let $\mathcal{W} = \text{span}(\{\mathbf{v}_1, \ldots, \mathbf{v}_k\})$. Now, $\mathbf{v} = \mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{w}_1 = \mathbf{proj}_{\mathcal{W}}\mathbf{v} \in \mathcal{W}$ and $\mathbf{w}_2 \in \mathcal{W}^{\perp}$. Expand $\langle \mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{w}_1 + \mathbf{w}_2, \mathbf{w}_1 + \mathbf{w}_2 \rangle$. Show that $\|\mathbf{v}\|^2 \geq \|\mathbf{w}_1\|^2$, and use the definition of $\mathbf{proj}_{\mathcal{W}}\mathbf{v}$.)

29. Let $\mathcal{W}$ be a finite dimensional subspace of an inner product space $\mathcal{V}$. Consider the mapping $L: \mathcal{V} \to \mathcal{W}$ given by $L(\mathbf{v}) = \mathbf{proj}_{\mathcal{W}}\mathbf{v}$.

(a) Prove that L is a linear transformation.

★(b) What are the kernel and range of L?

(c) Show that $L \circ L = L$.

★30. True or False:

(a) If $\mathcal{V}$ is a complex inner product space, then for all $\mathbf{x} \in \mathcal{V}$ and all $k \in \mathbb{C}$, $\|k\mathbf{x}\| = \overline{k}\|\mathbf{x}\|$.

(b) In a complex inner product space, the distance between two distinct vectors can be a pure imaginary number.

(c) Every linearly independent set of unit vectors in an inner product space is an orthonormal set.

(d) It is possible to define more than one inner product on the same vector space.

(e) The uniqueness proof of the Projection Theorem shows that if $\mathcal{W}$ is a subspace of $\mathbb{R}^n$, then $\mathbf{proj}_{\mathcal{W}}\mathbf{v}$ is independent of the particular inner product used on $\mathbb{R}^n$.

REVIEW EXERCISES FOR CHAPTER 7

1. Let $\mathbf{v}$, $\mathbf{w}$, and $\mathbf{z} \in \mathbb{C}^3$ be given by $\mathbf{v} = [i, 3-i, 2+3i]$, $\mathbf{w} = [-4-4i,\ 1+2i,\ 3-i]$, and $\mathbf{z} = [2+5i,\ 2-5i,\ -i]$.

★(a) Compute $\mathbf{v} \cdot \mathbf{w}$.

★(b) Compute $(1+2i)(\mathbf{v} \cdot \mathbf{z})$, $((1+2i)\mathbf{v}) \cdot \mathbf{z}$, and $\mathbf{v} \cdot ((1+2i)\mathbf{z})$.

(c) Explain why not all of the answers to part (b) are identical.

(d) Compute $\mathbf{w} \cdot \mathbf{z}$ and $\mathbf{w} \cdot (\mathbf{v} + \mathbf{z})$.

2. **(a)** Compute $\mathbf{H} = \mathbf{A}^*\mathbf{A}$, where $\mathbf{A} = \begin{bmatrix} 1-i & 2+i & 3-4i \\ 0 & 5-2i & -2+i \end{bmatrix}$ and show that $\mathbf{H}$ is Hermitian.

(b) Show that $\mathbf{AA}^*$ is also Hermitian.

3. Prove that if $\mathbf{A}$ is a skew-Hermitian $n \times n$ matrix and $\mathbf{w}, \mathbf{z} \in \mathbb{C}^n$, then $(\mathbf{Az}) \cdot \mathbf{w} = -\mathbf{z} \cdot (\mathbf{Aw})$.

4. In each part, solve the given system of linear equations.

★(a) $\begin{cases} (i)w + (1+i)z = -1+2i \\ (1+i)w + (5+2i)z = 5-3i \\ (2-i)w + (2-5i)z = 1-2i \end{cases}$

CHAPTER 9

Vector Differential Calculus. Grad, Div, Curl

Engineering, physics, and computer sciences, in general, but particularly solid mechanics, aerodynamics, aeronautics, fluid flow, heat flow, electrostatics, quantum physics, laser technology, robotics as well as other areas have applications that require an understanding of **vector calculus**. This field encompasses vector differential calculus and vector integral calculus. Indeed, the engineer, physicist, and mathematician need a good grounding in these areas as provided by the carefully chosen material of Chaps. 9 and 10.

Forces, velocities, and various other quantities may be thought of as vectors. Vectors appear frequently in the applications above and also in the biological and social sciences, so it is natural that problems are modeled in **3-space**. This is the space of three dimensions with the usual measurement of distance, as given by the Pythagorean theorem. Within that realm, **2-space** (the plane) is a special case. Working in 3-space requires that we extend the common differential calculus to vector differential calculus, that is, the calculus that deals with vector functions and vector fields and is explained in this chapter.

Chapter 9 is arranged in three groups of sections. Sections 9.1–9.3 extend the basic algebraic operations of vectors into 3-space. These operations include the inner product and the cross product. Sections 9.4 and 9.5 form the heart of vector differential calculus. Finally, Secs. 9.7–9.9 discuss three physically important concepts related to scalar and vector fields: gradient (Sec. 9.7), divergence (Sec. 9.8), and curl (Sec. 9.9). They are expressed in Cartesian coordinates in this chapter and, if desired, expressed in ***curvilinear coordinates*** in a short section in App. A3.4.

We shall keep this chapter ***independent of Chaps. 7 and 8***. Our present approach is in harmony with Chap. 7, with the restriction to two and three dimensions providing for a richer theory with basic physical, engineering, and geometric applications.

Prerequisite: Elementary use of second- and third-order determinants in Sec. 9.3.
Sections that may be omitted in a shorter course: 9.5, 9.6.
References and Answers to Problems: App. 1 Part B, App. 2.

9.1 Vectors in 2-Space and 3-Space

In engineering, physics, mathematics, and other areas we encounter two kinds of quantities. They are scalars and vectors.

A **scalar** is a quantity that is determined by its magnitude. It takes on a numerical value, i.e., a number. Examples of scalars are time, temperature, length, distance, speed, density, energy, and voltage.

In contrast, a **vector** is a quantity that has both magnitude and direction. We can say that a vector is an ***arrow*** or a ***directed line segment***. For example, a velocity vector has length or magnitude, which is speed, and direction, which indicates the direction of motion. Typical examples of vectors are displacement, velocity, and force, see Fig. 164 as an illustration.

More formally, we have the following. We denote vectors by lowercase boldface letters **a**, **b**, **v**, etc. In handwriting you may use arrows, for instance, $\vec{a}$ (in place of **a**), $\vec{b}$, etc.

A vector (arrow) has a tail, called its **initial point**, and a tip, called its **terminal point**. This is motivated in the **translation** (displacement without rotation) of the triangle in Fig. 165, where the initial point P of the vector **a** is the original position of a point, and the terminal point Q is the terminal position of that point, its position *after* the translation. The length of the arrow equals the distance between P and Q. This is called the **length** (or *magnitude*) of the vector **a** and is denoted by $|\mathbf{a}|$. Another name for *length* is **norm** (or *Euclidean norm*).

A vector of length 1 is called a **unit vector**.

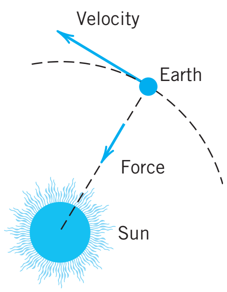


Fig. 164. Force and velocity

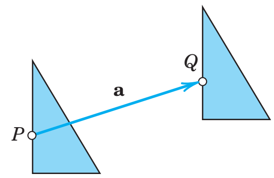


Fig. 165. Translation

Of course, we would like to calculate with vectors. For instance, we want to find the resultant of forces or compare parallel forces of different magnitude. This motivates our next ideas: to define *components* of a vector, and then the two basic algebraic operations of *vector addition* and *scalar multiplication.*

For this we must first define *equality of vectors* in a way that is practical in connection with forces and other applications.

DEFINITION

Equality of Vectors

Two vectors **a** and **b** are equal, written $\mathbf{a} = \mathbf{b}$, if they have the same length and the same direction [as explained in Fig. 166; in particular, note (B)]. Hence a vector can be arbitrarily translated; that is, its initial point can be chosen arbitrarily.

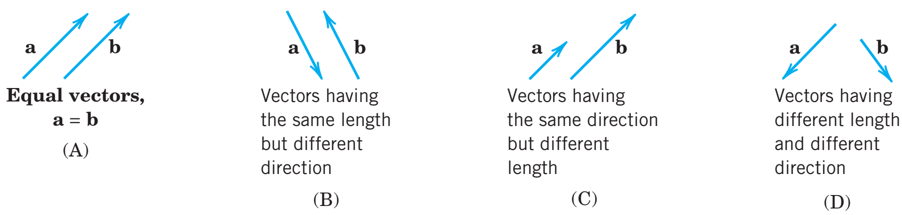


Fig. 166. (A) Equal vectors. (B)–(D) Different vectors

Components of a Vector

We choose an xyz **Cartesian coordinate system**[1] in space (Fig. 167), that is, a usual rectangular coordinate system with the same scale of measurement on the three mutually perpendicular coordinate axes. Let $\mathbf{a}$ be a given vector with initial point P: (x_1, y_1, z_1) and terminal point Q: (x_2, y_2, z_2). Then the three coordinate differences

$$a_1 = x_2 - x_1, \qquad a_2 = y_2 - y_1, \qquad a_3 = z_2 - z_1 \tag{1}$$

are called the **components** of the vector $\mathbf{a}$ with respect to that coordinate system, and we write simply $\mathbf{a} = [a_1, a_2, a_3]$. See Fig. 168.

The **length** $|\mathbf{a}|$ of $\mathbf{a}$ can now readily be expressed in terms of components because from (1) and the Pythagorean theorem we have

$$|\mathbf{a}| = \sqrt{a_1^2 + a_2^2 + a_3^2}. \tag{2}$$

EXAMPLE 1 Components and Length of a Vector

The vector $\mathbf{a}$ with initial point P: $(4, 0, 2)$ and terminal point Q: $(6, -1, 2)$ has the components

$$a_1 = 6 - 4 = 2, \qquad a_2 = -1 - 0 = -1, \qquad a_3 = 2 - 2 = 0.$$

Hence $\mathbf{a} = [2, -1, 0]$. (Can you sketch $\mathbf{a}$, as in Fig. 168?) Equation (2) gives the length

$$|\mathbf{a}| = \sqrt{2^2 + (-1)^2 + 0^2} = \sqrt{5}.$$

If we choose $(-1, 5, 8)$ as the initial point of $\mathbf{a}$, the corresponding terminal point is $(1, 4, 8)$.

If we choose the origin $(0, 0, 0)$ as the initial point of $\mathbf{a}$, the corresponding terminal point is $(2, -1, 0)$; its coordinates equal the components of $\mathbf{a}$. This suggests that we can determine each point in space by a vector, called the *position vector* of the point, as follows. ■

A Cartesian coordinate system being given, the **position vector r** of a point A: (x, y, z) is the vector with the origin $(0, 0, 0)$ as the initial point and A as the terminal point (see Fig. 169). Thus in components, $\mathbf{r} = [x, y, z]$. This can be seen directly from (1) with $x_1 = y_1 = z_1 = 0$.

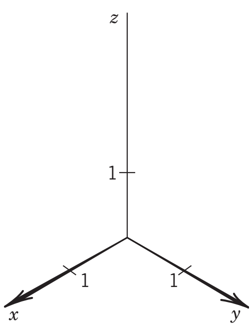


Fig. 167. Cartesian coordinate system

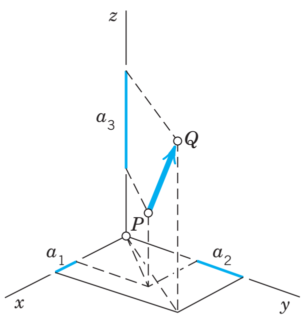


Fig. 168. Components of a vector

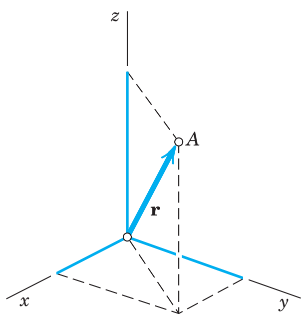


Fig. 169. Position vector **r** of a point A: (x, y, z)

[1]Named after the French philosopher and mathematician RENATUS CARTESIUS, latinized for RENÉ DESCARTES (1596–1650), who invented analytic geometry. His basic work *Géométrie* appeared in 1637, as an appendix to his *Discours de la méthode*.

Furthermore, if we translate a vector **a**, with initial point P and terminal point Q, then corresponding coordinates of P and Q change by the same amount, so that the differences in (1) remain unchanged. This proves

THEOREM 1

Vectors as Ordered Triples of Real Numbers

A fixed Cartesian coordinate system being given, each vector is uniquely determined by its ordered triple of corresponding components. Conversely, to each ordered triple of real numbers (a_1, a_2, a_3) *there corresponds precisely one vector* $\mathbf{a} = [a_1, a_2, a_3]$, *with* $(0, 0, 0)$ *corresponding to the* **zero vector 0**, *which has length* 0 *and no direction.*

Hence a vector equation $\mathbf{a} = \mathbf{b}$ *is equivalent to the three equations* $a_1 = b_1$, $a_2 = b_2$, $a_3 = b_3$ *for the components.*

We now see that from our "geometric" definition of a vector as an arrow we have arrived at an "algebraic" characterization of a vector by Theorem 1. We could have started from the latter and reversed our process. This shows that the two approaches are equivalent.

Vector Addition, Scalar Multiplication

Calculations with vectors are very useful and are almost as simple as the arithmetic for real numbers. Vector arithmetic follows almost naturally from applications. We first define how to add vectors and later on how to multiply a vector by a number.

DEFINITION

Addition of Vectors

The **sum a + b** of two vectors $\mathbf{a} = [a_1, a_2, a_3]$ and $\mathbf{b} = [b_1, b_2, b_3]$ is obtained by adding the corresponding components,

$$\mathbf{a} + \mathbf{b} = [a_1 + b_1, \quad a_2 + b_2, \quad a_3 + b_3]. \tag{3}$$

Geometrically, place the vectors as in Fig. 170 (the initial point of **b** at the terminal point of **a**); then **a** + **b** is the vector drawn from the initial point of **a** to the terminal point of **b**.

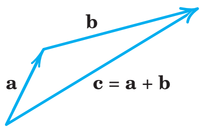


Fig. 170. Vector addition

For forces, this addition is the parallelogram law by which we obtain the **resultant** of two forces in mechanics. See Fig. 171.

Figure 172 shows (for the plane) that the "algebraic" way and the "geometric way" of vector addition give the same vector.

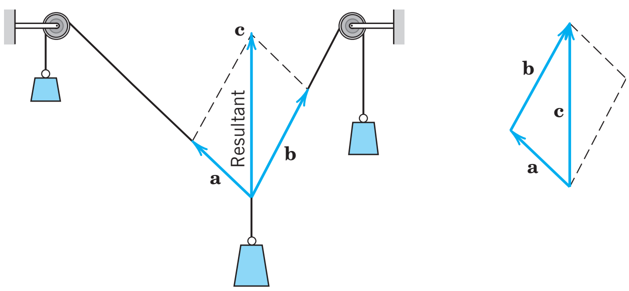


Fig. 171. Resultant of two forces (**parallelogram law**)

Basic Properties of Vector Addition. Familiar laws for real numbers give immediately

$$
\textbf{(4)}\qquad
\begin{aligned}
&\text{(a)} && \mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a} && \textit{(Commutativity)}\\
&\text{(b)} && (\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w}) && \textit{(Associativity)}\\
&\text{(c)} && \mathbf{a} + \mathbf{0} = \mathbf{0} + \mathbf{a} = \mathbf{a} &&\\
&\text{(d)} && \mathbf{a} + (-\mathbf{a}) = \mathbf{0}. &&
\end{aligned}
$$

Properties (a) and (b) are verified geometrically in Figs. 173 and 174. Furthermore, $-\mathbf{a}$ denotes the vector having the length $|\mathbf{a}|$ and the direction opposite to that of $\mathbf{a}$.

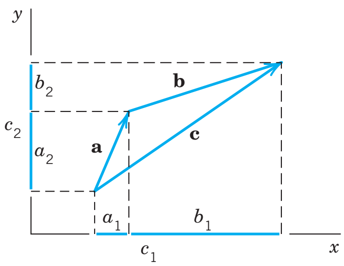


Fig. 172. Vector addition

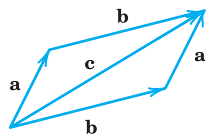


Fig. 173. Cummutativity of vector addition

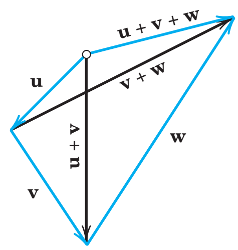


Fig. 174. Associativity of vector addition

In (4b) we may simply write $\mathbf{u} + \mathbf{v} + \mathbf{w}$, and similarly for sums of more than three vectors. Instead of $\mathbf{a} + \mathbf{a}$ we also write $2\mathbf{a}$, and so on. This (and the notation $-\mathbf{a}$ used just before) motivates defining the second algebraic operation for vectors as follows.

DEFINITION

Scalar Multiplication (Multiplication by a Number)

The product $c\mathbf{a}$ of any vector $\mathbf{a} = [a_1, a_2, a_3]$ and any scalar c (real number c) is the vector obtained by multiplying each component of $\mathbf{a}$ by c,

$$
\textbf{(5)}\qquad c\mathbf{a} = [ca_1, ca_2, ca_3].
$$

Geometrically, if $\mathbf{a} \neq \mathbf{0}$, then $c\mathbf{a}$ with $c > 0$ has the direction of $\mathbf{a}$ and with $c < 0$ the direction opposite to $\mathbf{a}$. In any case, the length of $c\mathbf{a}$ is $|c\mathbf{a}| = |c||\mathbf{a}|$, and $c\mathbf{a} = \mathbf{0}$ if $\mathbf{a} = \mathbf{0}$ or $c = 0$ (or both). (See Fig. 175.)

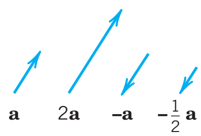


Fig. 175. Scalar multiplication [multiplication of vectors by scalars (numbers)]

Basic Properties of Scalar Multiplication. From the definitions we obtain directly

$$
\textbf{(6)}\qquad
\begin{aligned}
&\text{(a)} && c(\mathbf{a} + \mathbf{b}) = c\mathbf{a} + c\mathbf{b} &&\\
&\text{(b)} && (c + k)\mathbf{a} = c\mathbf{a} + k\mathbf{a} &&\\
&\text{(c)} && c(k\mathbf{a}) = (ck)\mathbf{a} && \text{(written } ck\mathbf{a}\text{)}\\
&\text{(d)} && 1\mathbf{a} = \mathbf{a}. &&
\end{aligned}
$$

You may prove that (4) and (6) imply for any vector **a**

$$\textbf{(7)} \qquad \begin{aligned} &\text{(a)} \quad 0\mathbf{a} = \mathbf{0} \\ &\text{(b)} \quad (-1)\mathbf{a} = -\mathbf{a}. \end{aligned}$$

Instead of $\mathbf{b} + (-\mathbf{a})$ we simply write $\mathbf{b} - \mathbf{a}$ (Fig. 176).

EXAMPLE 2 Vector Addition. Multiplication by Scalars

With respect to a given coordinate system, let

$$\mathbf{a} = [4, 0, 1] \qquad \text{and} \qquad \mathbf{b} = [2, -5, \tfrac{1}{3}].$$

Then $-\mathbf{a} = [-4, 0, -1]$, $7\mathbf{a} = [28, 0, 7]$, $\mathbf{a} + \mathbf{b} = [6, -5, \tfrac{4}{3}]$, and

$$2(\mathbf{a} - \mathbf{b}) = 2[2, 5, \tfrac{2}{3}] = [4, 10, \tfrac{4}{3}] = 2\mathbf{a} - 2\mathbf{b}.$$

Unit Vectors i, j, k. Besides $\mathbf{a} = [a_1, a_2, a_3]$ another popular way of writing vectors is

$$\textbf{(8)} \qquad \mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}.$$

In this representation, **i**, **j**, **k** are the unit vectors in the positive directions of the axes of a Cartesian coordinate system (Fig. 177). Hence, in components,

$$\textbf{(9)} \qquad \mathbf{i} = [1, 0, 0], \qquad \mathbf{j} = [0, 1, 0], \qquad \mathbf{k} = [0, 0, 1]$$

and the right side of (8) is a sum of three vectors parallel to the three axes.

EXAMPLE 3 ijk Notation for Vectors

In Example 2 we have $\mathbf{a} = 4\mathbf{i} + \mathbf{k}$, $\mathbf{b} = 2\mathbf{i} - 5\mathbf{j} + \tfrac{1}{3}\mathbf{k}$, and so on.

All the vectors $\mathbf{a} = [a_1, a_2, a_3] = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ (with real numbers as components) form the **real vector space** R^3 with the two *algebraic operations* of vector addition and scalar multiplication as just defined. R^3 has **dimension** 3. The triple of vectors **i**, **j**, **k** is called a **standard basis** of R^3. Given a Cartesian coordinate system, the representation (8) of a given vector is unique.

Vector space R^3 is a model of a general vector space, as discussed in Sec. 7.9, but is not needed in this chapter.

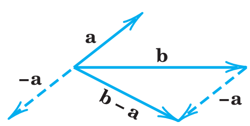


Fig. 176. Difference of vectors

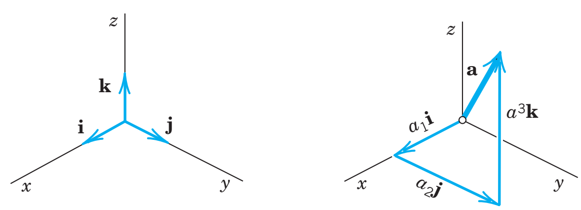


Fig. 177. The unit vectors **i**, **j**, **k** and the representation (8)

PROBLEM SET 9.1

1–5 COMPONENTS AND LENGTH

Find the components of the vector $\mathbf{v}$ with initial point P and terminal point Q. Find $|\mathbf{v}|$. Sketch $|\mathbf{v}|$. Find the unit vector $\mathbf{u}$ in the direction of $\mathbf{v}$.

1. P: (1, 1, 0), Q: (6, 2, 0)
2. P: (1, 1, 1), Q: (2, 2, 0)
3. P: (−3.0, 4,0, −0.5), Q: (5.5, 0, 1.2)
4. P: (1, 4, 2), Q: (−1, −4, −2)
5. P: (0, 0, 0), Q: (2, 1, −2)

6–10

Find the terminal point Q of the vector $\mathbf{v}$ with components as given and initial point P. Find $|\mathbf{v}|$.

6. 4, 0, 0; P: (0, 2, 13)
7. $\frac{1}{2}, 3, -\frac{1}{4}$; P: $(\frac{7}{2}, -3, \frac{3}{4})$
8. 13.1, 0.8, −2.0; P: (0, 0, 0)
9. 6, 1, −4; P: (−6, −1, −4)
10. 0, −3, 3; P: (0, 3, −3)

11–18 ADDITION, SCALAR MULTIPLICATION

Let $\mathbf{a} = [3, 2, 0] = 3\mathbf{i} + 2\mathbf{j}$; $\mathbf{b} = [-4, 6, 0] = 4\mathbf{i} + 6\mathbf{j}$, $\mathbf{c} = [5, -1, 8] = 5\mathbf{i} - \mathbf{j} + 8\mathbf{k}$, $\mathbf{d} = [0, 0, 4] = 4\mathbf{k}$.
Find:

11. $2\mathbf{a}$, $\frac{1}{2}\mathbf{a}$, $-\mathbf{a}$
12. $(\mathbf{a} + \mathbf{b}) + \mathbf{c}$, $\mathbf{a} + (\mathbf{b} + \mathbf{c})$
13. $\mathbf{b} + \mathbf{c}$, $\mathbf{c} + \mathbf{b}$
14. $3\mathbf{c} - 6\mathbf{d}$, $3(\mathbf{c} - 2\mathbf{d})$
15. $7(\mathbf{c} - \mathbf{b})$, $7\mathbf{c} - 7\mathbf{b}$
16. $\frac{9}{2}\mathbf{a} - 3\mathbf{c}$, $9(\frac{1}{2}\mathbf{a} - \frac{1}{3}\mathbf{c})$
17. $(7 - 3)\mathbf{a}$, $7\mathbf{a} - 3\mathbf{a}$
18. $4\mathbf{a} + 3\mathbf{b}$, $-4\mathbf{a} - 3\mathbf{b}$
19. What laws do Probs. 12–16 illustrate?
20. Prove Eqs. (4) and (6).

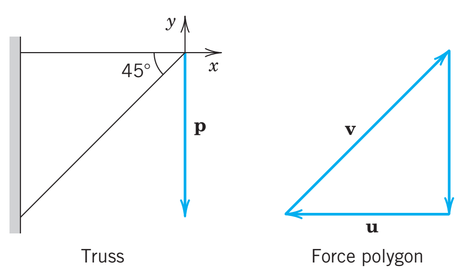
21–25 FORCES, RESULTANT

Find the resultant in terms of components and its magnitude.

21. $\mathbf{p} = [2, 3, 0]$, $\mathbf{q} = [0, 6, 1]$, $\mathbf{u} = [2, 0, -4]$
22. $\mathbf{p} = [1, -2, 3]$, $\mathbf{q} = [3, 21, -16]$, $\mathbf{u} = [-4, -19, 13]$
23. $\mathbf{u} = [8, -1, 0]$, $\mathbf{v} = [\frac{1}{2}, 0, \frac{4}{3}]$, $\mathbf{w} = [-\frac{17}{2}, 1, \frac{11}{3}]$
24. $\mathbf{p} = [-1, 2, -3]$, $\mathbf{q} = [1, 1, 1]$, $\mathbf{u} = [1, -2, 2]$
25. $\mathbf{u} = [3, 1, -6]$, $\mathbf{v} = [0, 2, 5]$, $\mathbf{w} = [3, -1, -13]$

26–37 FORCES, VELOCITIES

26. **Equilibrium.** Find $\mathbf{v}$ such that $\mathbf{p}$, $\mathbf{q}$, $\mathbf{u}$ in Prob. 21 and $\mathbf{v}$ are in equilibrium.
27. Find $\mathbf{p}$ such that $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ in Prob. 23 and $\mathbf{p}$ are in equilibrium.
28. **Unit vector.** Find the unit vector in the direction of the resultant in Prob. 24.
29. **Restricted resultant.** Find all $\mathbf{v}$ such that the resultant of $\mathbf{v}$, $\mathbf{p}$, $\mathbf{q}$, $\mathbf{u}$ with $\mathbf{p}$, $\mathbf{q}$, $\mathbf{u}$ as in Prob. 21 is parallel to the xy-plane.
30. Find $\mathbf{v}$ such that the resultant of $\mathbf{p}$, $\mathbf{q}$, $\mathbf{u}$, $\mathbf{v}$ with $\mathbf{p}$, $\mathbf{q}$, $\mathbf{u}$ as in Prob. 24 has no components in x- and y-directions.
31. For what k is the resultant of $[2, 0, -7]$, $[1, 2, -3]$, and $[0, 3, k]$ parallel to the xy-plane?
32. If $|\mathbf{p}| = 6$ and $|\mathbf{q}| = 4$, what can you say about the magnitude and direction of the resultant? Can you think of an application to robotics?
33. Same question as in Prob. 32 if $|\mathbf{p}| = 9$, $|\mathbf{q}| = 6$, $|\mathbf{u}| = 3$.
34. **Relative velocity.** If airplanes A and B are moving southwest with speed $|\mathbf{v}_A| = 550$ mph, and northwest with speed $|\mathbf{v}_B| = 450$ mph, respectively, what is the relative velocity $\mathbf{v} = \mathbf{v}_B - \mathbf{v}_A$ of B with respect to A?
35. Same question as in Prob. 34 for two ships moving northeast with speed $|\mathbf{v}_A| = 22$ knots and west with speed $|\mathbf{v}_B| = 19$ knots.
36. **Reflection.** If a ray of light is reflected once in each of two mutually perpendicular mirrors, what can you say about the reflected ray?
37. **Force polygon. Truss.** Find the forces in the system of two rods (*truss*) in the figure, where $|\mathbf{p}| = 1000$ nt. *Hint.* Forces in equilibrium form a polygon, the *force polygon*.



Problem 37

38. TEAM PROJECT. Geometric Applications. To increase your skill in dealing with vectors, use vectors to prove the following (see the figures).

(a) The diagonals of a parallelogram bisect each other.

(b) The line through the midpoints of adjacent sides of a parallelogram bisects one of the diagonals in the ratio 1 : 3.

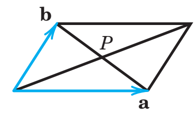
(c) Obtain (b) from (a).

(d) The three medians of a triangle (the segments from a vertex to the midpoint of the opposite side) meet at a single point, which divides the medians in the ratio 2 : 1.

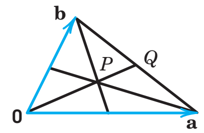
(e) The quadrilateral whose vertices are the midpoints of the sides of an arbitrary quadrilateral is a parallelogram.

(f) The four space diagonals of a parallelepiped meet and bisect each other.

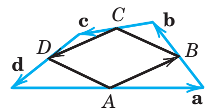
(g) The sum of the vectors drawn from the center of a regular polygon to its vertices is the zero vector.



Team Project 38(a)



Team Project 38(d)



Team Project 38(e)

9.2 Inner Product (Dot Product)

Orthogonality

The inner product or dot product can be motivated by calculating work done by a constant force, determining components of forces, or other applications. It involves the length of vectors and the angle between them. The inner product is a kind of multiplication of two vectors, defined in such a way that the outcome is a scalar. Indeed, another term for inner product is scalar product, a term we shall not use here. The definition of the inner product is as follows.

DEFINITION

Inner Product (Dot Product) of Vectors

The **inner product** or **dot product** $\mathbf{a} \bullet \mathbf{b}$ (read "**a** dot **b**") of two vectors **a** and **b** is the product of their lengths times the cosine of their angle (see Fig. 178),

$$\text{(1)} \qquad \begin{aligned} &\mathbf{a} \bullet \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos \gamma \qquad &&\text{if} \quad \mathbf{a} \neq \mathbf{0}, \mathbf{b} \neq \mathbf{0} \\ &\mathbf{a} \bullet \mathbf{b} = 0 &&\text{if} \quad \mathbf{a} = \mathbf{0} \text{ or } \mathbf{b} = \mathbf{0}. \end{aligned}$$

The angle γ, $0 \leqq \gamma \leqq \pi$, between **a** and **b** is measured when the initial points of the vectors coincide, as in Fig. 178. In components, $\mathbf{a} = [a_1, a_2, a_3]$, $\mathbf{b} = [b_1, b_2, b_3]$, and

$$\text{(2)} \qquad \mathbf{a} \bullet \mathbf{b} = a_1 b_1 + a_2 b_2 + a_3 b_3.$$

The second line in (1) is needed because γ is undefined when $\mathbf{a} = \mathbf{0}$ or $\mathbf{b} = \mathbf{0}$. The derivation of (2) from (1) is shown below.

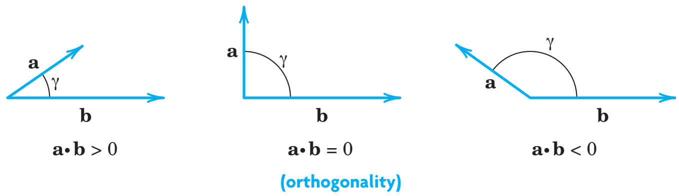


Fig. 178. Angle between vectors and value of inner product

Orthogonality. Since the cosine in (1) may be positive, 0, or negative, so may be the inner product (Fig. 178). The case that the inner product is zero is of particular practical interest and suggests the following concept.

A vector $\mathbf{a}$ is called **orthogonal** to a vector $\mathbf{b}$ if $\mathbf{a} \cdot \mathbf{b} = 0$. Then $\mathbf{b}$ is also orthogonal to $\mathbf{a}$, and we call $\mathbf{a}$ and $\mathbf{b}$ **orthogonal vectors**. Clearly, this happens for nonzero vectors if and only if $\cos \gamma = 0$; thus $\gamma = \pi/2$ (90°). This proves the important

THEOREM 1

Orthogonality Criterion

The inner product of two nonzero vectors is 0 *if and only if these vectors are perpendicular.*

Length and Angle. Equation (1) with $\mathbf{b} = \mathbf{a}$ gives $\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2$. Hence

$$|\mathbf{a}| = \sqrt{\mathbf{a} \cdot \mathbf{a}}. \tag{3}$$

From (3) and (1) we obtain for the angle γ between two nonzero vectors

$$\cos \gamma = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}||\mathbf{b}|} = \frac{\mathbf{a} \cdot \mathbf{b}}{\sqrt{\mathbf{a} \cdot \mathbf{a}}\sqrt{\mathbf{b} \cdot \mathbf{b}}}. \tag{4}$$

EXAMPLE 1 **Inner Product. Angle Between Vectors**

Find the inner product and the lengths of $\mathbf{a} = [1, 2, 0]$ and $\mathbf{b} = [3, -2, 1]$ as well as the angle between these vectors.

Solution. $\mathbf{a} \cdot \mathbf{b} = 1 \cdot 3 + 2 \cdot (-2) + 0 \cdot 1 = -1$, $|\mathbf{a}| = \sqrt{\mathbf{a} \cdot \mathbf{a}} = \sqrt{5}$, $|\mathbf{b}| = \sqrt{\mathbf{b} \cdot \mathbf{b}} = \sqrt{14}$, and (4) gives the angle

$$\gamma = \arccos \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}||\mathbf{b}|} = \arccos(-0.11952) = 1.69061 = 96.865°. \qquad \blacksquare$$

From the definition we see that the inner product has the following properties. For any vectors **a**, **b**, **c** and scalars q_1, q_2,

$$\textbf{(5)} \quad \begin{array}{lll} \text{(a)} & (q_1\mathbf{a} + q_2\mathbf{b}) \cdot \mathbf{c} = q_1\mathbf{a} \cdot \mathbf{c} + q_1\mathbf{b} \cdot \mathbf{c} & (\textit{Linearity}) \\ \text{(b)} & \mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a} & (\textit{Symmetry}) \\ \text{(c)} & \left.\begin{array}{c} \mathbf{a} \cdot \mathbf{a} \geqq 0 \\ \mathbf{a} \cdot \mathbf{a} = 0 \quad \text{if and only if} \quad \mathbf{a} = \mathbf{0} \end{array}\right\} & (\textit{Positive-definiteness}). \end{array}$$

Hence *dot multiplication is commutative* as shown by (5b). Furthermore, it is *distributive with respect to vector addition.* This follows from (5a) with $q_1 = 1$ and $q_2 = 1$:

$$\textbf{(5a*)} \qquad (\mathbf{a} + \mathbf{b}) \cdot \mathbf{c} = \mathbf{a} \cdot \mathbf{c} + \mathbf{b} \cdot \mathbf{c} \qquad (\textit{Distributivity}).$$

Furthermore, from (1) and $|\cos \gamma| \leqq 1$ we see that

$$\textbf{(6)} \qquad |\mathbf{a} \cdot \mathbf{b}| \leqq |\mathbf{a}||\mathbf{b}| \qquad (\textit{Cauchy–Schwarz inequality}).$$

Using this and (3), you may prove (see Prob. 16)

$$\textbf{(7)} \qquad |\mathbf{a} + \mathbf{b}| \leqq |\mathbf{a}| + |\mathbf{b}| \qquad (\textbf{Triangle inequality}).$$

Geometrically, (7) with $<$ says that one side of a triangle must be shorter than the other two sides together; this motivates the name of (7).

A simple direct calculation with inner products shows that

$$(8) \qquad |\mathbf{a} + \mathbf{b}|^2 + |\mathbf{a} - \mathbf{b}|^2 = 2(|\mathbf{a}|^2 + |\mathbf{b}|^2) \quad (\textit{Parallelogram equality}).$$

Equations (6)–(8) play a basic role in so-called *Hilbert spaces*, which are abstract inner product spaces. Hilbert spaces form the basis of quantum mechanics, for details see [GenRef7] listed in App. 1.

Derivation of (2) from (1). We write $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ and $\mathbf{b} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$, as in (8) of Sec. 9.1. If we substitute this into $\mathbf{a} \cdot \mathbf{b}$ and use (5a*), we first have a sum of $3 \times 3 = 9$ products

$$\mathbf{a} \cdot \mathbf{b} = a_1b_1\mathbf{i} \cdot \mathbf{i} + a_1b_2\mathbf{i} \cdot \mathbf{j} + \cdots + a_3b_3\mathbf{k} \cdot \mathbf{k}.$$

Now **i**, **j**, **k** are unit vectors, so that $\mathbf{i} \cdot \mathbf{i} = \mathbf{j} \cdot \mathbf{j} = \mathbf{k} \cdot \mathbf{k} = 1$ by (3). Since the coordinate axes are perpendicular, so are **i**, **j**, **k**, and Theorem 1 implies that the other six of those nine products are 0, namely, $\mathbf{i} \cdot \mathbf{j} = \mathbf{j} \cdot \mathbf{i} = \mathbf{j} \cdot \mathbf{k} = \mathbf{k} \cdot \mathbf{j} = \mathbf{k} \cdot \mathbf{i} = \mathbf{i} \cdot \mathbf{k} = 0$. But this reduces our sum for $\mathbf{a} \cdot \mathbf{b}$ to (2). ■

Applications of Inner Products

Typical applications of inner products are shown in the following examples and in Problem Set 9.2.

EXAMPLE 2 Work Done by a Force Expressed as an Inner Product

This is a major application. It concerns a body on which a *constant* force $\mathbf{p}$ acts. (For a *variable* force, see Sec. 10.1.) Let the body be given a displacement $\mathbf{d}$. Then the work done by $\mathbf{p}$ in the displacement is defined as

$$W = |\mathbf{p}||\mathbf{d}| \cos \alpha = \mathbf{p} \cdot \mathbf{d}, \tag{9}$$

that is, magnitude $|\mathbf{p}|$ of the force times length $|\mathbf{d}|$ of the displacement times the cosine of the angle α between $\mathbf{p}$ and $\mathbf{d}$ (Fig. 179). If $\alpha < 90°$, as in Fig. 179, then $W > 0$. If $\mathbf{p}$ and $\mathbf{d}$ are orthogonal, then the work is zero (why?). If $\alpha > 90°$, then $W < 0$, which means that in the displacement one has to do work against the force. For example, think of swimming across a river at some angle α against the current.

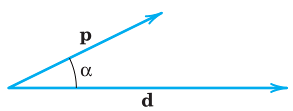


Fig. 179. Work done by a force

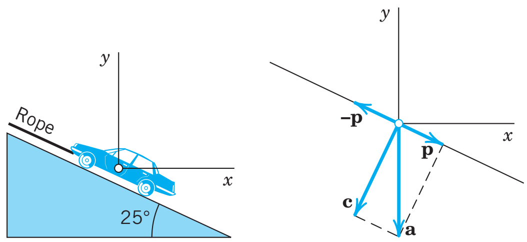


Fig. 180. Example 3

EXAMPLE 3 Component of a Force in a Given Direction

What force in the rope in Fig. 180 will hold a car of 5000 lb in equilibrium if the ramp makes an angle of 25° with the horizontal?

Solution. Introducing coordinates as shown, the weight is $\mathbf{a} = [0, -5000]$ because this force points downward, in the negative y-direction. We have to represent $\mathbf{a}$ as a sum (resultant) of two forces, $\mathbf{a} = \mathbf{c} + \mathbf{p}$, where $\mathbf{c}$ is the force the car exerts on the ramp, which is of no interest to us, and $\mathbf{p}$ is parallel to the rope. A vector in the direction of the rope is (see Fig. 180)

$$\mathbf{b} = [-1, \tan 25°] = [-1, 0.46631], \qquad \text{thus} \qquad |\mathbf{b}| = 1.10338,$$

The direction of the unit vector $\mathbf{u}$ is opposite to the direction of the rope so that

$$\mathbf{u} = -\frac{1}{|\mathbf{b}|}\mathbf{b} = [0.90631, -0.42262].$$

Since $|\mathbf{u}| = 1$ and $\cos \gamma > 0$, we see that we can write our result as

$$|\mathbf{p}| = (|\mathbf{a}| \cos \gamma)|\mathbf{u}| = \mathbf{a} \cdot \mathbf{u} = -\frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{b}|} = \frac{5000 \cdot 0.46631}{1.10338} = 2113 \text{ [lb]}.$$

We can also note that $\gamma = 90° - 25° = 65°$ is the angle between $\mathbf{a}$ and $\mathbf{p}$ so that

$$|\mathbf{p}| = |\mathbf{a}| \cos \gamma = 5000 \cos 65° = 2113 \text{ [lb]}.$$

Answer: About 2100 lb. ■

Example 3 is typical of applications that deal with the **component** or **projection** *of a vector* $\mathbf{a}$ *in the direction of a vector* $\mathbf{b}\,(\neq \mathbf{0})$. If we denote by p the length of the orthogonal projection of $\mathbf{a}$ on a straight line l parallel to $\mathbf{b}$ as shown in Fig. 181, then

$$p = |\mathbf{a}| \cos \gamma. \tag{10}$$

Here p is taken with the plus sign if $p\mathbf{b}$ has the direction of $\mathbf{b}$ and with the minus sign if $p\mathbf{b}$ has the direction opposite to $\mathbf{b}$.

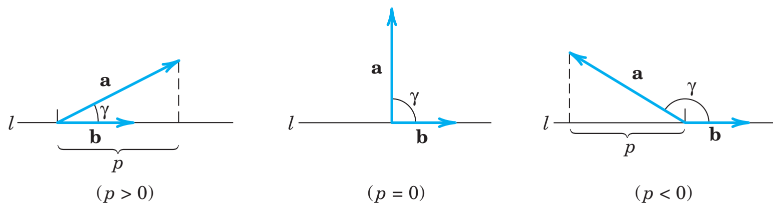


Fig. 181. Component of a vector **a** in the direction of a vector **b**

Multiplying (10) by $|\mathbf{b}|/|\mathbf{b}| = 1$, we have $\mathbf{a} \bullet \mathbf{b}$ in the numerator and thus

$$p = \frac{\mathbf{a} \bullet \mathbf{b}}{|\mathbf{b}|} \qquad (\mathbf{b} \neq \mathbf{0}). \tag{11}$$

If $\mathbf{b}$ is a unit vector, as it is often used for fixing a direction, then (11) simply gives

$$p = \mathbf{a} \bullet \mathbf{b} \qquad (|\mathbf{b}| = 1). \tag{12}$$

Figure 182 shows the projection p of $\mathbf{a}$ in the direction of $\mathbf{b}$ (as in Fig. 181) and the projection $q = |\mathbf{b}| \cos \gamma$ of $\mathbf{b}$ in the direction of $\mathbf{a}$.

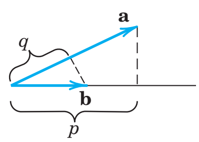


Fig. 182. Projections p of **a** on **b** and q of **b** on **a**

EXAMPLE 4 Orthonormal Basis

By definition, an *orthonormal basis* for 3-space is a basis $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ consisting of orthogonal unit vectors. It has the great advantage that the determination of the coefficients in representations $\mathbf{v} = l_1\mathbf{a} + l_2\mathbf{b} + l_3\mathbf{c}$ of a given vector $\mathbf{v}$ is very simple. We claim that $l_1 = \mathbf{a} \bullet \mathbf{v}$, $l_2 = \mathbf{b} \bullet \mathbf{v}$, $l_3 = \mathbf{c} \bullet \mathbf{v}$. Indeed, this follows simply by taking the inner products of the representation with $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, respectively, and using the orthonormality of the basis, $\mathbf{a} \bullet \mathbf{v} = l_1\mathbf{a} \bullet \mathbf{a} + l_2\mathbf{a} \bullet \mathbf{b} + l_3\mathbf{a} \bullet \mathbf{c} = l_1$, etc.

For example, the unit vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ in (8), Sec. 9.1, associated with a Cartesian coordinate system form an orthonormal basis, called the **standard basis** with respect to the given coordinate system. ■

EXAMPLE 5 **Orthogonal Straight Lines in the Plane**

Find the straight line L_1 through the point P: (1, 3) in the xy-plane and perpendicular to the straight line $L_2: x - 2y + 2 = 0$; see Fig. 183.

Solution. The idea is to write a general straight line $L_1: a_1x + a_2y = c$ as $\mathbf{a} \cdot \mathbf{r} = c$ with $\mathbf{a} = [a_1, a_2] \neq \mathbf{0}$ and $\mathbf{r} = [x, y]$, according to (2). Now the line L_1^* through the origin and parallel to L_1 is $\mathbf{a} \cdot \mathbf{r} = 0$. Hence, by Theorem 1, the vector $\mathbf{a}$ is perpendicular to $\mathbf{r}$. Hence it is perpendicular to L_1^* and also to L_1 because L_1 and L_1^* are parallel. $\mathbf{a}$ is called a **normal vector** of L_1 (and of L_1^*).

Now a normal vector of the given line $x - 2y + 2 = 0$ is $\mathbf{b} = [1, -2]$. Thus L_1 is perpendicular to L_2 if $\mathbf{b} \cdot \mathbf{a} = a_1 - 2a_2 = 0$, for instance, if $\mathbf{a} = [2, 1]$. Hence L_1 is given by $2x + y = c$. It passes through P: (1, 3) when $2 \cdot 1 + 3 = c = 5$. *Answer:* $y = -2x + 5$. Show that the point of intersection is $(x, y) = (1.6, 1.8)$. ■

EXAMPLE 6 **Normal Vector to a Plane**

Find a unit vector perpendicular to the plane $4x + 2y + 4z = -7$.

Solution. Using (2), we may write any plane in space as

$$\mathbf{a} \cdot \mathbf{r} = a_1x + a_2y + a_3z = c \tag{13}$$

where $\mathbf{a} = [a_1, a_2, a_3] \neq \mathbf{0}$ and $\mathbf{r} = [x, y, z]$. The unit vector in the direction of $\mathbf{a}$ is (Fig. 184)

$$\mathbf{n} = \frac{1}{|\mathbf{a}|}\mathbf{a}.$$

Dividing by $|\mathbf{a}|$, we obtain from (13)

$$\mathbf{n} \cdot \mathbf{r} = p \qquad \text{where} \qquad p = \frac{c}{|\mathbf{a}|}. \tag{14}$$

From (12) we see that p is the projection of $\mathbf{r}$ in the direction of $\mathbf{n}$. This projection has the same constant value $c/|\mathbf{a}|$ for the position vector $\mathbf{r}$ of any point in the plane. Clearly this holds if and only if $\mathbf{n}$ is perpendicular to the plane. $\mathbf{n}$ is called a **unit normal vector** of the plane (the other being $-\mathbf{n}$).

Furthermore, from this and the definition of projection, it follows that $|p|$ is the distance of the plane from the origin. Representation (14) is called **Hesse's[2] normal form** of a plane. In our case, $\mathbf{a} = [4, 2, 4]$, $c = -7$, $|\mathbf{a}| = 6$, $\mathbf{n} = \frac{1}{6}\mathbf{a} = [\frac{2}{3}, \frac{1}{3}, \frac{2}{3}]$, and the plane has the distance $\frac{7}{6}$ from the origin. ■

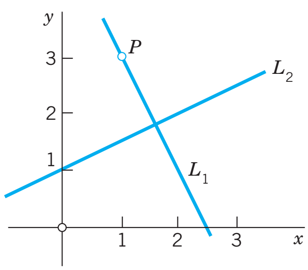


Fig. 183. Example 5

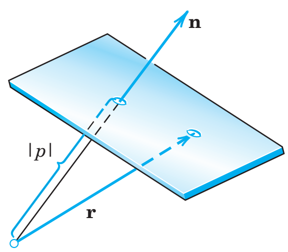


Fig. 184. Normal vector to a plane

[2]LUDWIG OTTO HESSE (1811–1874), German mathematician who contributed to the theory of curves and surfaces.

9.3 Vector Product (Cross Product)

We shall define another form of multiplication of vectors, inspired by applications, whose result will be a *vector*. This is in contrast to the dot product of Sec. 9.2 where multiplication resulted in a *scalar*. We can construct a vector $\mathbf{v}$ that is perpendicular to two vectors $\mathbf{a}$ and $\mathbf{b}$, which are two sides of a parallelogram on a plane in space as indicated in Fig. 185, such that the length $|\mathbf{v}|$ is numerically equal to the area of that parallelogram. Here then is the new concept.

DEFINITION

Vector Product (Cross Product, Outer Product) of Vectors

The **vector product** or **cross product** $\mathbf{a} \times \mathbf{b}$ (read "$\mathbf{a}$ cross $\mathbf{b}$") of two vectors $\mathbf{a}$ and $\mathbf{b}$ is the vector $\mathbf{v}$ denoted by

$$\mathbf{v} = \mathbf{a} \times \mathbf{b}$$

I. If $\mathbf{a} = \mathbf{0}$ or $\mathbf{b} = \mathbf{0}$, then we define $\mathbf{v} = \mathbf{a} \times \mathbf{b} = \mathbf{0}$.

II. If both vectors are nonzero vectors, then vector $\mathbf{v}$ has the length

$$|\mathbf{v}| = |\mathbf{a} \times \mathbf{b}| = |\mathbf{a}||\mathbf{b}| \sin \gamma, \tag{1}$$

where γ is the angle between $\mathbf{a}$ and $\mathbf{b}$ as in Sec. 9.2.

Furthermore, by design, $\mathbf{a}$ and $\mathbf{b}$ form the sides of a parallelogram on a plane in space. The parallelogram is shaded in blue in Fig. 185. The area of this blue parallelogram is precisely given by Eq. (1), so that the length $|\mathbf{v}|$ of the vector $\mathbf{v}$ is equal to the area of that parallelogram.

III. If $\mathbf{a}$ and $\mathbf{b}$ lie in the same straight line, i.e., $\mathbf{a}$ and $\mathbf{b}$ have the same or opposite directions, then γ is 0° or 180° so that $\sin \gamma = 0$. In that case $|\mathbf{v}| = 0$ so that $\mathbf{v} = \mathbf{a} \times \mathbf{b} = \mathbf{0}$.

IV. If cases I and III do not occur, then $\mathbf{v}$ is a nonzero vector. The direction of $\mathbf{v} = \mathbf{a} \times \mathbf{b}$ is perpendicular to both $\mathbf{a}$ and $\mathbf{b}$ such that $\mathbf{a}$, $\mathbf{b}$, $\mathbf{v}$—precisely in this order (!)—form a right-handed triple as shown in Figs. 185–187 and explained below.

Another term for vector product is outer product.

Remark. Note that I and III completely characterize the exceptional case when the cross product is equal to the zero vector, and II and IV the regular case where the cross product is perpendicular to two vectors.

Just as we did with the dot product, we would also like to express the cross product in components. Let $\mathbf{a} = [a_1, a_2, a_3]$ and $\mathbf{b} = [b_1, b_2, b_3]$. Then $\mathbf{v} = [v_1, v_2, v_3] = \mathbf{a} \times \mathbf{b}$ has the components

$$v_1 = a_2b_3 - a_3b_2, \qquad v_2 = a_3b_1 - a_1b_3, \qquad v_3 = a_1b_2 - a_2b_1. \tag{2}$$

Here the Cartesian coordinate system is *right-handed*, as explained below (see also Fig. 188). (For a left-handed system, each component of $\mathbf{v}$ must be multiplied by -1. Derivation of (2) in App. 4.)

Right-Handed Triple. A triple of vectors **a**, **b**, **v** is *right-handed* if the vectors in the given order assume the same sort of orientation as the thumb, index finger, and middle finger of the right hand when these are held as in Fig. 186. We may also say that if **a** is rotated into the direction of **b** through the angle $\gamma\,(<\pi)$, then **v** advances in the same direction as a right-handed screw would if turned in the same way (Fig. 187).

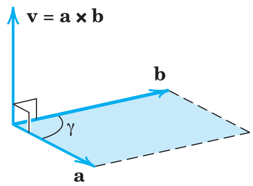


Fig. 185. Vector product

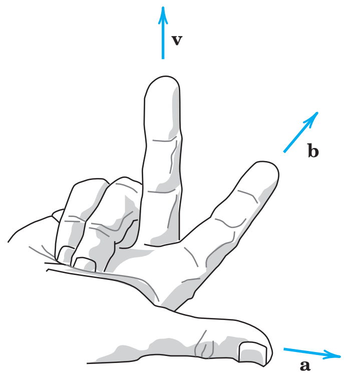


Fig. 186. Right-handed triple of vectors **a, b, v**

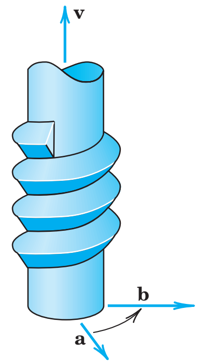
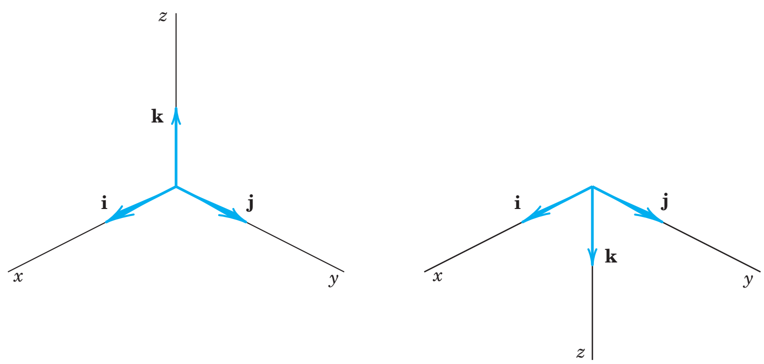


Fig. 187. Right-handed screw

Right-Handed Cartesian Coordinate System. The system is called **right-handed** if the corresponding unit vectors **i**, **j**, **k** in the positive directions of the axes (see Sec. 9.1) form a right-handed triple as in Fig. 188a. The system is called **left-handed** if the sense of **k** is reversed, as in Fig. 188b. In applications, we prefer right-handed systems.



(a) Right-handed (b) Left-handed

Fig. 188. The two types of Cartesian coordinate systems

How to Memorize (2). If you know second- and third-order determinants, you see that (2) can be written

$$(2^*)\qquad v_1 = \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}, \qquad v_2 = -\begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} = +\begin{vmatrix} a_3 & a_1 \\ b_3 & b_1 \end{vmatrix}, \qquad v_3 = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$$

and $\mathbf{v} = [v_1, v_2, v_3] = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$ is the expansion of the following symbolic determinant by its first row. (We call the determinant "symbolic" because the first row consists of vectors rather than of numbers.)

$$(2^{**}) \qquad \mathbf{v} = \mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} \mathbf{i} - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} \mathbf{j} + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \mathbf{k}.$$

For a left-handed system the determinant has a minus sign in front.

EXAMPLE 1 Vector Product

For the vector product $\mathbf{v} = \mathbf{a} \times \mathbf{b}$ of $\mathbf{a} = [1, 1, 0]$ and $\mathbf{b} = [3, 0, 0]$ in right-handed coordinates we obtain from (2)

$$v_1 = 0, \qquad v_2 = 0, \qquad v_3 = 1 \cdot 0 - 1 \cdot 3 = -3.$$

We confirm this by (2**):

$$\mathbf{v} = \mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 1 & 0 \\ 3 & 0 & 0 \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & 0 \end{vmatrix} \mathbf{i} - \begin{vmatrix} 1 & 0 \\ 3 & 0 \end{vmatrix} \mathbf{j} + \begin{vmatrix} 1 & 1 \\ 3 & 0 \end{vmatrix} \mathbf{k} = -3\mathbf{k} = [0, 0, -3].$$

To check the result in this simple case, sketch $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{v}$. Can you see that two vectors in the xy-plane must always have their vector product parallel to the z-axis (or equal to the zero vector)? ■

EXAMPLE 2 Vector Products of the Standard Basis Vectors

$$(3) \qquad \begin{aligned} \mathbf{i} \times \mathbf{j} &= \mathbf{k}, & \mathbf{j} \times \mathbf{k} &= \mathbf{i}, & \mathbf{k} \times \mathbf{i} &= \mathbf{j} \\ \mathbf{j} \times \mathbf{i} &= -\mathbf{k}, & \mathbf{k} \times \mathbf{j} &= -\mathbf{i}, & \mathbf{i} \times \mathbf{k} &= -\mathbf{j}. \end{aligned}$$

We shall use this in the next proof. ■

THEOREM 1 General Properties of Vector Products

(a) *For every scalar l,*

$$(4) \qquad (l\mathbf{a}) \times \mathbf{b} = l(\mathbf{a} \times \mathbf{b}) = \mathbf{a} \times (l\mathbf{b}).$$

(b) *Cross multiplication is distributive with respect to vector addition; that is,*

$$(5) \qquad \begin{aligned} &(\alpha) \quad \mathbf{a} \times (\mathbf{b} + \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) + (\mathbf{a} \times \mathbf{c}), \\ &(\beta) \quad (\mathbf{a} + \mathbf{b}) \times \mathbf{c} = (\mathbf{a} \times \mathbf{c}) + (\mathbf{b} \times \mathbf{c}). \end{aligned}$$

(c) *Cross multiplication is **not commutative** but **anticommutative;** that is,*

$$(6) \qquad \boxed{\mathbf{b} \times \mathbf{a} = -(\mathbf{a} \times \mathbf{b})} \qquad \text{(Fig. 189).}$$

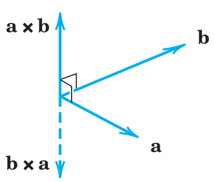


Fig. 189. Anticommutativity of cross multiplication

(d) *Cross multiplication is* ***not associative***; *that is, in general,*

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) \neq (\mathbf{a} \times \mathbf{b}) \times \mathbf{c} \tag{7}$$

so that the parentheses cannot be omitted.

PROOF Equation (4) follows directly from the definition. In (5α), formula (2*) gives for the first component on the left

$$\begin{aligned}\begin{vmatrix} a_2 & a_3 \\ b_2 + c_2 & b_3 + c_3 \end{vmatrix} &= a_2(b_3 + c_3) - a_3(b_2 + c_2) \\ &= (a_2b_3 - a_3b_2) + (a_2c_3 - a_3c_2) \\ &= \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} + \begin{vmatrix} a_2 & a_3 \\ c_2 & c_3 \end{vmatrix}.\end{aligned}$$

By (2*) the sum of the two determinants is the first component of $(\mathbf{a} \times \mathbf{b}) + (\mathbf{a} \times \mathbf{c})$, the right side of (5α). For the other components in (5α) and in 5(β), equality follows by the same idea.

Anticommutativity (6) follows from (2**) by noting that the interchange of Rows 2 and 3 multiplies the determinant by -1. We can confirm this geometrically if we set $\mathbf{a} \times \mathbf{b} = \mathbf{v}$ and $\mathbf{b} \times \mathbf{a} = \mathbf{w}$; then $|\mathbf{v}| = |\mathbf{w}|$ by (1), and for $\mathbf{b}$, $\mathbf{a}$, $\mathbf{w}$ to form a *right-handed* triple, we must have $\mathbf{w} = -\mathbf{v}$.

Finally, $\mathbf{i} \times (\mathbf{i} \times \mathbf{j}) = \mathbf{i} \times \mathbf{k} = -\mathbf{j}$, whereas $(\mathbf{i} \times \mathbf{i}) \times \mathbf{j} = \mathbf{0} \times \mathbf{j} = \mathbf{0}$ (see Example 2). This proves (7). ■

Typical Applications of Vector Products

EXAMPLE 3 **Moment of a Force**

In mechanics the moment m of a force $\mathbf{p}$ about a point Q is defined as the product $m = |\mathbf{p}|d$, where d is the (perpendicular) distance between Q and the line of action L of $\mathbf{p}$ (Fig. 190). If $\mathbf{r}$ is the vector from Q to any point A on L, then $d = |\mathbf{r}| \sin \gamma$, as shown in Fig. 190, and

$$m = |\mathbf{r}||\mathbf{p}| \sin \gamma.$$

Since γ is the angle between $\mathbf{r}$ and $\mathbf{p}$, we see from (1) that $m = |\mathbf{r} \times \mathbf{p}|$. The vector

$$\mathbf{m} = \mathbf{r} \times \mathbf{p} \tag{8}$$

is called the **moment vector** or **vector moment** of $\mathbf{p}$ about Q. Its magnitude is m. If $\mathbf{m} \neq \mathbf{0}$, its direction is that of the axis of the rotation about Q that $\mathbf{p}$ has the tendency to produce. This axis is perpendicular to both $\mathbf{r}$ and $\mathbf{p}$. ■

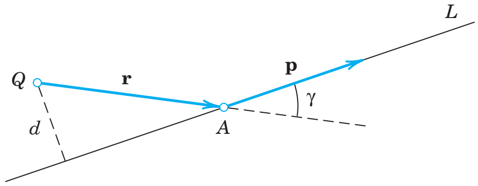


Fig. 190. Moment of a force $\mathbf{p}$

EXAMPLE 4 Moment of a Force

Find the moment of the force $\mathbf{p}$ about the center Q of a wheel, as given in Fig. 191.

Solution. Introducing coordinates as shown in Fig. 191, we have

$$\mathbf{p} = [1000 \cos 30°, \quad 1000 \sin 30°, \quad 0] = [866, \quad 500, \quad 0], \qquad \mathbf{r} = [0, \quad 1.5, \quad 0].$$

(Note that the center of the wheel is at $y = -1.5$ on the y-axis.) Hence (8) and (2**) give

$$\mathbf{m} = \mathbf{r} \times \mathbf{p} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & 1.5 & 0 \\ 866 & 500 & 0 \end{vmatrix} = 0\mathbf{i} - 0\mathbf{j} + \begin{vmatrix} 0 & 1.5 \\ 866 & 500 \end{vmatrix}\mathbf{k} = [0, 0, -1299].$$

This moment vector $\mathbf{m}$ is normal, i.e., perpendicular to the plane of the wheel. Hence it has the direction of the axis of rotation about the center Q of the wheel that the force $\mathbf{p}$ has the tendency to produce. The moment $\mathbf{m}$ points in the negative z-direction, This is, the direction in which a right-handed screw would advance if turned in that way. ■

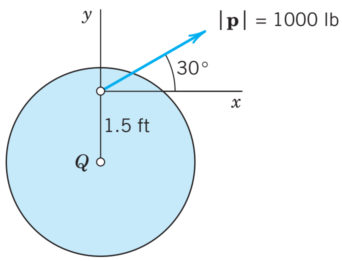


Fig. 191. Moment of a force **p**

EXAMPLE 5 Velocity of a Rotating Body

A rotation of a rigid body B in space can be simply and uniquely described by a vector $\mathbf{w}$ as follows. The direction of $\mathbf{w}$ is that of the axis of rotation and such that the rotation appears clockwise if one looks from the initial point of $\mathbf{w}$ to its terminal point. The length of $\mathbf{w}$ is equal to the **angular speed** $\omega(>0)$ of the rotation, that is, the linear (or tangential) speed of a point of B divided by its distance from the axis of rotation.

Let P be any point of B and d its distance from the axis. Then P has the speed ωd. Let $\mathbf{r}$ be the position vector of P referred to a coordinate system with origin 0 on the axis of rotation. Then $d = |\mathbf{r}| \sin \gamma$, where γ is the angle between $\mathbf{w}$ and $\mathbf{r}$. Therefore,

$$\omega d = |\mathbf{w}||\mathbf{r}| \sin \gamma = |\mathbf{w} \times \mathbf{r}|.$$

From this and the definition of vector product we see that the velocity vector $\mathbf{v}$ of P can be represented in the form (Fig. 192)

$$\mathbf{v} = \mathbf{w} \times \mathbf{r}. \tag{9}$$

This simple formula is useful for determining $\mathbf{v}$ at any point of B. ■

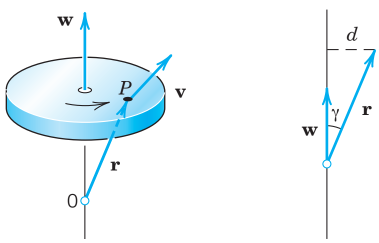


Fig. 192. Rotation of a rigid body

Scalar Triple Product

Certain products of vectors, having three or more factors, occur in applications. The most important of these products is the scalar triple product or mixed product of three vectors **a**, **b**, **c**.

$$(\mathbf{a}\quad\mathbf{b}\quad\mathbf{c}) = \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}). \tag{10*}$$

The scalar triple product is indeed a scalar since (10*) involves a dot product, which in turn is a scalar. We want to express the scalar triple product in components and as a third-order determinant. To this end, let $\mathbf{a} = [a_1, a_2, a_3]$, $\mathbf{b} = [b_1, b_2, b_3]$, and $\mathbf{c} = [c_1, c_2, c_3]$. Also set $\mathbf{b}\times\mathbf{c} = \mathbf{v} = [v_1, v_2, v_3]$. Then from the dot product in components [formula (2) in Sec. 9.2] and from (2*) with **b** and **c** instead of **a** and **b** we first obtain

$$\begin{aligned}\mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) &= \mathbf{a}\cdot\mathbf{v} = a_1v_1 + a_2v_2 + a_3v_3\\ &= a_1\begin{vmatrix} b_2 & b_3\\ c_2 & c_3\end{vmatrix} + a_2\begin{vmatrix} b_3 & b_1\\ c_3 & c_1\end{vmatrix} + a_3\begin{vmatrix} b_1 & b_2\\ c_1 & c_2\end{vmatrix}.\end{aligned}$$

The sum on the right is the expansion of a third-order determinant by its first row. Thus we obtain the desired formula for the scalar triple product, that is,

$$(\mathbf{a}\quad\mathbf{b}\quad\mathbf{c}) = \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = \begin{vmatrix} a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\\ c_1 & c_2 & c_3\end{vmatrix}. \tag{10}$$

The most important properties of the scalar triple product are as follows.

THEOREM 2

Properties and Applications of Scalar Triple Products

(a) *In* (10) *the dot and cross can be interchanged:*

$$(\mathbf{a}\quad\mathbf{b}\quad\mathbf{c}) = \mathbf{a}\cdot(\mathbf{b}\times\mathbf{c}) = (\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}. \tag{11}$$

(b) Geometric interpretation. *The absolute value* $|(\mathbf{a}\quad\mathbf{b}\quad\mathbf{c})|$ *of* (10) *is the volume of the parallelepiped* (oblique box) *with* **a**, **b**, **c** *as edge vectors* (Fig. 193).

(c) Linear independence. *Three vectors in* R^3 *are linearly independent if and only if their scalar triple product is not zero.*

PROOF **(a)** Dot multiplication is commutative, so that by (10)

$$(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c} = \mathbf{c}\cdot(\mathbf{a}\times\mathbf{b}) = \begin{vmatrix} c_1 & c_2 & c_3\\ a_1 & a_2 & a_3\\ b_1 & b_2 & b_3\end{vmatrix}.$$

From this we obtain the determinant in (10) by interchanging Rows 1 and 2 and in the result Rows 2 and 3. But this does not change the value of the determinant because each interchange produces a factor -1, and $(-1)(-1) = 1$. This proves (11).

(b) The volume of that box equals the height $h = |\mathbf{a}||\cos\gamma|$ (Fig. 193) times the area of the base, which is the area $|\mathbf{b} \times \mathbf{c}|$ of the parallelogram with sides $\mathbf{b}$ and $\mathbf{c}$. Hence the volume is

$$|\mathbf{a}||\mathbf{b} \times \mathbf{c}|\,|\cos\gamma| = |\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})| \qquad \text{(Fig. 193)}$$

as given by the absolute value of (11).

(c) Three nonzero vectors, whose initial points coincide, are linearly independent if and only if the vectors do not lie in the same plane nor lie on the same straight line.

This happens if and only if the triple product in (b) is not zero, so that the independence criterion follows. (The case of one of the vectors being the zero vector is trivial.)

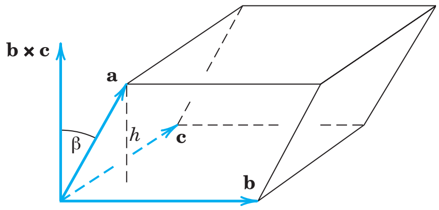


Fig. 193. Geometric interpretation of a scalar triple product

EXAMPLE 6 **Tetrahedron**

A tetrahedron is determined by three edge vectors **a**, **b**, **c**, as indicated in Fig. 194. Find the volume of the tetrahedron in Fig. 194, when $\mathbf{a} = [2, 0, 3]$, $\mathbf{b} = [0, 4, 1]$, $c = [5, 6, 0]$.

Solution. The volume V of the parallelepiped with these vectors as edge vectors is the absolute value of the scalar triple product

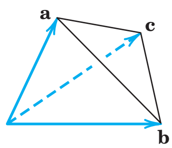


Fig. 194. Tetrahedron

$$(\mathbf{a} \quad \mathbf{b} \quad \mathbf{c}) = \begin{vmatrix} 2 & 0 & 3 \\ 0 & 4 & 1 \\ 5 & 6 & 0 \end{vmatrix} = 2\begin{vmatrix} 4 & 1 \\ 6 & 0 \end{vmatrix} + 3\begin{vmatrix} 0 & 4 \\ 5 & 6 \end{vmatrix} = -12 - 60 = -72.$$

Hence $V = 72$. The minus sign indicates that if the coordinates are right-handed, the triple **a**, **b**, **c** is left-handed. The volume of a tetrahedron is $\frac{1}{6}$ of that of the parallelepiped (can you prove it?), hence 12.

Can you sketch the tetrahedron, choosing the origin as the common initial point of the vectors? What are the coordinates of the four vertices? ■

This is the end of vector *algebra* (in space R^3 and in the plane). Vector *calculus* (differentiation) begins in the next section.

PROBLEM SET 9.3

1–10 **GENERAL PROBLEMS**

1. Give the details of the proofs of Eqs. (4) and (5).

2. What does $\mathbf{a} \times \mathbf{b} = \mathbf{a} \times \mathbf{c}$ with $\mathbf{a} \neq \mathbf{0}$ imply?

3. Give the details of the proofs of Eqs. (6) and (11).

4. Lagrange's identity for $|\mathbf{a} \times \mathbf{b}|$. Verify it for $\mathbf{a} = [3, 4, 2]$ and $\mathbf{b} = [1, 0, 2]$. Prove it, using $\sin^2\gamma = 1 - \cos^2\gamma$. The identity is

$$\textbf{(12)} \qquad |\mathbf{a} \times \mathbf{b}| = \sqrt{(\mathbf{a} \cdot \mathbf{a})\,(\mathbf{b} \cdot \mathbf{b}) - (\mathbf{a} \cdot \mathbf{b})^2}.$$

5. What happens in Example 3 of the text if you replace **p** by $-$**p**?
6. What happens in Example 5 if you choose a P at distance $2d$ from the axis of rotation?
7. **Rotation.** A wheel is rotating about the y-axis with angular speed $\omega = 20 \sec^{-1}$. The rotation appears clockwise if one looks from the origin in the positive y-direction. Find the velocity and speed at the point $[8, 6, 0]$. Make a sketch.
8. **Rotation.** What are the velocity and speed in Prob. 7 at the point $(4, 2, -2)$ if the wheel rotates about the line $y = x, z = 0$ with $\omega = 10 \sec^{-1}$?
9. **Scalar triple product.** What does $(\mathbf{a}\ \mathbf{b}\ \mathbf{c}) = 0$ imply with respect to these vectors?
10. **WRITING REPORT.** Summarize the most important applications discussed in this section. Give examples. No proofs.

11–23 VECTOR AND SCALAR TRIPLE PRODUCTS

With respect to right-handed Cartesian coordinates, let $\mathbf{a} = [2, 1, 0]$, $\mathbf{b} = [-3, 2, 0]$, $\mathbf{c} = [1, 4, -2]$, and $\mathbf{d} = [5, -1, 3]$. Showing details, find:

11. $\mathbf{a} \times \mathbf{b}$, $\mathbf{b} \times \mathbf{a}$, $\mathbf{a} \cdot \mathbf{b}$
12. $3\mathbf{c} \times 5\mathbf{d}$, $15\mathbf{d} \times \mathbf{c}$, $15\mathbf{d} \cdot \mathbf{c}$, $15\mathbf{c} \cdot \mathbf{d}$
13. $\mathbf{c} \times (\mathbf{a} + \mathbf{b})$, $\mathbf{a} \times \mathbf{c} + \mathbf{b} \times \mathbf{c}$
14. $4\mathbf{b} \times 3\mathbf{c} + 12\mathbf{c} \times \mathbf{b}$
15. $(\mathbf{a} + \mathbf{d}) \times (\mathbf{d} + \mathbf{a})$
16. $(\mathbf{b} \times \mathbf{c}) \cdot \mathbf{d}$, $\mathbf{b} \cdot (\mathbf{c} \times \mathbf{d})$
17. $(\mathbf{b} \times \mathbf{c}) \times \mathbf{d}$, $\mathbf{b} \times (\mathbf{c} \times \mathbf{d})$
18. $(\mathbf{a} \times \mathbf{b}) \times \mathbf{a}$, $\mathbf{a} \times (\mathbf{b} \times \mathbf{a})$
19. $(\mathbf{i}\ \mathbf{j}\ \mathbf{k})$, $(\mathbf{i}\ \mathbf{k}\ \mathbf{j})$
20. $(\mathbf{a} \times \mathbf{b}) \times (\mathbf{c} \times \mathbf{d})$, $(\mathbf{a}\ \mathbf{b}\ \mathbf{d})\mathbf{c} - (\mathbf{a}\ \mathbf{b}\ \mathbf{c})\mathbf{d}$
21. $4\mathbf{b} \times 3\mathbf{c}$, $12|\mathbf{b} \times \mathbf{c}|$, $12|\mathbf{c} \times \mathbf{b}|$
22. $(\mathbf{a} - \mathbf{b}\ \ \mathbf{c} - \mathbf{b}\ \ \mathbf{d} - \mathbf{b})$, $(\mathbf{a}\ \mathbf{c}\ \mathbf{d})$
23. $\mathbf{b} \times \mathbf{b}$, $(\mathbf{b} - \mathbf{c}) \times (\mathbf{c} - \mathbf{b})$, $\mathbf{b} \cdot \mathbf{b}$
24. **TEAM PROJECT. Useful Formulas for Three and Four Vectors.** Prove (13)–(16), which are often useful in practical work, and illustrate each formula with two examples. *Hint.* For (13) choose Cartesian coordinates such that $\mathbf{d} = [d_1, 0, 0]$ and $\mathbf{c} = [c_1, c_2, 0]$. Show that each side of (13) then equals $[-b_2c_2d_1, b_1c_2d_1, 0]$, and give reasons why the two sides are then equal in any Cartesian coordinate system. For (14) and (15) use (13).

$$(13)\quad \mathbf{b} \times (\mathbf{c} \times \mathbf{d}) = (\mathbf{b} \cdot \mathbf{d})\mathbf{c} - (\mathbf{b} \cdot \mathbf{c})\mathbf{d}$$

$$(14)\quad (\mathbf{a} \times \mathbf{b}) \times (\mathbf{c} \times \mathbf{d}) = (\mathbf{a}\ \mathbf{b}\ \mathbf{d})\mathbf{c} - (\mathbf{a}\ \mathbf{b}\ \mathbf{c})\mathbf{d}$$

$$(15)\quad (\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c})$$

$$(16)\quad (\mathbf{a}\ \mathbf{b}\ \mathbf{c}) = (\mathbf{b}\ \mathbf{c}\ \mathbf{a}) = (\mathbf{c}\ \mathbf{a}\ \mathbf{b}) = -(\mathbf{c}\ \mathbf{b}\ \mathbf{a}) = -(\mathbf{a}\ \mathbf{c}\ \mathbf{b})$$

25–35 APPLICATIONS

25. **Moment m of a force p.** Find the moment vector **m** and m of $\mathbf{p} = [2, 3, 0]$ about Q: $(2, 1, 0)$ acting on a line through A: $(0, 3, 0)$. Make a sketch.
26. **Moment.** Solve Prob. 25 if $\mathbf{p} = [1, 0, 3]$, Q: $(2, 0, 3)$, and A: $(4, 3, 5)$.
27. **Parallelogram.** Find the area if the vertices are $(4, 2, 0)$, $(10, 4, 0)$, $(5, 4, 0)$, and $(11, 6, 0)$. Make a sketch.
28. **A remarkable parallelogram.** Find the area of the quadrangle Q whose vertices are the midpoints of the sides of the quadrangle P with vertices A: $(2, 1, 0)$, B: $(5, -1. 0)$, C: $(8, 2, 0)$, and D: $(4, 3, 0)$. Verify that Q is a parallelogram.
29. **Triangle.** Find the area if the vertices are $(0, 0, 1)$, $(2, 0, 5)$, and $(2, 3, 4)$.
30. **Plane.** Find the plane through the points A: $(1, 2, \frac{1}{4})$, B: $(4, 2, -2)$, and C: $(0, 8, 4)$.
31. **Plane.** Find the plane through $(1, 3, 4)$, $(1, -2, 6)$, and $(4, 0, 7)$.
32. **Parallelepiped.** Find the volume if the edge vectors are $\mathbf{i} + \mathbf{j}$, $-2\mathbf{i} + 2\mathbf{k}$, and $-2\mathbf{i} - 3\mathbf{k}$. Make a sketch.
33. **Tetrahedron.** Find the volume if the vertices are $(1, 1, 1)$, $(5, -7, 3)$, $(7, 4, 8)$, and $(10, 7, 4)$.
34. **Tetrahedron.** Find the volume if the vertices are $(1, 3, 6)$, $(3, 7, 12)$, $(8, 8, 9)$, and $(2, 2, 8)$.
35. **WRITING PROJECT. Applications of Cross Products.** Summarize the most important applications we have discussed in this section and give a few simple examples. No proofs.

9.4 Vector and Scalar Functions and Their Fields. Vector Calculus: Derivatives

Our discussion of vector calculus begins with identifying the two types of functions on which it operates. Let P be any point in a domain of definition. Typical domains in applications are three-dimensional, or a surface or a curve in space. Then we define a **vector function** **v**, whose values are vectors, that is,

$$\mathbf{v} = \mathbf{v}(P) = [v_1(P), v_2(P), v_3(P)]$$

that depends on points P in space. We say that a vector function defines a **vector field** in a domain of definition. Typical domains were just mentioned. Examples of vector fields are the field of tangent vectors of a curve (shown in Fig. 195), normal vectors of a surface (Fig. 196), and velocity field of a rotating body (Fig. 197). Note that vector functions may also depend on time t or on some other parameters.

Similarly, we define a **scalar function** f, whose values are scalars, that is,

$$f = f(P)$$

that depends on P. We say that a scalar function defines a scalar field in that three-dimensional domain or surface or curve in space. Two representative examples of scalar fields are the temperature field of a body and the pressure field of the air in Earth's atmosphere. Note that scalar functions may also depend on some parameter such as time t.

Notation. If we introduce Cartesian coordinates x, y, z, then, instead of writing $\mathbf{v}(P)$ for the vector function, we can write

$$\mathbf{v}(x, y, z) = [v_1(x, y, z), \quad v_2(x, y, z), \quad v_3(x, y, z)].$$

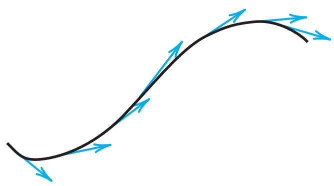


Fig. 195. Field of tangent vectors of a curve

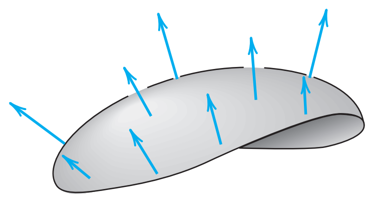


Fig. 196. Field of normal vectors of a surface

We have to keep in mind that the components depend on our choice of coordinate system, whereas a vector field that has a physical or geometric meaning should have magnitude and direction depending only on P, not on the choice of coordinate system.

Similarly, for a scalar function, we write

$$f(P) = f(x, y, z).$$

We illustrate our discussion of vector functions, scalar functions, vector fields, and scalar fields by the following three examples.

EXAMPLE 1 Scalar Function (Euclidean Distance in Space)

The distance $f(P)$ of any point P from a fixed point P_0 in space is a scalar function whose domain of definition is the whole space. $f(P)$ defines a scalar field in space. If we introduce a Cartesian coordinate system and P_0 has the coordinates x_0, y_0, z_0, then f is given by the well-known formula

$$f(P) = f(x, y, z) = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}$$

where x, y, z are the coordinates of P. If we replace the given Cartesian coordinate system with another such system by translating and rotating the given system, then the values of the coordinates of P and P_0 will in general change, but $f(P)$ will have the same value as before. Hence $f(P)$ is a scalar function. The direction cosines of the straight line through P and P_0 are not scalars because their values depend on the choice of the coordinate system. ■

EXAMPLE 2 Vector Field (Velocity Field)

At any instant the velocity vectors $\mathbf{v}(P)$ of a rotating body B constitute a vector field, called the **velocity field** of the rotation. If we introduce a Cartesian coordinate system having the origin on the axis of rotation, then (see Example 5 in Sec. 9.3)

$$\mathbf{v}(x, y, z) = \mathbf{w} \times \mathbf{r} = \mathbf{w} \times [x, y, z] = \mathbf{w} \times (x\mathbf{i} + y\mathbf{j} + z\mathbf{k}) \tag{1}$$

where x, y, z are the coordinates of any point P of B at the instant under consideration. If the coordinates are such that the z-axis is the axis of rotation and $\mathbf{w}$ points in the positive z-direction, then $\mathbf{w} = \omega\mathbf{k}$ and

$$\mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & 0 & \omega \\ x & y & z \end{vmatrix} = \omega[-y, x, 0] = \omega(-y\mathbf{i} + x\mathbf{j}).$$

An example of a rotating body and the corresponding velocity field are shown in Fig. 197. ■

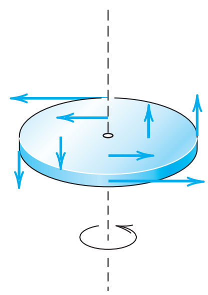


Fig. 197. Velocity field of a rotating body

EXAMPLE 3 Vector Field (Field of Force, Gravitational Field)

Let a particle A of mass M be fixed at a point P_0 and let a particle B of mass m be free to take up various positions P in space. Then A attracts B. According to **Newton's law of gravitation** the corresponding gravitational force $\mathbf{p}$ is directed from P to P_0, and its magnitude is proportional to $1/r^2$, where r is the distance between P and P_0, say,

$$|\mathbf{p}| = \frac{c}{r^2}, \qquad c = GMm. \tag{2}$$

Here $G = 6.67 \cdot 10^{-8}\ \mathrm{cm}^3/(\mathrm{g} \cdot \mathrm{sec}^2)$ is the gravitational constant. Hence $\mathbf{p}$ defines a vector field in space. If we introduce Cartesian coordinates such that P_0 has the coordinates x_0, y_0, z_0 and P has the coordinates x, y, z, then by the Pythagorean theorem,

$$r = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2} \qquad (\geqq 0).$$

Assuming that $r > 0$ and introducing the vector

$$\mathbf{r} = [x - x_0,\quad y - y_0,\quad z - z_0] = (x - x_0)\mathbf{i} + (y - y_0)\mathbf{j} + (z - z_0)\mathbf{k},$$

we have $|\mathbf{r}| = r$, and $(-1/r)\mathbf{r}$ is a unit vector in the direction of $\mathbf{p}$; the minus sign indicates that $\mathbf{p}$ is directed from P to P_0 (Fig. 198). From this and (2) we obtain

$$\begin{aligned} \mathbf{p} = |\mathbf{p}|\left(-\frac{1}{r}\mathbf{r}\right) &= -\frac{c}{r^3}\mathbf{r} = \left[-c\,\frac{x - x_0}{r^3},\quad -c\,\frac{y - y_0}{r^3},\quad -c\,\frac{z - z_0}{r^3}\right] \\ &= -c\,\frac{x - x_0}{r^3}\,\mathbf{i} - c\,\frac{y - y_0}{r^3}\,\mathbf{j} - c\,\frac{z - z_0}{r^3}\,\mathbf{k}. \end{aligned} \tag{3}$$

This vector function describes the gravitational force acting on B. ■

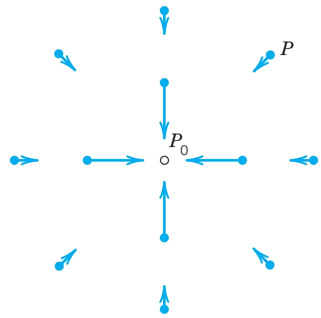


Fig. 198. Gravitational field in Example 3

Vector Calculus

The student may be pleased to learn that many of the concepts covered in (regular) calculus carry over to vector calculus. Indeed, we show how the basic concepts of convergence, continuity, and differentiability from calculus can be defined for vector functions in a simple and natural way. Most important of these is the derivative of a vector function.

Convergence. An infinite sequence of vectors $\mathbf{a}_{(n)}$, $n = 1, 2, \cdots,$ is said to **converge** if there is a vector $\mathbf{a}$ such that

$$\lim_{n\to\infty} |\mathbf{a}_{(n)} - \mathbf{a}| = 0. \tag{4}$$

$\mathbf{a}$ is called the **limit vector** of that sequence, and we write

$$\lim_{n\to\infty} \mathbf{a}_{(n)} = \mathbf{a}. \tag{5}$$

If the vectors are given in Cartesian coordinates, then this sequence of vectors converges to $\mathbf{a}$ if and only if the three sequences of components of the vectors converge to the corresponding components of $\mathbf{a}$. We leave the simple proof to the student.

Similarly, a vector function $\mathbf{v}(t)$ of a real variable t is said to have the **limit** $\boldsymbol{l}$ as t approaches t_0, if $\mathbf{v}(t)$ is defined in some neighborhood of t_0 (possibly except at t_0) and

$$\lim_{t\to t_0} |\mathbf{v}(t) - \boldsymbol{l}| = 0. \tag{6}$$

Then we write

$$\lim_{t\to t_0} \mathbf{v}(t) = \boldsymbol{l}. \tag{7}$$

Here, a *neighborhood* of t_0 is an interval (segment) on the t-axis containing t_0 as an interior point (not as an endpoint).

Continuity. A vector function $\mathbf{v}(t)$ is said to be **continuous** at $t = t_0$ if it is defined in some neighborhood of t_0 (including at t_0 itself!) and

$$\lim_{t\to t_0} \mathbf{v}(t) = \mathbf{v}(t_0). \tag{8}$$

If we introduce a Cartesian coordinate system, we may write

$$\mathbf{v}(t) = [v_1(t), v_2(t), v_3(t)] = v_1(t)\mathbf{i} + v_2(t)\mathbf{j} + v_3(t)\mathbf{k}.$$

Then $\mathbf{v}(t)$ is continuous at t_0 if and only if its three components are continuous at t_0.

We now state the most important of these definitions.

DEFINITION

Derivative of a Vector Function

A vector function $\mathbf{v}(t)$ is said to be **differentiable** at a point t if the following limit exists:

$$\mathbf{v}'(t) = \lim_{\Delta t \to 0} \frac{\mathbf{v}(t + \Delta t) - \mathbf{v}(t)}{\Delta t}. \tag{9}$$

This vector $\mathbf{v}'(t)$ is called the **derivative** of $\mathbf{v}(t)$. See Fig. 199.

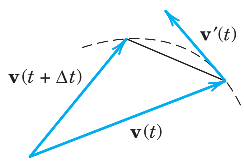


Fig. 199. Derivative of a vector function

In components with respect to a given Cartesian coordinate system,

$$\mathbf{v}'(t) = [v_1'(t), \quad v_2'(t), \quad v_3'(t)]. \tag{10}$$

Hence the derivative $\mathbf{v}'(t)$ ***is obtained by differentiating each component separately.*** For instance, if $\mathbf{v} = [t, t^2, 0]$, then $\mathbf{v}' = [1, 2t, 0]$.

Equation (10) follows from (9) and conversely because (9) is a "vector form" of the usual formula of calculus by which the derivative of a function of a single variable is defined. [The curve in Fig. 199 is the locus of the terminal points representing $\mathbf{v}(t)$ for values of the independent variable in some interval containing t and $t + \Delta t$ in (9)]. It follows that the familiar differentiation rules continue to hold for differentiating vector functions, for instance,

$$(c\mathbf{v})' = c\mathbf{v}' \qquad (c \text{ constant}),$$

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

and in particular

$$(\mathbf{u} \bullet \mathbf{v})' = \mathbf{u}' \bullet \mathbf{v} + \mathbf{u} \bullet \mathbf{v}' \tag{11}$$

$$(\mathbf{u} \times \mathbf{v})' = \mathbf{u}' \times \mathbf{v} + \mathbf{u} \times \mathbf{v}' \tag{12}$$

$$(\mathbf{u} \quad \mathbf{v} \quad \mathbf{w})' = (\mathbf{u}' \quad \mathbf{v} \quad \mathbf{w}) + (\mathbf{u} \quad \mathbf{v}' \quad \mathbf{w}) + (\mathbf{u} \quad \mathbf{v} \quad \mathbf{w}'). \tag{13}$$

The simple proofs are left to the student. In (12), note the order of the vectors carefully because cross multiplication is not commutative.

EXAMPLE 4 Derivative of a Vector Function of Constant Length

Let $\mathbf{v}(t)$ be a vector function whose length is constant, say, $|\mathbf{v}(t)| = c$. Then $|\mathbf{v}|^2 = \mathbf{v} \cdot \mathbf{v} = c^2$, and $(\mathbf{v} \cdot \mathbf{v})' = 2\mathbf{v} \cdot \mathbf{v}' = 0$, by differentiation [see (11)]. This yields the following result. *The derivative of a vector function* $\mathbf{v}(t)$ *of constant length is either the zero vector or is perpendicular to* $\mathbf{v}(t)$. ■

Partial Derivatives of a Vector Function

Our present discussion shows that partial differentiation of vector functions of two or more variables can be introduced as follows. Suppose that the components of a vector function

$$\mathbf{v} = [v_1, \quad v_2, \quad v_3] = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$$

are differentiable functions of n variables $t_1, \cdots, t_n$. Then the **partial derivative** of $\mathbf{v}$ with respect to t_m is denoted by $\partial\mathbf{v}/\partial t_m$ and is defined as the vector function

$$\frac{\partial \mathbf{v}}{\partial t_m} = \frac{\partial v_1}{\partial t_m}\mathbf{i} + \frac{\partial v_2}{\partial t_m}\mathbf{j} + \frac{\partial v_3}{\partial t_m}\mathbf{k}.$$

Similarly, second partial derivatives are

$$\frac{\partial^2 \mathbf{v}}{\partial t_l \partial t_m} = \frac{\partial^2 v_1}{\partial t_l \partial t_m}\mathbf{i} + \frac{\partial^2 v_2}{\partial t_l \partial t_m}\mathbf{j} + \frac{\partial^2 v_3}{\partial t_l \partial t_m}\mathbf{k},$$

and so on.

EXAMPLE 5 Partial Derivatives

Let $\mathbf{r}(t_1, t_2) = a \cos t_1 \,\mathbf{i} + a \sin t_1 \,\mathbf{j} + t_2 \,\mathbf{k}$. Then $\dfrac{\partial \mathbf{r}}{\partial t_1} = -a \sin t_1 \,\mathbf{i} + a \cos t_1 \,\mathbf{j}$ and $\dfrac{\partial \mathbf{r}}{\partial t_2} = \mathbf{k}$. ■

Various physical and geometric applications of derivatives of vector functions will be discussed in the next sections as well as in Chap. 10.

PROBLEM SET 9.4

1–8 SCALAR FIELDS IN THE PLANE

Let the temperature T in a body be independent of z so that it is given by a scalar function $T = T(x, t)$. Identify the isotherms $T(x, y) = $ const. Sketch some of them.

1. $T = x^2 - y^2$
2. $T = xy$
3. $T = 3x - 4y$
4. $T = \arctan(y/x)$
5. $T = y/(x^2 + y^2)$
6. $T = x/(x^2 + y^2)$
7. $T = 9x^2 + 4y^2$
8. CAS PROJECT. Scalar Fields in the Plane. Sketch or graph isotherms of the following fields and describe what they look like.

(a) $x^2 - 4x - y^2$
(b) $x^2y - y^3/3$
(c) $\cos x \sinh y$
(d) $\sin x \sinh y$
(e) $e^x \sin y$
(f) $e^{2x} \cos 2y$
(g) $x^4 - 6x^2y^2 + y^4$
(h) $x^2 - 2x - y^2$

9–14 SCALAR FIELDS IN SPACE

What kind of surfaces are the **level surfaces** $f(x, y, z) = $ const?

9. $f = 4x - 3y + 2z$
10. $f = 9(x^2 + y^2) + z^2$
11. $f = 5x^2 + 2y^2$
12. $f = z - \sqrt{x^2 + y^2}$
13. $f = z - (x^2 + y^2)$
14. $f = x - y^2$

Mean Value Theorems

THEOREM 2

Mean Value Theorem

Let $f(x, y, z)$ be continuous and have continuous first partial derivatives in a domain D in xyz-space. Let P_0: (x_0, y_0, z_0) and P: $(x_0 + h, y_0 + k, z_0 + l)$ be points in D such that the straight line segment P_0P joining these points lies entirely in D. Then

$$f(x_0 + h, y_0 + k, z_0 + l) - f(x_0, y_0, z_0) = h\frac{\partial f}{\partial x} + k\frac{\partial f}{\partial y} + l\frac{\partial f}{\partial z}, \tag{7}$$

the partial derivatives being evaluated at a suitable point of that segment.

Special Cases

For a function $f(x, y)$ of two variables (satisfying assumptions as in the theorem), formula (7) reduces to (Fig. 214)

$$f(x_0 + h, y_0 + k) - f(x_0, y_0) = h\frac{\partial f}{\partial x} + k\frac{\partial f}{\partial y}, \tag{8}$$

and, for a function $f(x)$ of a single variable, (7) becomes

$$f(x_0 + h) - f(x_0) = h\frac{\partial f}{\partial x}, \tag{9}$$

where in (9), the domain D is a segment of the x-axis and the derivative is taken at a suitable point between x_0 and $x_0 + h$.

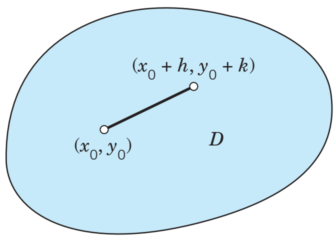


Fig. 214. Mean value theorem for a function of two variables [Formula (8)]

9.7 Gradient of a Scalar Field. Directional Derivative

We shall see that *some* of the vector fields that occur in applications—not all of them!—can be obtained from scalar fields. Using scalar fields instead of vector fields is of a considerable advantage because scalar fields are easier to use than vector fields. It is the "gradient" that allows us to obtain vector fields from scalar fields, and thus the gradient is of great practical importance to the engineer.

DEFINITION 1

Gradient

The setting is that we are given a scalar function $f(x, y, z)$ that is defined and differentiable in a domain in 3-space with Cartesian coordinates x, y, z. We denote the **gradient** of that function by grad f or ∇f (read **nabla** f). Then the qradient of $f(x, y, z)$ is defined as the vector function

$$\text{grad}\, f = \nabla f = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right] = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} + \frac{\partial f}{\partial z}\mathbf{k}. \tag{1}$$

Remarks. For a definition of the gradient in curvilinear coordinates, see App. 3.4. As a quick example, if $f(x, y, z) = 2y^3 + 4xz + 3x$, then grad $f = [4z + 3, 6y^2, 4x]$. Furthermore, we will show later in this section that (1) actually does define a vector.

The notation ∇f is suggested by the *differential operator* ∇ (read *nabla*) defined by

$$\nabla = \frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}. \tag{1*}$$

Gradients are useful in several ways, notably in giving the rate of change of $f(x, y, z)$ in any direction in space, in obtaining surface normal vectors, and in deriving vector fields from scalar fields, as we are going to show in this section.

Directional Derivative

From calculus we know that the partial derivatives in (1) give the rates of change of $f(x, y, z)$ in the directions of the three coordinate axes. It seems natural to extend this and ask for the rate of change of f *in an arbitrary direction* in space. This leads to the following concept.

DEFINITION 2

Directional Derivative

The directional derivative $D_{\mathbf{b}}f$ or df/ds of a function $f(x, y, z)$ at a point P in the direction of a vector $\mathbf{b}$ is defined by (see Fig. 215)

$$D_{\mathbf{b}}f = \frac{df}{ds} = \lim_{s\to 0} \frac{f(Q) - f(P)}{s}. \tag{2}$$

Here Q is a variable point on the straight line L in the direction of $\mathbf{b}$, and $|s|$ is the distance between P and Q. Also, $s > 0$ if Q lies in the direction of $\mathbf{b}$ (as in Fig. 215), $s < 0$ if Q lies in the direction of $-\mathbf{b}$, and $s = 0$ if $Q = P$.

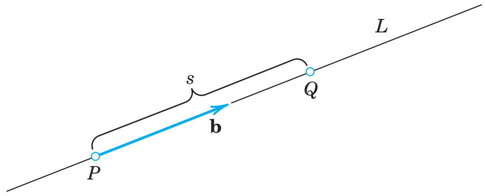


Fig. 215. Directional derivative

The next idea is to use Cartesian xyz-coordinates and for $\mathbf{b}$ a unit vector. Then the line L is given by

$$\mathbf{r}(s) = x(s)\mathbf{i} + y(s)\mathbf{j} + z(s)\mathbf{k} = \mathbf{p}_0 + s\mathbf{b} \qquad (|\mathbf{b}| = 1) \tag{3}$$

where $\mathbf{p}_0$ the position vector of P. Equation (2) now shows that $D_{\mathbf{b}}f = df/ds$ is the derivative of the function $f(x(s), y(s), z(s))$ with respect to the arc length s of L. Hence, assuming that f has continuous partial derivatives and applying the chain rule [formula (3) in the previous section], we obtain

$$D_{\mathbf{b}}f = \frac{df}{ds} = \frac{\partial f}{\partial x}x' + \frac{\partial f}{\partial y}y' + \frac{\partial f}{\partial z}z' \tag{4}$$

where primes denote derivatives with respect to s (which are taken at $s = 0$). But here, differentiating (3) gives $\mathbf{r}' = x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k} = \mathbf{b}$. Hence (4) is simply the inner product of grad f and $\mathbf{b}$ [see (2), Sec. 9.2]; that is,

$$\boxed{D_{\mathbf{b}}f = \frac{df}{ds} = \mathbf{b} \bullet \operatorname{grad} f} \qquad (|\mathbf{b}| = 1). \tag{5}$$

ATTENTION! If the direction is given by a vector $\mathbf{a}$ of any length ($\neq 0$), then

$$\boxed{D_{\mathbf{a}}f = \frac{df}{ds} = \frac{1}{|\mathbf{a}|}\mathbf{a} \bullet \operatorname{grad} f.} \tag{5*}$$

EXAMPLE 1 Gradient. Directional Derivative

Find the directional derivative of $f(x, y, z) = 2x^2 + 3y^2 + z^2$ at P: (2, 1, 3) in the direction of $\mathbf{a} = [1, 0, -2]$.

Solution. grad $f = [4x, 6y, 2z]$ gives at P the vector grad $f(P) = [8, 6, 6]$. From this and (5*) we obtain, since $|\mathbf{a}| = \sqrt{5}$,

$$D_{\mathbf{a}}f(P) = \frac{1}{\sqrt{5}}[1, 0, -2] \bullet [8, 6, 6] = \frac{1}{\sqrt{5}}(8 + 0 - 12) = -\frac{4}{\sqrt{5}} = -1.789.$$

The minus sign indicates that at P the function f is decreasing in the direction of $\mathbf{a}$. ■

Gradient Is a Vector. Maximum Increase

Here is a finer point of mathematics that concerns the consistency of our theory: grad f in (1) *looks* like a vector—after all, it has three components! But to prove that it *actually is* a vector, since it is defined in terms of components depending on the Cartesian coordinates, we must show that grad f has a length and direction independent of the choice of those coordinates. See proof of Theorem 1. In contrast, $[\partial f/\partial x, 2\partial f/\partial y, \partial f/\partial z]$ also looks like a vector but does not have a length and direction independent of the choice of Cartesian coordinates.

Incidentally, the direction makes the gradient eminently useful: grad f *points in the direction of maximum increase of* f.

THEOREM 1

Use of Gradient: Direction of Maximum Increase

Let $f(P) = f(x, y, z)$ *be a scalar function having continuous first partial derivatives in some domain* B *in space. Then* grad f *exists in* B *and is a vector, that is, its length and direction are independent of the particular choice of Cartesian coordinates. If* grad $f(P) \neq \mathbf{0}$ *at some point* P*, it has the direction of maximum increase of* f *at* P.

PROOF From (5) and the definition of inner product [(1) in Sec. 9.2] we have

$$D_{\mathbf{b}}f = |\mathbf{b}||\text{grad } f| \cos\gamma = |\text{grad } f| \cos\gamma \tag{6}$$

where γ is the angle between $\mathbf{b}$ and grad f. Now f is a scalar function. Hence its value at a point P depends on P but not on the particular choice of coordinates. The same holds for the arc length s of the line L in Fig. 215, hence also for $D_{\mathbf{b}}f$. Now (6) shows that $D_{\mathbf{b}}f$ is maximum when $\cos\gamma = 1$, $\gamma = 0$, and then $D_{\mathbf{b}}f = |\text{grad } f|$. It follows that the length and direction of grad f are independent of the choice of coordinates. Since $\gamma = 0$ if and only if $\mathbf{b}$ has the direction of grad f, the latter is the direction of maximum increase of f at P, provided grad $f \neq \mathbf{0}$ at P. Make sure that you understood the proof to get a good feel for mathematics.

Gradient as Surface Normal Vector

Gradients have an important application in connection with surfaces, namely, as surface normal vectors, as follows. Let S be a surface represented by $f(x, y, z) = c = \text{const}$, where f is differentiable. Such a surface is called a **level surface** of f, and for different c we get different level surfaces. Now let C be a curve on S through a point P of S. As a curve in space, C has a representation $\mathbf{r}(t) = [x(t), y(t), z(t)]$. For C to lie on the surface S, the components of $\mathbf{r}(t)$ must satisfy $f(x, y, z) = c$, that is,

$$f(x(t), y(t), z(t)) = c. \tag{7}$$

Now a tangent vector of C is $\mathbf{r}'(t) = [x'(t), y'(t), z'(t)]$. And the tangent vectors of all curves on S passing through P will generally form a plane, called the **tangent plane** of S at P. (Exceptions occur at edges or cusps of S, for instance, at the apex of the cone in Fig. 217.) The normal of this plane (the straight line through P perpendicular to the tangent plane) is called the **surface normal** to S at P. A vector in the direction of the surface normal is called a **surface normal vector** of S at P. We can obtain such a vector quite simply by differentiating (7) with respect to t. By the chain rule,

$$\frac{\partial f}{\partial x}x' + \frac{\partial f}{\partial y}y' + \frac{\partial f}{\partial z}z' = (\text{grad } f) \bullet \mathbf{r}' = 0.$$

Hence grad f is orthogonal to all the vectors $\mathbf{r}'$ in the tangent plane, so that it is a normal vector of S at P. Our result is as follows (see Fig. 216).

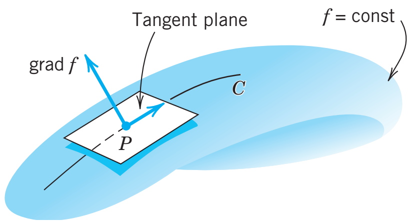


Fig. 216. Gradient as surface normal vector

THEOREM 2

Gradient as Surface Normal Vector

Let f be a differentiable scalar function in space. Let $f(x, y, z) = c = const$ represent a surface S. Then if the gradient of f at a point P of S is not the zero vector, it is a normal vector of S at P.

EXAMPLE 2

Gradient as Surface Normal Vector. Cone

Find a unit normal vector $\mathbf{n}$ of the cone of revolution $z^2 = 4(x^2 + y^2)$ at the point P: $(1, 0, 2)$.

Solution. The cone is the level surface $f = 0$ of $f(x, y, z) = 4(x^2 + y^2) - z^2$. Thus (Fig. 217)

$$\text{grad } f = [8x, \quad 8y, \quad -2z], \qquad \text{grad } f(P) = [8, \quad 0, \quad -4]$$

$$\mathbf{n} = \frac{1}{|\text{grad } f(P)|}\text{grad } f(P) = \left[\frac{2}{\sqrt{5}}, \quad 0, \quad -\frac{1}{\sqrt{5}}\right].$$

$\mathbf{n}$ points downward since it has a negative z-component. The other unit normal vector of the cone at P is $-\mathbf{n}$. ■

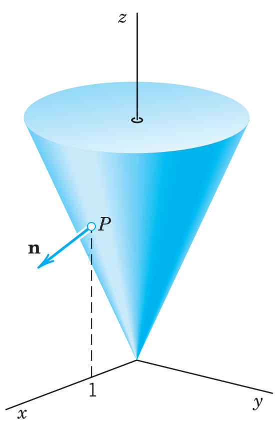


Fig. 217. Cone and unit normal vector **n**

Vector Fields That Are Gradients of Scalar Fields ("Potentials")

At the beginning of this section we mentioned that some vector fields have the advantage that they can be obtained from scalar fields, which can be worked with more easily. Such a vector field is given by a vector function $\mathbf{v}(P)$, which is obtained as the gradient of a scalar function, say, $\mathbf{v}(P) = \operatorname{grad} f(P)$. The function $f(P)$ is called a *potential function* or a **potential** of $\mathbf{v}(P)$. Such a $\mathbf{v}(P)$ and the corresponding vector field are called **conservative** because in such a vector field, energy is conserved; that is, no energy is lost (or gained) in displacing a body (or a charge in the case of an electrical field) from a point P to another point in the field and back to P. We show this in Sec. 10.2.

Conservative fields play a central role in physics and engineering. A basic application concerns the gravitational force (see Example 3 in Sec. 9.4) and we show that it has a potential which satisfies Laplace's equation, the most important partial differential equation in physics and its applications.

THEOREM 3

Gravitational Field. Laplace's Equation

The force of attraction

$$\mathbf{p} = -\frac{c}{r^3}\mathbf{r} = -c\left[\frac{x-x_0}{r^3}, \frac{y-y_0}{r^3}, \frac{z-z_0}{r^3}\right] \tag{8}$$

between two particles at points P_0: (x_0, y_0, z_0) and P: (x, y, z) (as given by Newton's law of gravitation) *has the potential* $f(x, y, z) = c/r$, *where* $r\ (>0)$ *is the distance between* P_0 *and* P.

Thus $\mathbf{p} = \operatorname{grad} f = \operatorname{grad}(c/r)$. *This potential* f *is a solution of* **Laplace's equation**

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = 0. \tag{9}$$

[$\nabla^2 f$ (read *nabla squared* f) is called the **Laplacian** of f.]

PROOF That distance is $r = ((x-x_0)^2 + (y-y_0)^2 + (z-z_2)^2)^{1/2}$. The key observation now is that for the components of $\mathbf{p} = [p_1, p_2, p_3]$ we obtain by partial differentiation

$$\frac{\partial}{\partial x}\left(\frac{1}{r}\right) = \frac{-2(x-x_0)}{2[(x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2]^{3/2}} = -\frac{x-x_0}{r^3} \tag{10a}$$

and similarly

$$\begin{aligned} \frac{\partial}{\partial y}\left(\frac{1}{r}\right) &= -\frac{y-y_0}{r^3}, \\ \frac{\partial}{\partial z}\left(\frac{1}{r}\right) &= -\frac{z-z_0}{r^3}. \end{aligned} \tag{10b}$$

From this we see that, indeed, **p** is the gradient of the scalar function $f = c/r$. The second statement of the theorem follows by partially differentiating (10), that is,

$$\frac{\partial^2}{\partial x^2}\left(\frac{1}{r}\right) = -\frac{1}{r^3} + \frac{3(x - x_0)^2}{r^5},$$

$$\frac{\partial^2}{\partial y^2}\left(\frac{1}{r}\right) = -\frac{1}{r^3} + \frac{3(y - y_0)^2}{r^5},$$

$$\frac{\partial^2}{\partial z^2}\left(\frac{1}{r}\right) = -\frac{1}{r^3} + \frac{3(z - z_0)^2}{r^5},$$

and then adding these three expressions. Their common denominator is r^5. Hence the three terms $-1/r^3$ contribute $-3r^2$ to the numerator, and the three other terms give the sum

$$3(x - x_0)^2 + 3(y - y_0)^2 + 3(z - z_0)^2 = 3r^2,$$

so that the numerator is 0, and we obtain (9). ∎

$\nabla^2 f$ is also denoted by Δf. The differential operator

(11) $$\boxed{\nabla^2 = \Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}}$$

(read "nabla squared" or "delta") is called the **Laplace operator**. It can be shown that the field of force produced by any distribution of masses is given by a vector function that is the gradient of a scalar function f, and f satisfies (9) in any region that is free of matter.

The great importance of the Laplace equation also results from the fact that there are other laws in physics that are of the same form as Newton's law of gravitation. For instance, in electrostatics the force of attraction (or repulsion) between two particles of opposite (or like) charge Q_1 and Q_2 is

(12) $$\boxed{\mathbf{p} = \frac{k}{r^3}\mathbf{r}}$$ (**Coulomb's law**[6]).

Laplace's equation will be discussed in detail in Chaps. 12 and 18.

A method for finding out whether a given vector field has a potential will be explained in Sec. 9.9.

---

[6]CHARLES AUGUSTIN DE COULOMB (1736–1806), French physicist and engineer. Coulomb's law was derived by him from his own very precise measurements.

PROBLEM SET 9.7

1–6 CALCULATION OF GRADIENTS

Find grad f. Graph some level curves $f = \text{const}$. Indicate ∇f by arrows at some points of these curves.

1. $f = (x + 1)(2y - 1)$
2. $f = 9x^2 + 4y^2$
3. $f = y/x$
4. $(y + 6)^2 + (x - 4)^2$
5. $f = x^4 + y^4$
6. $f = (x^2 - y^2)/(x^2 + y^2)$

7–10 USEFUL FORMULAS FOR GRADIENT AND LAPLACIAN

Prove and illustrate by an example.

7. $\nabla(f^n) = nf^{n-1}\nabla f$
8. $\nabla(fg) = f\nabla g + g\nabla f$
9. $\nabla(f/g) = (1/g^2)(g\nabla f - f\nabla g)$
10. $\nabla^2(fg) = g\nabla^2 f + 2\nabla f \cdot \nabla g + f\nabla^2 g$

11–15 USE OF GRADIENTS. ELECTRIC FORCE

The force in an electrostatic field given by $f(x, y, z)$ has the direction of the gradient. Find ∇f and its value at P.

11. $f = xy$, P: $(-4, 5)$
12. $f = x/(x^2 + y^2)$, P: $(1, 1)$
13. $f = \ln(x^2 + y^2)$, P: $(8, 6)$
14. $f = (x^2 + y^2 + z^2)^{-1/2}$ P: $(12, 0, 16)$
15. $f = 4x^2 + 9y^2 + z^2$, P: $(5, -1, -11)$
16. For what points P: (x, y, z) does ∇f with $f = 25x^2 + 9y^2 + 16z^2$ have the direction from P to the origin?
17. Same question as in Prob. 16 when $f = 25x^2 + 4y^2$.

18–23 VELOCITY FIELDS

Given the velocity potential f of a flow, find the velocity $\mathbf{v} = \nabla f$ of the field and its value $\mathbf{v}(P)$ at P. Sketch $\mathbf{v}(P)$ and the curve $f = \text{const}$ passing through P.

18. $f = x^2 - 6x - y^2$, P: $(-1, 5)$
19. $f = \cos x \cosh y$, P: $(\frac{1}{2}\pi, \ln 2)$
20. $f = x(1 + (x^2 + y^2)^{-1})$, P: $(1, 1)$
21. $f = e^x \cos y$, P: $(1, \frac{1}{2}\pi)$
22. At what points is the flow in Prob. 21 directed vertically upward?
23. At what points is the flow in Prob. 21 horizontal?

24–27 HEAT FLOW

Experiments show that in a temperature field, heat flows in the direction of maximum decrease of temperature T. Find this direction in general and at the given point P. Sketch that direction at P as an arrow.

24. $T = 3x^2 - 2y^2$, P: $(2.5, 1.8)$
25. $T = z/(x^2 + y^2)$, P: $(0, 1, 2)$
26. $T = x^2 + y^2 + 4z^2$, P: $(2, -1, 2)$
27. **CAS PROJECT. Isotherms.** Graph some curves of constant temperature ("isotherms") and indicate directions of heat flow by arrows when the temperature equals **(a)** $x^3 - 3xy^2$, **(b)** $\sin x \sinh y$, and **(c)** $e^x \cos y$.
28. **Steepest ascent.** If $z(x, y) = 3000 - x^2 - 9y^2$ [meters] gives the elevation of a mountain at sea level, what is the direction of steepest ascent at P: $(4, 1)$?
29. **Gradient.** What does it mean if $|\nabla f(P)| > |\nabla f(Q)|$ at two points P and Q in a scalar field?

9.8 Divergence of a Vector Field

Vector calculus owes much of its importance in engineering and physics to the gradient, divergence, and curl. From a scalar field we can obtain a vector field by the gradient (Sec. 9.7). Conversely, from a vector field we can obtain a scalar field by the divergence or another vector field by the curl (to be discussed in Sec. 9.9). These concepts were suggested by basic physical applications. This will be evident from our examples.

To begin, let $\mathbf{v}(x, y, z)$ be a differentiable vector function, where x, y, z are Cartesian coordinates, and let v_1, v_2, v_3 be the components of $\mathbf{v}$. Then the function

$$\text{div}\,\mathbf{v} = \frac{\partial v_1}{\partial x} + \frac{\partial v_2}{\partial y} + \frac{\partial v_3}{\partial z} \tag{1}$$

is called the **divergence** of $\mathbf{v}$ or the *divergence of the vector field defined by* $\mathbf{v}$. For example, if

$$\mathbf{v} = [3xz, 2xy, -yz^2] = 3xz\mathbf{i} + 2xy\mathbf{j} - yz^2\mathbf{k}, \qquad \text{then} \qquad \operatorname{div} \mathbf{v} = 3z + 2x - 2yz.$$

Another common notation for the divergence is

$$\begin{aligned} \operatorname{div} \mathbf{v} = \nabla \bullet \mathbf{v} &= \left[\frac{\partial}{\partial x}, \frac{\partial}{\partial \mathbf{y}}, \frac{\partial}{\partial \mathbf{z}}\right] \bullet [v_1, v_2, v_3] \\ &= \left(\frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}\right) \bullet (v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}) \\ &= \frac{\partial v_1}{\partial x} + \frac{\partial v_2}{\partial y} + \frac{\partial v_3}{\partial z}, \end{aligned}$$

with the understanding that the "product" $(\partial/\partial x)v_1$ in the dot product means the partial derivative $\partial v_1/\partial x$, etc. This is a convenient notation, but nothing more. Note that $\nabla \bullet \mathbf{v}$ means the scalar div $\mathbf{v}$, whereas ∇f means the vector grad f defined in Sec. 9.7.

In Example 2 we shall see that the divergence has an important physical meaning. Clearly, the values of a function that characterizes a physical or geometric property must be independent of the particular choice of coordinates. In other words, these values must be invariant with respect to coordinate transformations. Accordingly, the following theorem should hold.

THEOREM 1

Invariance of the Divergence

The divergence div $\mathbf{v}$ *is a scalar function, that is, its values depend only on the points in space (and, of course, on* $\mathbf{v}$*) but not on the choice of the coordinates in* (1)*, so that with respect to other Cartesian coordinates* x^*, y^*, z^* *and corresponding components* v_1^*, v_2^*, v_3^* *of* $\mathbf{v}$,

$$\operatorname{div} \mathbf{v} = \frac{\partial v_1^*}{\partial x^*} + \frac{\partial v_2^*}{\partial y^*} + \frac{\partial v_3^*}{\partial z^*}. \tag{2}$$

We shall prove this theorem in Sec. 10.7, using integrals.

Presently, let us turn to the more immediate practical task of gaining a feel for the significance of the divergence. Let $f(x, y, z)$ be a twice differentiable scalar function. Then its gradient exists,

$$\mathbf{v} = \operatorname{grad} f = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right] = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} + \frac{\partial f}{\partial z}\mathbf{k}$$

and we can differentiate once more, the first component with respect to x, the second with respect to y, the third with respect to z, and then form the divergence,

$$\operatorname{div} \mathbf{v} = \operatorname{div}(\operatorname{grad} f) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}.$$

Hence we have the basic result that *the divergence of the gradient is the Laplacian* (Sec. 9.7),

$$\text{div}\,(\text{grad}\,f) = \nabla^2 f. \tag{3}$$

EXAMPLE 1 Gravitational Force. Laplace's Equation

The gravitational force $\mathbf{p}$ in Theorem 3 of the last section is the gradient of the scalar function $f(x, y, z) = c/r$, which satisfies Laplaces equation $\nabla^2 f = 0$. According to (3) this implies that div $\mathbf{p} = 0$ $(r > 0)$. ■

The following example from hydrodynamics shows the physical significance of the divergence of a vector field. We shall get back to this topic in Sec. 10.8 and add further physical details.

EXAMPLE 2 Flow of a Compressible Fluid. Physical Meaning of the Divergence

We consider the motion of a fluid in a region R having no **sources** or **sinks** in R, that is, no points at which fluid is produced or disappears. The concept of **fluid state** is meant to cover also gases and vapors. Fluids in the restricted sense, or liquids, such as water or oil, have very small compressibility, which can be neglected in many problems. In contrast, gases and vapors have high compressibility. Their density ρ (= mass per unit volume) depends on the coordinates x, y, z in space and may also depend on time t. We assume that our fluid is compressible. We consider the flow through a rectangular box B of small edges Δx, Δy, Δz parallel to the coordinate axes as shown in Fig. 218. (Here Δ is a standard notation for small quantities and, of course, has nothing to do with the notation for the Laplacian in (11) of Sec. 9.7.) The box B has the volume $\Delta V = \Delta x\,\Delta y\,\Delta z$. Let $\mathbf{v} = [v_1, v_2, v_3] = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$ be the velocity vector of the motion. We set

$$\mathbf{u} = \rho\mathbf{v} = [u_1, u_2, u_3] = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k} \tag{4}$$

and assume that $\mathbf{u}$ and $\mathbf{v}$ are continuously differentiable vector functions of x, y, z, and t, that is, they have first partial derivatives which are continuous. Let us calculate the change in the mass included in B by considering the **flux** across the boundary, that is, the total loss of mass leaving B per unit time. Consider the flow through the left of the three faces of B that are visible in Fig. 218, whose area is $\Delta x\,\Delta z$. Since the vectors $v_1\mathbf{i}$ and $v_3\,\mathbf{k}$ are parallel to that face, the components v_1 and v_3 of $\mathbf{v}$ contribute nothing to this flow. Hence the mass of fluid entering through that face during a short time interval Δt is given approximately by

$$(\rho v_2)_y\,\Delta x\,\Delta z\,\Delta t = (u_2)_y\,\Delta x\,\Delta z\,\Delta t,$$

where the subscript y indicates that this expression refers to the left face. The mass of fluid leaving the box B through the opposite face during the same time interval is approximately $(u_2)_{y+\Delta y}\,\Delta x\,\Delta z\,\Delta t$, where the subscript $y + \Delta y$ indicates that this expression refers to the right face (which is not visible in Fig. 218). The difference

$$\Delta u_2\,\Delta x\,\Delta z\,\Delta t = \frac{\Delta u_2}{\Delta y}\,\Delta V\,\Delta t \qquad [\Delta u_2 = (u_2)_{y+\Delta y} - (u_2)_y]$$

is the approximate loss of mass. Two similar expressions are obtained by considering the other two pairs of parallel faces of B. If we add these three expressions, we find that the total loss of mass in B during the time interval Δt is approximately

$$\left(\frac{\Delta u_1}{\Delta x} + \frac{\Delta u_2}{\Delta y} + \frac{\Delta u_3}{\Delta z}\right)\Delta V\,\Delta t,$$

where

$$\Delta u_1 = (u_1)_{x+\Delta x} - (u_1)_x \qquad \text{and} \qquad \Delta u_3 = (u_3)_{z+\Delta z} - (u_3)_z.$$

This loss of mass in B is caused by the time rate of change of the density and is thus equal to

$$-\frac{\partial\rho}{\partial t}\,\Delta V\,\Delta t.$$

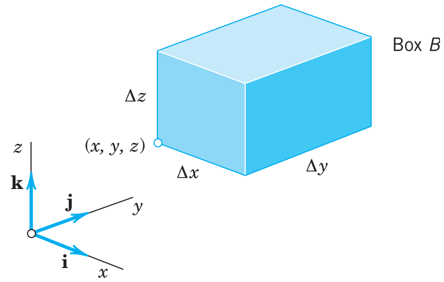


Fig. 218. Physical interpretation of the divergence

If we equate both expressions, divide the resulting equation by $\Delta V \Delta t$, and let Δx, Δy, Δz, and Δt approach zero, then we obtain

$$\operatorname{div} \mathbf{u} = \operatorname{div} (\rho \mathbf{v}) = -\frac{\partial \rho}{\partial t}$$

or

$$\frac{\partial \rho}{\partial t} + \operatorname{div} (\rho \mathbf{v}) = 0. \tag{5}$$

This important relation is called the *condition for the conservation of mass* or the **continuity equation** *of a compressible fluid flow.*

If the flow is **steady**, that is, independent of time, then $\partial \rho / \partial t = 0$ and the continuity equation is

$$\operatorname{div} (\rho \mathbf{v}) = 0. \tag{6}$$

If the density ρ is constant, so that the fluid is incompressible, then equation (6) becomes

$$\operatorname{div} \mathbf{v} = 0. \tag{7}$$

This relation is known as the **condition of incompressibility**. It expresses the fact that the balance of outflow and inflow for a given volume element is zero at any time. Clearly, the assumption that the flow has no sources or sinks in R is essential to our argument. $\mathbf{v}$ is also referred to as **solenoidal**.

From this discussion you should conclude and remember that, roughly speaking, ***the divergence measures outflow minus inflow***. ■

Comment. The **divergence theorem** of Gauss, an integral theorem involving the divergence, follows in the next chapter (Sec. 10.7).

PROBLEM SET 9.8

1–6 CALCULATION OF THE DIVERGENCE

Find div $\mathbf{v}$ and its value at P.

1. $\mathbf{v} = [x^2, 4y^2, 9z^2]$, P: $(-1, 0, \frac{1}{2})$

2. $\mathbf{v} = [0, \cos xyz, \sin xyz]$, P: $(2, \frac{1}{2}\pi, 0]$

3. $\mathbf{v} = (x^2 + y^2)^{-1}[x, y]$

4. $\mathbf{v} = [v_1(y, z), v_2(z, x), v_3(x, y)]$, P: $(3, 1, -1)]$

5. $\mathbf{v} = x^2y^2z^2[x, y, z]$, P: $(3, -1, 4)$

6. $\mathbf{v} = (x^2 + y^2 + z^2)^{-3/2}[x, y, z]$

7. For what v_3 is $\mathbf{v} = [e^x \cos y, e^x \sin y, v_3]$ solenoidal?

8. Let $\mathbf{v} = [x, y, v_3]$. Find a v_3 such that **(a)** div $\mathbf{v} > 0$ everywhere, **(b)** div $\mathbf{v} > 0$ if $|z| < 1$ and div $\mathbf{v} < 0$ if $|z| > 1$.

9. PROJECT. Useful Formulas for the Divergence. Prove

(a) $\operatorname{div}(k\mathbf{v}) = k \operatorname{div} \mathbf{v}$ (k constant)

(b) $\operatorname{div}(f\mathbf{v}) = f \operatorname{div} \mathbf{v} + \mathbf{v} \cdot \nabla f$

(c) $\operatorname{div}(f\nabla g) = f\nabla^2 g + \nabla f \cdot \nabla g$

(d) $\operatorname{div}(f\nabla g) - \operatorname{div}(g\nabla f) = f\nabla^2 g - g\nabla^2 f$

Verify (b) for $f = e^{xyz}$ and $\mathbf{v} = ax\mathbf{i} + by\mathbf{j} + cz\mathbf{k}$. Obtain the answer to Prob. 6 from (b). Verify (c) for $f = x^2 - y^2$ and $g = e^{x+y}$. Give examples of your own for which (a)–(d) are advantageous.

10. CAS EXPERIMENT. Visualizing the Divergence. Graph the given velocity field $\mathbf{v}$ of a fluid flow in a square centered at the origin with sides parallel to the coordinate axes. Recall that the divergence measures outflow minus inflow. By looking at the flow near the sides of the square, can you see whether div $\mathbf{v}$ must be positive or negative or may perhaps be zero? Then calculate div $\mathbf{v}$. First do the given flows and then do some of your own. Enjoy it.

(a) $\mathbf{v} = \mathbf{i}$

(b) $\mathbf{v} = x\mathbf{i}$

(c) $\mathbf{v} = x\mathbf{i} - y\mathbf{j}$

(d) $\mathbf{v} = x\mathbf{i} + y\mathbf{j}$

(e) $\mathbf{v} = -x\mathbf{i} - y\mathbf{j}$

(f) $\mathbf{v} = (x^2 + y^2)^{-1}(-y\mathbf{i} + x\mathbf{j})$

11. Incompressible flow. Show that the flow with velocity vector $\mathbf{v} = y\mathbf{i}$ is incompressible. Show that the particles that at time $t = 0$ are in the cube whose faces are portions of the planes $x = 0, x = 1, y = 0, y = 1, z = 0, z = 1$ occupy at $t = 1$ the volume 1.

12. Compressible flow. Consider the flow with velocity vector $\mathbf{v} = x\mathbf{i}$. Show that the individual particles have the position vectors $\mathbf{r}(t) = c_1e^t\mathbf{i} + c_2\mathbf{j} + c_3\mathbf{k}$ with constant c_1, c_2, c_3. Show that the particles that at $t = 0$ are in the cube of Prob. 11 at $t = 1$ occupy the volume e.

13. Rotational flow. The velocity vector $\mathbf{v}(x, y, z)$ of an incompressible fluid rotating in a cylindrical vessel is of the form $\mathbf{v} = \mathbf{w} \times \mathbf{r}$, where $\mathbf{w}$ is the (constant) rotation vector; see Example 5 in Sec. 9.3. Show that div $\mathbf{v} = 0$. Is this plausible because of our present Example 2?

14. Does div $\mathbf{u}$ = div $\mathbf{v}$ imply $\mathbf{u} = \mathbf{v}$ or $\mathbf{u} = \mathbf{v} + \mathbf{k}$ ($\mathbf{k}$ constant)? Give reason.

15–20 LAPLACIAN

Calculate $\nabla^2 f$ by Eq. (3). Check by direct differentiation. Indicate when (3) is simpler. Show the details of your work.

15. $f = \cos^2 x + \sin^2 y$

16. $f = e^{xyz}$

17. $f = \ln(x^2 + y^2)$

18. $f = z - \sqrt{x^2 + y^2}$

19. $f = 1/(x^2 + y^2 + z^2)$

20. $f = e^{2x} \cosh 2y$

9.9 Curl of a Vector Field

The concepts of gradient (Sec. 9.7), divergence (Sec. 9.8), and curl are of fundamental importance in vector calculus and frequently applied in vector fields. In this section we define and discuss the concept of the curl and apply it to several engineering problems.

Let $\mathbf{v}(x, y, z) = [v_1, v_2, v_3] = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$ be a differentiable vector function of the Cartesian coordinates x, y, z. Then the **curl** *of the vector function* $\mathbf{v}$ or *of the vector field given by* $\mathbf{v}$ is defined by the "symbolic" determinant

$$\text{(1)} \qquad \operatorname{curl} \mathbf{v} = \nabla \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ v_1 & v_2 & v_3 \end{vmatrix}$$

$$= \left(\frac{\partial v_3}{\partial y} - \frac{\partial v_2}{\partial z}\right)\mathbf{i} + \left(\frac{\partial v_1}{\partial z} - \frac{\partial v_3}{\partial x}\right)\mathbf{j} + \left(\frac{\partial v_2}{\partial x} - \frac{\partial v_1}{\partial y}\right)\mathbf{k}.$$

This is the formula when x, y, z are *right-handed*. If they are *left-handed*, the determinant has a minus sign in front (just as in (2**) in Sec. 9.3).

Instead of curl $\mathbf{v}$ one also uses the notation rot $\mathbf{v}$. This is suggested by "rotation," an application explored in Example 2. Note that curl $\mathbf{v}$ is a vector, as shown in Theorem 3.

EXAMPLE 1 Curl of a Vector Function

Let $\mathbf{v} = [yz, \quad 3zx, \quad z] = yz\mathbf{i} + 3zx\mathbf{j} + z\mathbf{k}$ with right-handed x, y, z. Then (1) gives

$$\operatorname{curl} \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ yz & 3zx & z \end{vmatrix} = -3x\mathbf{i} + y\mathbf{j} + (3z - z)\mathbf{k} = -3x\mathbf{i} + y\mathbf{j} + 2z\mathbf{k}.$$

The curl has many applications. A typical example follows. More about the nature and significance of the curl will be considered in Sec. 10.9.

EXAMPLE 2 Rotation of a Rigid Body. Relation to the Curl

We have seen in Example 5, Sec. 9.3, that a rotation of a rigid body B about a fixed axis in space can be described by a vector $\mathbf{w}$ of magnitude ω in the direction of the axis of rotation, where ω (>0) is the angular speed of the rotation, and $\mathbf{w}$ is directed so that the rotation appears clockwise if we look in the direction of $\mathbf{w}$. According to (9), Sec. 9.3, the velocity field of the rotation can be represented in the form

$$\mathbf{v} = \mathbf{w} \times \mathbf{r}$$

where $\mathbf{r}$ is the position vector of a moving point with respect to a Cartesian coordinate system *having the origin on the axis of rotation*. Let us choose right-handed Cartesian coordinates such that the axis of rotation is the z-axis. Then (see Example 2 in Sec. 9.4)

$$\mathbf{w} = [0, \quad 0, \quad \omega] = \omega\mathbf{k}, \qquad \mathbf{v} = \mathbf{w} \times \mathbf{r} = [-\omega y, \quad \omega x, \quad 0] = -\omega y\mathbf{i} + \omega x\mathbf{j}.$$

Hence

$$\operatorname{curl} \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ -\omega y & \omega x & 0 \end{vmatrix} = [0, \quad 0, \quad 2\omega] = 2\omega\mathbf{k} = 2\mathbf{w}.$$

This proves the following theorem.

THEOREM 1 Rotating Body and Curl

The curl of the velocity field of a rotating rigid body has the direction of the axis of the rotation, and its magnitude equals twice the angular speed of the rotation.

Next we show how the grad, div, and curl are interrelated, thereby shedding further light on the nature of the curl.

THEOREM 2

Grad, Div, Curl

Gradient fields are ***irrotational***. *That is, if a continuously differentiable vector function is the gradient of a scalar function f, then its curl is the zero vector,*

$$\text{curl}\,(\text{grad}\, f) = \mathbf{0}. \tag{2}$$

Furthermore, the divergence of the curl of a twice continuously differentiable vector function **v** *is zero,*

$$\text{div}\,(\text{curl}\,\mathbf{v}) = 0. \tag{3}$$

PROOF Both (2) and (3) follow directly from the definitions by straightforward calculation. In the proof of (3) the six terms cancel in pairs. ■

EXAMPLE 3 **Rotational and Irrotational Fields**

The field in Example 2 is not irrotational. A similar velocity field is obtained by stirring tea or coffee in a cup. The gravitational field in Theorem 3 of Sec. 9.7 has curl $\mathbf{p} = \mathbf{0}$. It is an irrotational gradient field. ■

The term "irrotational" for curl $\mathbf{v} = \mathbf{0}$ is suggested by the use of the curl for characterizing the rotation in a field. If a gradient field occurs elsewhere, not as a velocity field, it is usually called **conservative** (see Sec. 9.7). Relation (3) is plausible because of the interpretation of the curl as a rotation and of the divergence as a flux (see Example 2 in Sec. 9.8).

Finally, since the curl is defined in terms of coordinates, we should do what we did for the gradient in Sec. 9.7, namely, to find out whether the curl is a vector. This is true, as follows.

THEOREM 3

Invariance of the Curl

curl **v** is a vector. It has a length and a direction that are independent of the particular choice of a Cartesian coordinate system in space.

PROOF The proof is quite involved and shown in App. 4.

We have completed our discussion of vector differential calculus. The companion Chap. 10 on vector integral calculus follows and makes use of many concepts covered in this chapter, including dot and cross products, parametric representation of curves C, along with grad, div, and curl.

PROBLEM SET 9.9

1. **WRITING REPORT. Grad, div, curl.** List the definitions and most important facts and formulas for grad, div, curl, and ∇^2. Use your list to write a corresponding report of 3–4 pages, with examples of your own. No proofs.
2. **(a)** What direction does curl **v** have if **v** is parallel to the yz-plane? **(b)** If, moreover, **v** is independent of x?
3. Prove Theorem 2. Give two examples for (2) and (3) each.

4–8 CALCULUTION OF CURL

Find curl **v** for **v** given with respect to right-handed Cartesian coordinates. Show the details of your work.

4. $\mathbf{v} = [2y^2, 5x, 0]$
5. $\mathbf{v} = xyz[x, y, z]$
6. $\mathbf{v} = (x^2 + y^2 + z^2)^{-3/2}[x, y, z]$
7. $\mathbf{v} = [0, 0, e^{-x} \sin y]$
8. $\mathbf{v} = [e^{-z^2}, e^{-x^2}, e^{-y^2}]$

9–13 FLUID FLOW

Let **v** be the velocity vector of a steady fluid flow. Is the flow irrotational? Incompressible? Find the streamlines (the paths of the particles). *Hint.* See the answers to Probs. 9 and 11 for a determination of a path.

9. $\mathbf{v} = [0, 3z^2, 0]$
10. $\mathbf{v} = [\sec x, \csc x, 0]$
11. $\mathbf{v} = [y, -2x, 0]$
12. $\mathbf{v} = [-y, x, \pi]$
13. $\mathbf{v} = [x, y, -z]$

14. **PROJECT. Useful Formulas for the Curl.** Assuming sufficient differentiability, show that
    (a) $\operatorname{curl}(\mathbf{u} + \mathbf{v}) = \operatorname{curl}\mathbf{u} + \operatorname{curl}\mathbf{v}$
    (b) $\operatorname{div}(\operatorname{curl}\mathbf{v}) = 0$
    (c) $\operatorname{curl}(f\mathbf{v}) = (\operatorname{grad} f) \times \mathbf{v} + f \operatorname{curl}\mathbf{v}$
    (d) $\operatorname{curl}(\operatorname{grad} f) = \mathbf{0}$
    (e) $\operatorname{div}(\mathbf{u} \times \mathbf{v}) = \mathbf{v} \cdot \operatorname{curl}\mathbf{u} - \mathbf{u} \cdot \operatorname{curl}\mathbf{v}$

15–20 DIV AND CURL

With respect to right-handed coordinates, let $\mathbf{u} = [y, z, x]$, $\mathbf{v} = [yz, zx, xy]$, $f = xyz$, and $g = x + y + z$. Find the given expressions. Check your result by a formula in Proj. 14 if applicable.

15. $\operatorname{curl}(\mathbf{u} + \mathbf{v})$, $\operatorname{curl}\mathbf{v}$
16. $\operatorname{curl}(g\mathbf{v})$
17. $\mathbf{v} \cdot \operatorname{curl}\mathbf{u}$, $\mathbf{u} \cdot \operatorname{curl}\mathbf{v}$, $\mathbf{u} \cdot \operatorname{curl}\mathbf{u}$
18. $\operatorname{div}(\mathbf{u} \times \mathbf{v})$
19. $\operatorname{curl}(g\mathbf{u} + \mathbf{v})$, $\operatorname{curl}(g\mathbf{u})$
20. $\operatorname{div}(\operatorname{grad}(fg))$

CHAPTER 9 REVIEW QUESTIONS AND PROBLEMS

1. What is a vector? A vector function? A vector field? A scalar? A scalar function? A scalar field? Give examples.
2. What is an inner product, a vector product, a scalar triple product? What applications motivate these products?
3. What are right-handed and left-handed coordinates? When is this distinction important?
4. When is a vector product the zero vector? What is orthogonality?
5. How is the derivative of a vector function defined? What is its significance in geometry and mechanics?
6. If $\mathbf{r}(t)$ represents a motion, what are $\mathbf{r}'(t)$, $|\mathbf{r}'(t)|$, $\mathbf{r}''(t)$, and $|\mathbf{r}''(t)|$?
7. Can a moving body have constant speed but variable velocity? Nonzero acceleration?
8. What do you know about directional derivatives? Their relation to the gradient?
9. Write down the definitions and explain the significance of grad, div, and curl.
10. Granted sufficient differentiability, which of the following expressions make sense? $f \operatorname{curl}\mathbf{v}$, $\mathbf{v} \operatorname{curl} f$, $\mathbf{u} \times \mathbf{v}$, $\mathbf{u} \times \mathbf{v} \times \mathbf{w}$, $f \cdot \mathbf{v}$, $f \cdot (\mathbf{v} \times \mathbf{w})$, $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$, $\mathbf{v} \times \operatorname{curl}\mathbf{v}$, $\operatorname{div}(f\mathbf{v})$, $\operatorname{curl}(f\mathbf{v})$, and $\operatorname{curl}(f \cdot \mathbf{v})$.

11–19 ALGEBRAIC OPERATIONS FOR VECTORS

Let $\mathbf{a} = [4, 7, 0]$, $\mathbf{b} = [3, -1, 5]$, $\mathbf{c} = [-6, 2, 0]$, and $\mathbf{d} = [1, -2, 8]$. Calculate the following expressions. Try to make a sketch.

11. $\mathbf{a} \cdot \mathbf{c}$, $3\mathbf{b} \cdot 8\mathbf{d}$, $24\mathbf{d} \cdot \mathbf{b}$, $\mathbf{a} \cdot \mathbf{a}$
12. $\mathbf{a} \times \mathbf{c}$, $\mathbf{b} \times \mathbf{d}$, $\mathbf{d} \times \mathbf{b}$, $\mathbf{a} \times \mathbf{a}$
13. $\mathbf{b} \times \mathbf{c}$, $\mathbf{c} \times \mathbf{b}$, $\mathbf{c} \times \mathbf{c}$, $\mathbf{c} \cdot \mathbf{c}$
14. $5(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$, $\mathbf{a} \cdot (5\mathbf{b} \times \mathbf{c})$, $(5\mathbf{a}\ \mathbf{b}\ \mathbf{c})$, $5(\mathbf{a} \cdot \mathbf{b}) \times \mathbf{c}$
15. $6(\mathbf{a} \times \mathbf{b}) \times \mathbf{d}$, $\mathbf{a} \times 6(\mathbf{b} \times \mathbf{d})$, $2\mathbf{a} \times 3\mathbf{b} \times \mathbf{d}$
16. $(1/|\mathbf{a}|)\mathbf{a}$, $(1/|\mathbf{b}|)\mathbf{b}$, $\mathbf{a} \cdot \mathbf{b}/|\mathbf{b}|$, $\mathbf{a} \cdot \mathbf{b}/|\mathbf{a}|$
17. $(\mathbf{a}\ \mathbf{b}\ \mathbf{d})$, $(\mathbf{b}\ \mathbf{a}\ \mathbf{d})$, $(\mathbf{b}\ \mathbf{d}\ \mathbf{a})$
18. $|\mathbf{a} + \mathbf{b}|$, $|\mathbf{a}| + |\mathbf{b}|$
19. $\mathbf{a} \times \mathbf{b} - \mathbf{b} \times \mathbf{a}$, $(\mathbf{a} \times \mathbf{c}) \cdot \mathbf{c}$, $|\mathbf{a} \times \mathbf{b}|$
20. **Commutativity.** When is $\mathbf{u} \times \mathbf{v} = \mathbf{v} \times \mathbf{u}$? When is $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$?
21. **Resultant, equilibrium.** Find **u** such that **u** and **a**, **b**, **c**, **d** above and **u** are in equilibrium.
22. **Resultant.** Find the most general **v** such that the resultant of **v**, **a**, **b**, **c** (see above) is parallel to the *yz*-plane.
23. **Angle.** Find the angle between **a** and **c**. Between **b** and **d**. Sketch **a** and **c**.
24. **Planes.** Find the angle between the two planes P_1: $4x - y + 3z = 12$ and P_2: $x + 2y + 4z = 4$. Make a sketch.
25. **Work.** Find the work done by $\mathbf{q} = [5, 2, 0]$ in the displacement from $(1, 1, 0)$ to $(4, 3, 0)$.
26. **Component.** When is the component of a vector **v** in the direction of a vector **w** equal to the component of **w** in the direction of **v**?
27. **Component.** Find the component of $\mathbf{v} = [4, 7, 0]$ in the direction of $\mathbf{w} = [2, 2, 0]$. Sketch it.

28. Moment. When is the moment of a force equal to zero?

29. Moment. A force $\mathbf{p} = [4, 2, 0]$ is acting in a line through (2, 3, 0). Find its moment vector about the center (5, 1, 0) of a wheel.

30. Velocity, acceleration. Find the velocity, speed, and acceleration of the motion given by $\mathbf{r}(t) = [3 \cos t, 3 \sin t, 4t]$ (t = time) at the point P: $(3/\sqrt{2}, 3/\sqrt{2}, \pi)$.

31. Tetrahedron. Find the volume if the vertices are (0, 0, 0), (3, 1, 2), (2, 4, 0), (5, 4, 0).

32–40 GRAD, DIV, CURL, ∇^2, $D_v f$

Let $f = xy - yz$, $\mathbf{v} = [2y, 2z, 4x + z]$, and $\mathbf{w} = [3z^2, x^2 - y^2, y^2]$. Find:

32. grad f and f grad f at P: (2, 7, 0)

33. div $\mathbf{v}$, div $\mathbf{w}$

34. curl $\mathbf{v}$, curl $\mathbf{w}$

35. div (grad f), $\nabla^2 f$, $\nabla^2(xyf)$

36. (curl $\mathbf{w}$) • $\mathbf{v}$ at (4, 0, 2)

37. grad (div $\mathbf{w}$)

38. $D_v f$ at P: (1, 1, 2)

39. $D_w f$ at P: (3, 0, 2)

40. $\mathbf{v} \cdot ((\text{curl } \mathbf{w}) \times \mathbf{v})$

SUMMARY OF CHAPTER 9
Vector Differential Calculus. Grad, Div, Curl

All vectors of the form $\mathbf{a} = [a_1, a_2, a_3] = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ constitute the **real vector space** R^3 with componentwise vector addition

$$(1) \qquad [a_1, a_2, a_3] + [b_1, b_2, b_3] = [a_1 + b_1, a_2 + b_2, a_3 + b_3]$$

and componentwise scalar multiplication (c a scalar, a real number)

$$(2) \qquad c[a_1, a_2, a_3] = [ca_1, ca_2, ca_3] \qquad \text{(Sec. 9.1).}$$

For instance, the *resultant* of forces $\mathbf{a}$ and $\mathbf{b}$ is the sum $\mathbf{a} + \mathbf{b}$.

The **inner product** or **dot product** of two vectors is defined by

$$(3) \qquad \mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos \gamma = a_1b_1 + a_2b_2 + a_3b_3 \qquad \text{(Sec. 9.2)}$$

where γ is the angle between $\mathbf{a}$ and $\mathbf{b}$. This gives for the **norm** or **length** $|\mathbf{a}|$ of $\mathbf{a}$

$$(4) \qquad |\mathbf{a}| = \sqrt{\mathbf{a} \cdot \mathbf{a}} = \sqrt{a_1^2 + a_2^2 + a_3^2}$$

as well as a formula for γ. If $\mathbf{a} \cdot \mathbf{b} = 0$, we call $\mathbf{a}$ and $\mathbf{b}$ **orthogonal**. The dot product is suggested by the *work* $W = \mathbf{p} \cdot \mathbf{d}$ done by a force $\mathbf{p}$ in a displacement $\mathbf{d}$.

The **vector product** or **cross product** $\mathbf{v} = \mathbf{a} \times \mathbf{b}$ is a vector of length

$$(5) \qquad |\mathbf{a} \times \mathbf{b}| = |\mathbf{a}||\mathbf{b}| \sin \gamma \qquad \text{(Sec. 9.3)}$$

and perpendicular to both $\mathbf{a}$ and $\mathbf{b}$ such that $\mathbf{a}$, $\mathbf{b}$, $\mathbf{v}$ form a *right-handed* triple. In terms of components with respect to right-handed coordinates,

$$(6) \qquad \mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \qquad \text{(Sec. 9.3).}$$

The vector product is suggested, for instance, by moments of forces or by rotations. CAUTION! This multiplication is *anti*commutative, $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$, and is *not* associative.

An (oblique) box with edges $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ has volume equal to the absolute value of the **scalar triple product**

$$(\mathbf{a} \quad \mathbf{b} \quad \mathbf{c}) = \mathbf{a} \bullet (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) \bullet \mathbf{c}. \tag{7}$$

Sections 9.4–9.9 extend differential calculus to vector functions

$$\mathbf{v}(t) = [v_1(t), v_2(t), v_3(t)] = v_1(t)\mathbf{i} + v_2(t)\mathbf{j} + v_3(t)\mathbf{k}$$

and to vector functions of more than one variable (see below). The derivative of $\mathbf{v}(t)$ is

$$\mathbf{v}' = \frac{d\mathbf{v}}{dt} = \lim_{\Delta t \to 0} \frac{\mathbf{v}(t + \Delta t) - \mathbf{v}(t)}{\Delta t} = [v_1', v_2', v_3'] = v_1'\mathbf{i} + v_2'\mathbf{j} + v_3'\mathbf{k}. \tag{8}$$

Differentiation rules are as in calculus. They imply (Sec. 9.4)

$$(\mathbf{u} \bullet \mathbf{v})' = \mathbf{u}' \bullet \mathbf{v} + \mathbf{u} \bullet \mathbf{v}', \qquad (\mathbf{u} \times \mathbf{v})' = \mathbf{u}' \times \mathbf{v} + \mathbf{u} \times \mathbf{v}'.$$

Curves C in space represented by the position vector $\mathbf{r}(t)$ have $\mathbf{r}'(t)$ as a **tangent vector** (the **velocity** in mechanics when t is time), $\mathbf{r}'(s)$ (s arc length, Sec. 9.5) as the *unit tangent vector,* and $|\mathbf{r}''(s)| = \kappa$ as the *curvature* (the *acceleration* in mechanics).

Vector functions $\mathbf{v}(x, y, z) = [v_1(x, y, z), v_2(x, y, z), v_3(x, y, z)]$ represent vector fields in space. Partial derivatives with respect to the Cartesian coordinates x, y, z are obtained componentwise, for instance,

$$\frac{\partial \mathbf{v}}{\partial x} = \left[\frac{\partial v_1}{\partial x}, \frac{\partial v_2}{\partial x}, \frac{\partial v_3}{\partial x}\right] = \frac{\partial v_1}{\partial x}\mathbf{i} + \frac{\partial v_2}{\partial x}\mathbf{j} + \frac{\partial v_3}{\partial x}\mathbf{k} \qquad \text{(Sec. 9.6).}$$

The **gradient** of a scalar function f is

$$\operatorname{grad} f = \nabla f = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right] \qquad \text{(Sec. 9.7).} \tag{9}$$

The **directional derivative** of f in the direction of a vector $\mathbf{a}$ is

$$D_{\mathbf{a}} f = \frac{df}{ds} = \frac{1}{|\mathbf{a}|} \mathbf{a} \bullet \nabla f \qquad \text{(Sec. 9.7).} \tag{10}$$

The **divergence** of a vector function $\mathbf{v}$ is

$$\operatorname{div} \mathbf{v} = \nabla \bullet \mathbf{v} = \frac{\partial v_1}{\partial x} + \frac{\partial v_2}{\partial y} + \frac{\partial v_3}{\partial z}. \qquad \text{(Sec. 9.8).} \tag{11}$$

CHAPTER 9

Vector Differential Calculus. Grad, Div, Curl

Engineering, physics, and computer sciences, in general, but particularly solid mechanics, aerodynamics, aeronautics, fluid flow, heat flow, electrostatics, quantum physics, laser technology, robotics as well as other areas have applications that require an understanding of **vector calculus**. This field encompasses vector differential calculus and vector integral calculus. Indeed, the engineer, physicist, and mathematician need a good grounding in these areas as provided by the carefully chosen material of Chaps. 9 and 10.

Forces, velocities, and various other quantities may be thought of as vectors. Vectors appear frequently in the applications above and also in the biological and social sciences, so it is natural that problems are modeled in **3-space**. This is the space of three dimensions with the usual measurement of distance, as given by the Pythagorean theorem. Within that realm, **2-space** (the plane) is a special case. Working in 3-space requires that we extend the common differential calculus to vector differential calculus, that is, the calculus that deals with vector functions and vector fields and is explained in this chapter.

Chapter 9 is arranged in three groups of sections. Sections 9.1–9.3 extend the basic algebraic operations of vectors into 3-space. These operations include the inner product and the cross product. Sections 9.4 and 9.5 form the heart of vector differential calculus. Finally, Secs. 9.7–9.9 discuss three physically important concepts related to scalar and vector fields: gradient (Sec. 9.7), divergence (Sec. 9.8), and curl (Sec. 9.9). They are expressed in Cartesian coordinates in this chapter and, if desired, expressed in ***curvilinear coordinates*** in a short section in App. A3.4.

We shall keep this chapter ***independent of Chaps. 7 and 8***. Our present approach is in harmony with Chap. 7, with the restriction to two and three dimensions providing for a richer theory with basic physical, engineering, and geometric applications.

Prerequisite: Elementary use of second- and third-order determinants in Sec. 9.3.
Sections that may be omitted in a shorter course: 9.5, 9.6.
References and Answers to Problems: App. 1 Part B, App. 2.

9.1 Vectors in 2-Space and 3-Space

In engineering, physics, mathematics, and other areas we encounter two kinds of quantities. They are scalars and vectors.

A **scalar** is a quantity that is determined by its magnitude. It takes on a numerical value, i.e., a number. Examples of scalars are time, temperature, length, distance, speed, density, energy, and voltage.

In contrast, a **vector** is a quantity that has both magnitude and direction. We can say that a vector is an ***arrow*** or a ***directed line segment***. For example, a velocity vector has length or magnitude, which is speed, and direction, which indicates the direction of motion. Typical examples of vectors are displacement, velocity, and force, see Fig. 164 as an illustration.

More formally, we have the following. We denote vectors by lowercase boldface letters **a**, **b**, **v**, etc. In handwriting you may use arrows, for instance, $\vec{a}$ (in place of **a**), $\vec{b}$, etc.

A vector (arrow) has a tail, called its **initial point**, and a tip, called its **terminal point**. This is motivated in the **translation** (displacement without rotation) of the triangle in Fig. 165, where the initial point P of the vector **a** is the original position of a point, and the terminal point Q is the terminal position of that point, its position *after* the translation. The length of the arrow equals the distance between P and Q. This is called the **length** (or *magnitude*) of the vector **a** and is denoted by $|\mathbf{a}|$. Another name for *length* is **norm** (or *Euclidean norm*).

A vector of length 1 is called a **unit vector**.

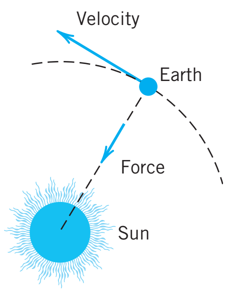


Fig. 164. Force and velocity

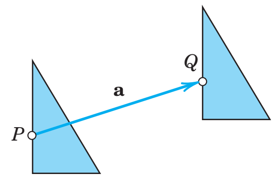


Fig. 165. Translation

Of course, we would like to calculate with vectors. For instance, we want to find the resultant of forces or compare parallel forces of different magnitude. This motivates our next ideas: to define *components* of a vector, and then the two basic algebraic operations of *vector addition* and *scalar multiplication.*

For this we must first define *equality of vectors* in a way that is practical in connection with forces and other applications.

DEFINITION

Equality of Vectors

Two vectors **a** and **b** are equal, written $\mathbf{a} = \mathbf{b}$, if they have the same length and the same direction [as explained in Fig. 166; in particular, note (B)]. Hence a vector can be arbitrarily translated; that is, its initial point can be chosen arbitrarily.

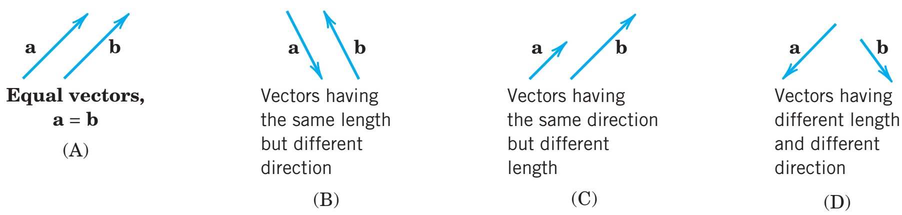


Fig. 166. (A) Equal vectors. (B)–(D) Different vectors

Components of a Vector

We choose an *xyz* **Cartesian coordinate system**[1] in space (Fig. 167), that is, a usual rectangular coordinate system with the same scale of measurement on the three mutually perpendicular coordinate axes. Let **a** be a given vector with initial point P: (x_1, y_1, z_1) and terminal point Q: (x_2, y_2, z_2). Then the three coordinate differences

$$a_1 = x_2 - x_1, \qquad a_2 = y_2 - y_1, \qquad a_3 = z_2 - z_1 \tag{1}$$

are called the **components** of the vector **a** with respect to that coordinate system, and we write simply $\mathbf{a} = [a_1, a_2, a_3]$. See Fig. 168.

The **length** $|\mathbf{a}|$ of **a** can now readily be expressed in terms of components because from (1) and the Pythagorean theorem we have

$$|\mathbf{a}| = \sqrt{a_1^2 + a_2^2 + a_3^2}. \tag{2}$$

EXAMPLE 1 Components and Length of a Vector

The vector **a** with initial point P: (4, 0, 2) and terminal point Q: (6, −1, 2) has the components

$$a_1 = 6 - 4 = 2, \qquad a_2 = -1 - 0 = -1, \qquad a_3 = 2 - 2 = 0.$$

Hence $\mathbf{a} = [2, -1, 0]$. (Can you sketch **a**, as in Fig. 168?) Equation (2) gives the length

$$|\mathbf{a}| = \sqrt{2^2 + (-1)^2 + 0^2} = \sqrt{5}.$$

If we choose (−1, 5, 8) as the initial point of **a**, the corresponding terminal point is (1, 4, 8).

If we choose the origin (0, 0, 0) as the initial point of **a**, the corresponding terminal point is (2, −1, 0); its coordinates equal the components of **a**. This suggests that we can determine each point in space by a vector, called the *position vector* of the point, as follows. ■

A Cartesian coordinate system being given, the **position vector r** of a point A: (x, y, z) is the vector with the origin (0, 0, 0) as the initial point and A as the terminal point (see Fig. 169). Thus in components, $\mathbf{r} = [x, y, z]$. This can be seen directly from (1) with $x_1 = y_1 = z_1 = 0$.

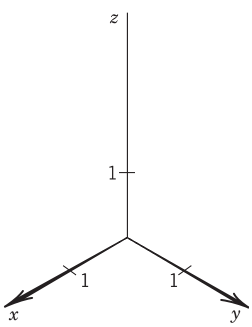


Fig. 167. Cartesian coordinate system

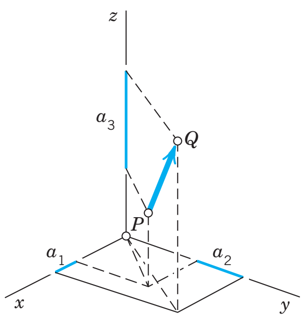


Fig. 168. Components of a vector

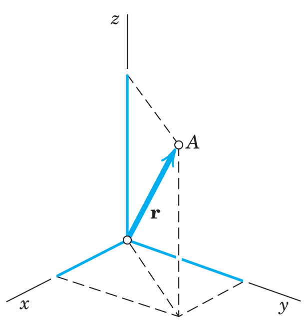


Fig. 169. Position vector **r** of a point A: (x, y, z)

[1]Named after the French philosopher and mathematician RENATUS CARTESIUS, latinized for RENÉ DESCARTES (1596–1650), who invented analytic geometry. His basic work *Géométrie* appeared in 1637, as an appendix to his *Discours de la méthode*.

Furthermore, if we translate a vector **a**, with initial point P and terminal point Q, then corresponding coordinates of P and Q change by the same amount, so that the differences in (1) remain unchanged. This proves

THEOREM 1

Vectors as Ordered Triples of Real Numbers

A fixed Cartesian coordinate system being given, each vector is uniquely determined by its ordered triple of corresponding components. Conversely, to each ordered triple of real numbers (a_1, a_2, a_3) *there corresponds precisely one vector* $\mathbf{a} = [a_1, a_2, a_3]$, *with* $(0, 0, 0)$ *corresponding to the* **zero vector 0**, *which has length* 0 *and no direction.*

Hence a vector equation $\mathbf{a} = \mathbf{b}$ *is equivalent to the three equations* $a_1 = b_1$, $a_2 = b_2$, $a_3 = b_3$ *for the components.*

We now see that from our "geometric" definition of a vector as an arrow we have arrived at an "algebraic" characterization of a vector by Theorem 1. We could have started from the latter and reversed our process. This shows that the two approaches are equivalent.

Vector Addition, Scalar Multiplication

Calculations with vectors are very useful and are almost as simple as the arithmetic for real numbers. Vector arithmetic follows almost naturally from applications. We first define how to add vectors and later on how to multiply a vector by a number.

DEFINITION

Addition of Vectors

The **sum** $\mathbf{a} + \mathbf{b}$ of two vectors $\mathbf{a} = [a_1, a_2, a_3]$ and $\mathbf{b} = [b_1, b_2, b_3]$ is obtained by adding the corresponding components,

$$\mathbf{a} + \mathbf{b} = [a_1 + b_1, \quad a_2 + b_2, \quad a_3 + b_3]. \tag{3}$$

Geometrically, place the vectors as in Fig. 170 (the initial point of **b** at the terminal point of **a**); then $\mathbf{a} + \mathbf{b}$ is the vector drawn from the initial point of **a** to the terminal point of **b**.

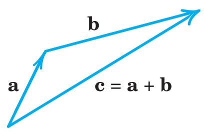


Fig. 170. Vector addition

For forces, this addition is the parallelogram law by which we obtain the **resultant** of two forces in mechanics. See Fig. 171.

Figure 172 shows (for the plane) that the "algebraic" way and the "geometric way" of vector addition give the same vector.

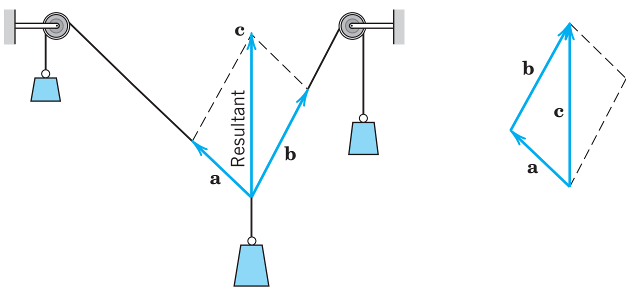


Fig. 171. Resultant of two forces (**parallelogram law**)

Basic Properties of Vector Addition. Familiar laws for real numbers give immediately

$$
\begin{aligned}
&\text{(a)} && \mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a} && \textit{(Commutativity)}\\
&\text{(b)} && (\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w}) && \textit{(Associativity)}\\
&\text{(c)} && \mathbf{a} + \mathbf{0} = \mathbf{0} + \mathbf{a} = \mathbf{a} &&\\
&\text{(d)} && \mathbf{a} + (-\mathbf{a}) = \mathbf{0}. &&
\end{aligned}
\tag{4}
$$

Properties (a) and (b) are verified geometrically in Figs. 173 and 174. Furthermore, $-\mathbf{a}$ denotes the vector having the length $|\mathbf{a}|$ and the direction opposite to that of $\mathbf{a}$.

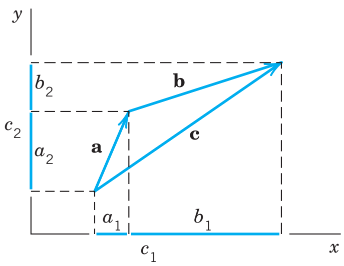


Fig. 172. Vector addition

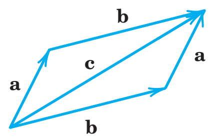


Fig. 173. Cummutativity of vector addition

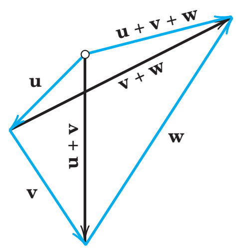


Fig. 174. Associativity of vector addition

In (4b) we may simply write $\mathbf{u} + \mathbf{v} + \mathbf{w}$, and similarly for sums of more than three vectors. Instead of $\mathbf{a} + \mathbf{a}$ we also write $2\mathbf{a}$, and so on. This (and the notation $-\mathbf{a}$ used just before) motivates defining the second algebraic operation for vectors as follows.

DEFINITION

Scalar Multiplication (Multiplication by a Number)

The product $c\mathbf{a}$ of any vector $\mathbf{a} = [a_1, a_2, a_3]$ and any scalar c (real number c) is the vector obtained by multiplying each component of $\mathbf{a}$ by c,

$$c\mathbf{a} = [ca_1, ca_2, ca_3]. \tag{5}$$

Geometrically, if $\mathbf{a} \neq \mathbf{0}$, then $c\mathbf{a}$ with $c > 0$ has the direction of $\mathbf{a}$ and with $c < 0$ the direction opposite to $\mathbf{a}$. In any case, the length of $c\mathbf{a}$ is $|c\mathbf{a}| = |c||\mathbf{a}|$, and $c\mathbf{a} = \mathbf{0}$ if $\mathbf{a} = \mathbf{0}$ or $c = 0$ (or both). (See Fig. 175.)

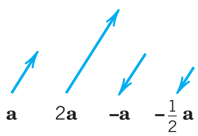


Fig. 175. Scalar multiplication [multiplication of vectors by scalars (numbers)]

Basic Properties of Scalar Multiplication. From the definitions we obtain directly

$$
\begin{aligned}
&\text{(a)} && c(\mathbf{a} + \mathbf{b}) = c\mathbf{a} + c\mathbf{b} &&\\
&\text{(b)} && (c + k)\mathbf{a} = c\mathbf{a} + k\mathbf{a} &&\\
&\text{(c)} && c(k\mathbf{a}) = (ck)\mathbf{a} && \text{(written } ck\mathbf{a}\text{)}\\
&\text{(d)} && 1\mathbf{a} = \mathbf{a}. &&
\end{aligned}
\tag{6}
$$

You may prove that (4) and (6) imply for any vector **a**

$$
\text{(7)} \qquad \begin{aligned} &\text{(a)} \quad && 0\mathbf{a} = \mathbf{0} \\ &\text{(b)} \quad && (-1)\mathbf{a} = -\mathbf{a}. \end{aligned}
$$

Instead of $\mathbf{b} + (-\mathbf{a})$ we simply write $\mathbf{b} - \mathbf{a}$ (Fig. 176).

EXAMPLE 2 Vector Addition. Multiplication by Scalars

With respect to a given coordinate system, let

$$\mathbf{a} = [4, 0, 1] \qquad \text{and} \qquad \mathbf{b} = [2, -5, \tfrac{1}{3}].$$

Then $-\mathbf{a} = [-4, 0, -1]$, $\quad 7\mathbf{a} = [28, 0, 7]$, $\quad \mathbf{a} + \mathbf{b} = [6, -5, \tfrac{4}{3}]$, and

$$2(\mathbf{a} - \mathbf{b}) = 2[2, 5, \tfrac{2}{3}] = [4, 10, \tfrac{4}{3}] = 2\mathbf{a} - 2\mathbf{b}.$$ ■

Unit Vectors i, j, k. Besides $\mathbf{a} = [a_1, a_2, a_3]$ another popular way of writing vectors is

$$\text{(8)} \qquad \mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}.$$

In this representation, **i**, **j**, **k** are the unit vectors in the positive directions of the axes of a Cartesian coordinate system (Fig. 177). Hence, in components,

$$\text{(9)} \qquad \mathbf{i} = [1, 0, 0], \qquad \mathbf{j} = [0, 1, 0], \qquad \mathbf{k} = [0, 0, 1]$$

and the right side of (8) is a sum of three vectors parallel to the three axes.

EXAMPLE 3 ijk Notation for Vectors

In Example 2 we have $\mathbf{a} = 4\mathbf{i} + \mathbf{k}$, $\mathbf{b} = 2\mathbf{i} - 5\mathbf{j} + \tfrac{1}{3}\mathbf{k}$, and so on. ■

All the vectors $\mathbf{a} = [a_1, a_2, a_3] = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ (with real numbers as components) form the **real vector space** R^3 with the two *algebraic operations* of vector addition and scalar multiplication as just defined. R^3 has **dimension** 3. The triple of vectors **i**, **j**, **k** is called a **standard basis** of R^3. Given a Cartesian coordinate system, the representation (8) of a given vector is unique.

Vector space R^3 is a model of a general vector space, as discussed in Sec. 7.9, but is not needed in this chapter.

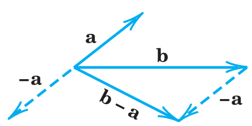


Fig. 176. Difference of vectors

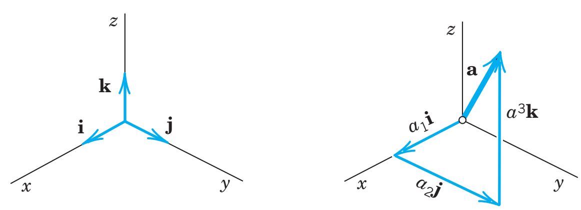


Fig. 177. The unit vectors **i**, **j**, **k** and the representation (8)

PROBLEM SET 9.1

1–5 COMPONENTS AND LENGTH

Find the components of the vector $\mathbf{v}$ with initial point P and terminal point Q. Find $|\mathbf{v}|$. Sketch $|\mathbf{v}|$. Find the unit vector $\mathbf{u}$ in the direction of $\mathbf{v}$.

1. P: (1, 1, 0), Q: (6, 2, 0)
2. P: (1, 1, 1), Q: (2, 2, 0)
3. P: (−3.0, 4,0, −0.5), Q: (5.5, 0, 1.2)
4. P: (1, 4, 2), Q: (−1, −4, −2)
5. P: (0, 0, 0), Q: (2, 1, −2)

6–10

Find the terminal point Q of the vector $\mathbf{v}$ with components as given and initial point P. Find $|\mathbf{v}|$.

6. 4, 0, 0; P: (0, 2, 13)
7. $\frac{1}{2}, 3, -\frac{1}{4}$; P: $(\frac{7}{2}, -3, \frac{3}{4})$
8. 13.1, 0.8, −2.0; P: (0, 0, 0)
9. 6, 1, −4; P: (−6, −1, −4)
10. 0, −3, 3; P: (0, 3, −3)

11–18 ADDITION, SCALAR MULTIPLICATION

Let $\mathbf{a} = [3, 2, 0] = 3\mathbf{i} + 2\mathbf{j}$; $\mathbf{b} = [-4, 6, 0] = 4\mathbf{i} + 6\mathbf{j}$, $\mathbf{c} = [5, -1, 8] = 5\mathbf{i} - \mathbf{j} + 8\mathbf{k}$, $\mathbf{d} = [0, 0, 4] = 4\mathbf{k}$.
Find:

11. $2\mathbf{a}$, $\frac{1}{2}\mathbf{a}$, $-\mathbf{a}$
12. $(\mathbf{a} + \mathbf{b}) + \mathbf{c}$, $\mathbf{a} + (\mathbf{b} + \mathbf{c})$
13. $\mathbf{b} + \mathbf{c}$, $\mathbf{c} + \mathbf{b}$
14. $3\mathbf{c} - 6\mathbf{d}$, $3(\mathbf{c} - 2\mathbf{d})$
15. $7(\mathbf{c} - \mathbf{b})$, $7\mathbf{c} - 7\mathbf{b}$
16. $\frac{9}{2}\mathbf{a} - 3\mathbf{c}$, $9(\frac{1}{2}\mathbf{a} - \frac{1}{3}\mathbf{c})$
17. $(7 - 3)\mathbf{a}$, $7\mathbf{a} - 3\mathbf{a}$
18. $4\mathbf{a} + 3\mathbf{b}$, $-4\mathbf{a} - 3\mathbf{b}$
19. What laws do Probs. 12–16 illustrate?
20. Prove Eqs. (4) and (6).

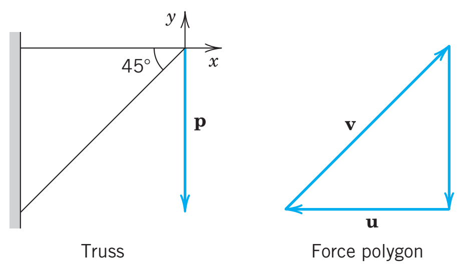
21–25 FORCES, RESULTANT

Find the resultant in terms of components and its magnitude.

21. $\mathbf{p} = [2, 3, 0]$, $\mathbf{q} = [0, 6, 1]$, $\mathbf{u} = [2, 0, -4]$
22. $\mathbf{p} = [1, -2, 3]$, $\mathbf{q} = [3, 21, -16]$, $\mathbf{u} = [-4, -19, 13]$
23. $\mathbf{u} = [8, -1, 0]$, $\mathbf{v} = [\frac{1}{2}, 0, \frac{4}{3}]$, $\mathbf{w} = [-\frac{17}{2}, 1, \frac{11}{3}]$
24. $\mathbf{p} = [-1, 2, -3]$, $\mathbf{q} = [1, 1, 1]$, $\mathbf{u} = [1, -2, 2]$
25. $\mathbf{u} = [3, 1, -6]$, $\mathbf{v} = [0, 2, 5]$, $\mathbf{w} = [3, -1, -13]$

26–37 FORCES, VELOCITIES

26. **Equilibrium.** Find $\mathbf{v}$ such that $\mathbf{p}$, $\mathbf{q}$, $\mathbf{u}$ in Prob. 21 and $\mathbf{v}$ are in equilibrium.
27. Find $\mathbf{p}$ such that $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ in Prob. 23 and $\mathbf{p}$ are in equilibrium.
28. **Unit vector.** Find the unit vector in the direction of the resultant in Prob. 24.
29. **Restricted resultant.** Find all $\mathbf{v}$ such that the resultant of $\mathbf{v}$, $\mathbf{p}$, $\mathbf{q}$, $\mathbf{u}$ with $\mathbf{p}$, $\mathbf{q}$, $\mathbf{u}$ as in Prob. 21 is parallel to the xy-plane.
30. Find $\mathbf{v}$ such that the resultant of $\mathbf{p}$, $\mathbf{q}$, $\mathbf{u}$, $\mathbf{v}$ with $\mathbf{p}$, $\mathbf{q}$, $\mathbf{u}$ as in Prob. 24 has no components in x- and y-directions.
31. For what k is the resultant of [2, 0, −7], [1, 2, −3], and [0, 3, k] parallel to the xy-plane?
32. If $|\mathbf{p}| = 6$ and $|\mathbf{q}| = 4$, what can you say about the magnitude and direction of the resultant? Can you think of an application to robotics?
33. Same question as in Prob. 32 if $|\mathbf{p}| = 9$, $|\mathbf{q}| = 6$, $|\mathbf{u}| = 3$.
34. **Relative velocity.** If airplanes A and B are moving southwest with speed $|\mathbf{v}_A| = 550$ mph, and northwest with speed $|\mathbf{v}_B| = 450$ mph, respectively, what is the relative velocity $\mathbf{v} = \mathbf{v}_B - \mathbf{v}_A$ of B with respect to A?
35. Same question as in Prob. 34 for two ships moving northeast with speed $|\mathbf{v}_A| = 22$ knots and west with speed $|\mathbf{v}_B| = 19$ knots.
36. **Reflection.** If a ray of light is reflected once in each of two mutually perpendicular mirrors, what can you say about the reflected ray?
37. **Force polygon. Truss.** Find the forces in the system of two rods (*truss*) in the figure, where $|\mathbf{p}| = 1000$ nt. *Hint.* Forces in equilibrium form a polygon, the *force polygon*.



Problem 37

38. TEAM PROJECT. Geometric Applications. To increase your skill in dealing with vectors, use vectors to prove the following (see the figures).

(a) The diagonals of a parallelogram bisect each other.

(b) The line through the midpoints of adjacent sides of a parallelogram bisects one of the diagonals in the ratio 1 : 3.

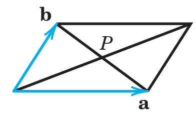
(c) Obtain (b) from (a).

(d) The three medians of a triangle (the segments from a vertex to the midpoint of the opposite side) meet at a single point, which divides the medians in the ratio 2 : 1.

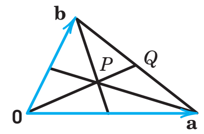
(e) The quadrilateral whose vertices are the midpoints of the sides of an arbitrary quadrilateral is a parallelogram.

(f) The four space diagonals of a parallelepiped meet and bisect each other.

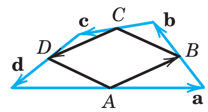
(g) The sum of the vectors drawn from the center of a regular polygon to its vertices is the zero vector.



Team Project 38(a)



Team Project 38(d)



Team Project 38(e)

9.2 Inner Product (Dot Product)

Orthogonality

The inner product or dot product can be motivated by calculating work done by a constant force, determining components of forces, or other applications. It involves the length of vectors and the angle between them. The inner product is a kind of multiplication of two vectors, defined in such a way that the outcome is a scalar. Indeed, another term for inner product is scalar product, a term we shall not use here. The definition of the inner product is as follows.

DEFINITION

Inner Product (Dot Product) of Vectors

The **inner product** or **dot product** $\mathbf{a} \cdot \mathbf{b}$ (read "$\mathbf{a}$ dot $\mathbf{b}$") of two vectors $\mathbf{a}$ and $\mathbf{b}$ is the product of their lengths times the cosine of their angle (see Fig. 178),

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos \gamma \qquad \text{if} \quad \mathbf{a} \neq \mathbf{0}, \mathbf{b} \neq \mathbf{0}$$
$$\mathbf{a} \cdot \mathbf{b} = 0 \qquad \text{if} \quad \mathbf{a} = \mathbf{0} \text{ or } \mathbf{b} = \mathbf{0}. \tag{1}$$

The angle γ, $0 \leqq \gamma \leqq \pi$, between $\mathbf{a}$ and $\mathbf{b}$ is measured when the initial points of the vectors coincide, as in Fig. 178. In components, $\mathbf{a} = [a_1, a_2, a_3]$, $\mathbf{b} = [b_1, b_2, b_3]$, and

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + a_3 b_3. \tag{2}$$

38. TEAM PROJECT. Geometric Applications. To increase your skill in dealing with vectors, use vectors to prove the following (see the figures).

(a) The diagonals of a parallelogram bisect each other.

(b) The line through the midpoints of adjacent sides of a parallelogram bisects one of the diagonals in the ratio 1 : 3.

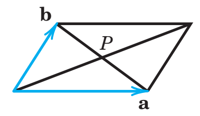
(c) Obtain (b) from (a).

(d) The three medians of a triangle (the segments from a vertex to the midpoint of the opposite side) meet at a single point, which divides the medians in the ratio 2 : 1.

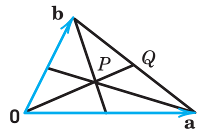
(e) The quadrilateral whose vertices are the midpoints of the sides of an arbitrary quadrilateral is a parallelogram.

(f) The four space diagonals of a parallelepiped meet and bisect each other.

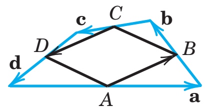
(g) The sum of the vectors drawn from the center of a regular polygon to its vertices is the zero vector.



Team Project 38(a)



Team Project 38(d)



Team Project 38(e)

9.2 Inner Product (Dot Product)

Orthogonality

The inner product or dot product can be motivated by calculating work done by a constant force, determining components of forces, or other applications. It involves the length of vectors and the angle between them. The inner product is a kind of multiplication of two vectors, defined in such a way that the outcome is a scalar. Indeed, another term for inner product is scalar product, a term we shall not use here. The definition of the inner product is as follows.

DEFINITION

Inner Product (Dot Product) of Vectors

The **inner product** or **dot product** $\mathbf{a} \cdot \mathbf{b}$ (read "$\mathbf{a}$ dot $\mathbf{b}$") of two vectors $\mathbf{a}$ and $\mathbf{b}$ is the product of their lengths times the cosine of their angle (see Fig. 178),

$$\text{(1)} \qquad \begin{aligned} \mathbf{a} \cdot \mathbf{b} &= |\mathbf{a}||\mathbf{b}| \cos \gamma \qquad &&\text{if} \quad \mathbf{a} \neq \mathbf{0}, \mathbf{b} \neq \mathbf{0} \\ \mathbf{a} \cdot \mathbf{b} &= 0 &&\text{if} \quad \mathbf{a} = \mathbf{0} \text{ or } \mathbf{b} = \mathbf{0}. \end{aligned}$$

The angle γ, $0 \leqq \gamma \leqq \pi$, between $\mathbf{a}$ and $\mathbf{b}$ is measured when the initial points of the vectors coincide, as in Fig. 178. In components, $\mathbf{a} = [a_1, a_2, a_3]$, $\mathbf{b} = [b_1, b_2, b_3]$, and

$$\text{(2)} \qquad \mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + a_3 b_3.$$

The second line in (1) is needed because γ is undefined when $\mathbf{a} = \mathbf{0}$ or $\mathbf{b} = \mathbf{0}$. The derivation of (2) from (1) is shown below.

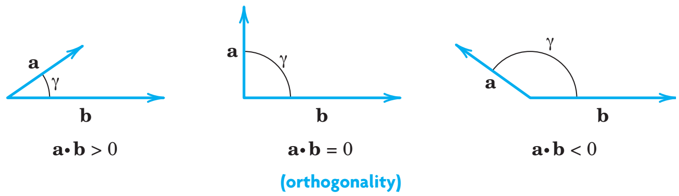


Fig. 178. Angle between vectors and value of inner product

Orthogonality. Since the cosine in (1) may be positive, 0, or negative, so may be the inner product (Fig. 178). The case that the inner product is zero is of particular practical interest and suggests the following concept.

A vector $\mathbf{a}$ is called **orthogonal** to a vector $\mathbf{b}$ if $\mathbf{a} \cdot \mathbf{b} = 0$. Then $\mathbf{b}$ is also orthogonal to $\mathbf{a}$, and we call $\mathbf{a}$ and $\mathbf{b}$ **orthogonal vectors**. Clearly, this happens for nonzero vectors if and only if $\cos \gamma = 0$; thus $\gamma = \pi/2$ (90°). This proves the important

THEOREM 1

Orthogonality Criterion

The inner product of two nonzero vectors is 0 *if and only if these vectors are perpendicular.*

Length and Angle. Equation (1) with $\mathbf{b} = \mathbf{a}$ gives $\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2$. Hence

$$|\mathbf{a}| = \sqrt{\mathbf{a} \cdot \mathbf{a}}. \tag{3}$$

From (3) and (1) we obtain for the angle γ between two nonzero vectors

$$\cos \gamma = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}||\mathbf{b}|} = \frac{\mathbf{a} \cdot \mathbf{b}}{\sqrt{\mathbf{a} \cdot \mathbf{a}}\sqrt{\mathbf{b} \cdot \mathbf{b}}}. \tag{4}$$

EXAMPLE 1

Inner Product. Angle Between Vectors

Find the inner product and the lengths of $\mathbf{a} = [1, 2, 0]$ and $\mathbf{b} = [3, -2, 1]$ as well as the angle between these vectors.

Solution. $\mathbf{a} \cdot \mathbf{b} = 1 \cdot 3 + 2 \cdot (-2) + 0 \cdot 1 = -1$, $|\mathbf{a}| = \sqrt{\mathbf{a} \cdot \mathbf{a}} = \sqrt{5}$, $|\mathbf{b}| = \sqrt{\mathbf{b} \cdot \mathbf{b}} = \sqrt{14}$, and (4) gives the angle

$$\gamma = \arccos \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}||\mathbf{b}|} = \arccos(-0.11952) = 1.69061 = 96.865°.$$ ■

From the definition we see that the inner product has the following properties. For any vectors **a**, **b**, **c** and scalars q_1, q_2,

$$
\textbf{(5)} \quad
\begin{array}{lll}
\text{(a)} & (q_1\mathbf{a} + q_2\mathbf{b}) \bullet \mathbf{c} = q_1\mathbf{a} \bullet \mathbf{c} + q_1\mathbf{b} \bullet \mathbf{c} & (\textit{Linearity}) \\
\text{(b)} & \mathbf{a} \bullet \mathbf{b} = \mathbf{b} \bullet \mathbf{a} & (\textit{Symmetry}) \\
\text{(c)} & \left.\begin{array}{c} \mathbf{a} \bullet \mathbf{a} \geqq 0 \\ \mathbf{a} \bullet \mathbf{a} = 0 \quad \text{if and only if} \quad \mathbf{a} = \mathbf{0} \end{array}\right\} & (\textit{Positive-definiteness}).
\end{array}
$$

Hence *dot multiplication is commutative* as shown by (5b). Furthermore, it is *distributive with respect to vector addition*. This follows from (5a) with $q_1 = 1$ and $q_2 = 1$:

$$\textbf{(5a*)} \qquad (\mathbf{a} + \mathbf{b}) \bullet \mathbf{c} = \mathbf{a} \bullet \mathbf{c} + \mathbf{b} \bullet \mathbf{c} \qquad (\textit{Distributivity}).$$

Furthermore, from (1) and $|\cos \gamma| \leqq 1$ we see that

$$\textbf{(6)} \qquad |\mathbf{a} \bullet \mathbf{b}| \leqq |\mathbf{a}||\mathbf{b}| \qquad (\textit{Cauchy–Schwarz inequality}).$$

Using this and (3), you may prove (see Prob. 16)

$$\textbf{(7)} \qquad |\mathbf{a} + \mathbf{b}| \leqq |\mathbf{a}| + |\mathbf{b}| \qquad (\textbf{Triangle inequality}).$$

Geometrically, (7) with $<$ says that one side of a triangle must be shorter than the other two sides together; this motivates the name of (7).

A simple direct calculation with inner products shows that

$$(8) \qquad |\mathbf{a} + \mathbf{b}|^2 + |\mathbf{a} - \mathbf{b}|^2 = 2(|\mathbf{a}|^2 + |\mathbf{b}|^2) \quad (\textit{Parallelogram equality}).$$

Equations (6)–(8) play a basic role in so-called *Hilbert spaces*, which are abstract inner product spaces. Hilbert spaces form the basis of quantum mechanics, for details see [GenRef7] listed in App. 1.

Derivation of (2) from (1). We write $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ and $\mathbf{b} = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$, as in (8) of Sec. 9.1. If we substitute this into $\mathbf{a} \bullet \mathbf{b}$ and use (5a*), we first have a sum of $3 \times 3 = 9$ products

$$\mathbf{a} \bullet \mathbf{b} = a_1b_1\mathbf{i} \bullet \mathbf{i} + a_1b_2\mathbf{i} \bullet \mathbf{j} + \cdots + a_3b_3\mathbf{k} \bullet \mathbf{k}.$$

Now **i**, **j**, **k** are unit vectors, so that $\mathbf{i} \bullet \mathbf{i} = \mathbf{j} \bullet \mathbf{j} = \mathbf{k} \bullet \mathbf{k} = 1$ by (3). Since the coordinate axes are perpendicular, so are **i**, **j**, **k**, and Theorem 1 implies that the other six of those nine products are 0, namely, $\mathbf{i} \bullet \mathbf{j} = \mathbf{j} \bullet \mathbf{i} = \mathbf{j} \bullet \mathbf{k} = \mathbf{k} \bullet \mathbf{j} = \mathbf{k} \bullet \mathbf{i} = \mathbf{i} \bullet \mathbf{k} = 0$. But this reduces our sum for $\mathbf{a} \bullet \mathbf{b}$ to (2). ■

Applications of Inner Products

Typical applications of inner products are shown in the following examples and in Problem Set 9.2.

EXAMPLE 2 Work Done by a Force Expressed as an Inner Product

This is a major application. It concerns a body on which a *constant* force $\mathbf{p}$ acts. (For a *variable* force, see Sec. 10.1.) Let the body be given a displacement $\mathbf{d}$. Then the work done by $\mathbf{p}$ in the displacement is defined as

$$W = |\mathbf{p}||\mathbf{d}| \cos \alpha = \mathbf{p} \cdot \mathbf{d}, \tag{9}$$

that is, magnitude $|\mathbf{p}|$ of the force times length $|\mathbf{d}|$ of the displacement times the cosine of the angle α between $\mathbf{p}$ and $\mathbf{d}$ (Fig. 179). If $\alpha < 90°$, as in Fig. 179, then $W > 0$. If $\mathbf{p}$ and $\mathbf{d}$ are orthogonal, then the work is zero (why?). If $\alpha > 90°$, then $W < 0$, which means that in the displacement one has to do work against the force. For example, think of swimming across a river at some angle α against the current.

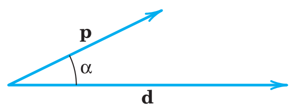


Fig. 179. Work done by a force

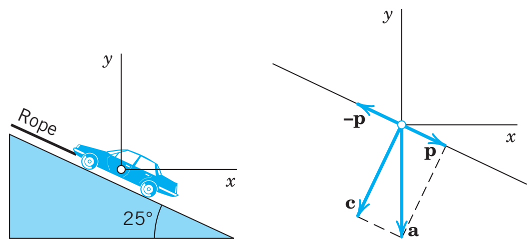


Fig. 180. Example 3

EXAMPLE 3 Component of a Force in a Given Direction

What force in the rope in Fig. 180 will hold a car of 5000 lb in equilibrium if the ramp makes an angle of 25° with the horizontal?

Solution. Introducing coordinates as shown, the weight is $\mathbf{a} = [0, -5000]$ because this force points downward, in the negative y-direction. We have to represent $\mathbf{a}$ as a sum (resultant) of two forces, $\mathbf{a} = \mathbf{c} + \mathbf{p}$, where $\mathbf{c}$ is the force the car exerts on the ramp, which is of no interest to us, and $\mathbf{p}$ is parallel to the rope. A vector in the direction of the rope is (see Fig. 180)

$$\mathbf{b} = [-1, \tan 25°] = [-1, 0.46631], \qquad \text{thus} \qquad |\mathbf{b}| = 1.10338,$$

The direction of the unit vector $\mathbf{u}$ is opposite to the direction of the rope so that

$$\mathbf{u} = -\frac{1}{|\mathbf{b}|}\mathbf{b} = [0.90631, -0.42262].$$

Since $|\mathbf{u}| = 1$ and $\cos \gamma > 0$, we see that we can write our result as

$$|\mathbf{p}| = (|\mathbf{a}| \cos \gamma)|\mathbf{u}| = \mathbf{a} \cdot \mathbf{u} = -\frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{b}|} = \frac{5000 \cdot 0.46631}{1.10338} = 2113 \text{ [lb]}.$$

We can also note that $\gamma = 90° - 25° = 65°$ is the angle between $\mathbf{a}$ and $\mathbf{p}$ so that

$$|\mathbf{p}| = |\mathbf{a}| \cos \gamma = 5000 \cos 65° = 2113 \text{ [lb]}.$$

Answer: About 2100 lb. ■

Example 3 is typical of applications that deal with the **component** or **projection** *of a vector* **a** *in the direction of a vector* **b** $(\neq \mathbf{0})$. If we denote by p the length of the orthogonal projection of **a** on a straight line l parallel to **b** as shown in Fig. 181, then

$$p = |\mathbf{a}| \cos \gamma. \tag{10}$$

Here p is taken with the plus sign if $p\mathbf{b}$ has the direction of **b** and with the minus sign if $p\mathbf{b}$ has the direction opposite to **b**.

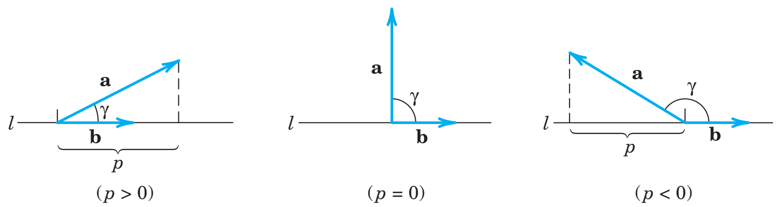


Fig. 181. Component of a vector **a** in the direction of a vector **b**

Multiplying (10) by $|\mathbf{b}|/|\mathbf{b}| = 1$, we have $\mathbf{a} \cdot \mathbf{b}$ in the numerator and thus

$$p = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{b}|} \qquad (\mathbf{b} \neq \mathbf{0}). \tag{11}$$

If **b** is a unit vector, as it is often used for fixing a direction, then (11) simply gives

$$p = \mathbf{a} \cdot \mathbf{b} \qquad (|\mathbf{b}| = 1). \tag{12}$$

Figure 182 shows the projection p of **a** in the direction of **b** (as in Fig. 181) and the projection $q = |\mathbf{b}| \cos \gamma$ of **b** in the direction of **a**.

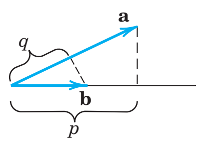


Fig. 182. Projections p of **a** on **b** and q of **b** on **a**

EXAMPLE 4 **Orthonormal Basis**

By definition, an *orthonormal basis* for 3-space is a basis $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ consisting of orthogonal unit vectors. It has the great advantage that the determination of the coefficients in representations $\mathbf{v} = l_1\mathbf{a} + l_2\mathbf{b} + l_3\mathbf{c}$ of a given vector **v** is very simple. We claim that $l_1 = \mathbf{a} \cdot \mathbf{v}$, $l_2 = \mathbf{b} \cdot \mathbf{v}$, $l_3 = \mathbf{c} \cdot \mathbf{v}$. Indeed, this follows simply by taking the inner products of the representation with **a**, **b**, **c**, respectively, and using the orthonormality of the basis, $\mathbf{a} \cdot \mathbf{v} = l_1\mathbf{a} \cdot \mathbf{a} + l_2\mathbf{a} \cdot \mathbf{b} + l_3\mathbf{a} \cdot \mathbf{c} = l_1$, etc.

For example, the unit vectors **i**, **j**, **k** in (8), Sec. 9.1, associated with a Cartesian coordinate system form an orthonormal basis, called the **standard basis** with respect to the given coordinate system. ■

EXAMPLE 5 **Orthogonal Straight Lines in the Plane**

Find the straight line L_1 through the point P: (1, 3) in the xy-plane and perpendicular to the straight line $L_2 : x - 2y + 2 = 0$; see Fig. 183.

Solution. The idea is to write a general straight line $L_1 : a_1x + a_2y = c$ as $\mathbf{a} \bullet \mathbf{r} = c$ with $\mathbf{a} = [a_1, a_2] \neq \mathbf{0}$ and $\mathbf{r} = [x, y]$, according to (2). Now the line L_1^* through the origin and parallel to L_1 is $\mathbf{a} \bullet \mathbf{r} = 0$. Hence, by Theorem 1, the vector $\mathbf{a}$ is perpendicular to $\mathbf{r}$. Hence it is perpendicular to L_1^* and also to L_1 because L_1 and L_1^* are parallel. $\mathbf{a}$ is called a **normal vector** of L_1 (and of L_1^*).

Now a normal vector of the given line $x - 2y + 2 = 0$ is $\mathbf{b} = [1, -2]$. Thus L_1 is perpendicular to L_2 if $\mathbf{b} \bullet \mathbf{a} = a_1 - 2a_2 = 0$, for instance, if $\mathbf{a} = [2, 1]$. Hence L_1 is given by $2x + y = c$. It passes through P: (1, 3) when $2 \cdot 1 + 3 = c = 5$. *Answer:* $y = -2x + 5$. Show that the point of intersection is $(x, y) = (1.6, 1.8)$. ■

EXAMPLE 6 **Normal Vector to a Plane**

Find a unit vector perpendicular to the plane $4x + 2y + 4z = -7$.

Solution. Using (2), we may write any plane in space as

$$\mathbf{a} \bullet \mathbf{r} = a_1x + a_2y + a_3z = c \tag{13}$$

where $\mathbf{a} = [a_1, a_2, a_3] \neq \mathbf{0}$ and $\mathbf{r} = [x, y, z]$. The unit vector in the direction of $\mathbf{a}$ is (Fig. 184)

$$\mathbf{n} = \frac{1}{|\mathbf{a}|}\mathbf{a}.$$

Dividing by $|\mathbf{a}|$, we obtain from (13)

$$\mathbf{n} \bullet \mathbf{r} = p \qquad \text{where} \qquad p = \frac{c}{|\mathbf{a}|}. \tag{14}$$

From (12) we see that p is the projection of $\mathbf{r}$ in the direction of $\mathbf{n}$. This projection has the same constant value $c/|\mathbf{a}|$ for the position vector $\mathbf{r}$ of any point in the plane. Clearly this holds if and only if $\mathbf{n}$ is perpendicular to the plane. $\mathbf{n}$ is called a **unit normal vector** of the plane (the other being $-\mathbf{n}$).

Furthermore, from this and the definition of projection, it follows that $|p|$ is the distance of the plane from the origin. Representation (14) is called **Hesse's**[2] **normal form** of a plane. In our case, $\mathbf{a} = [4, 2, 4]$, $c = -7$, $|\mathbf{a}| = 6$, $\mathbf{n} = \frac{1}{6}\mathbf{a} = [\frac{2}{3}, \frac{1}{3}, \frac{2}{3}]$, and the plane has the distance $\frac{7}{6}$ from the origin. ■

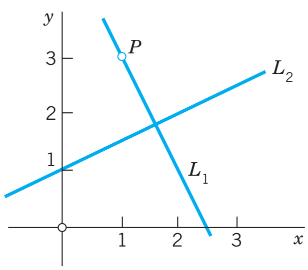


Fig. 183. Example 5

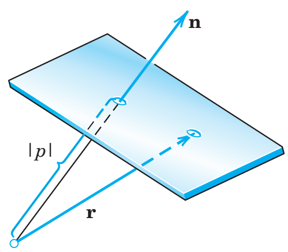


Fig. 184. Normal vector to a plane

[2]LUDWIG OTTO HESSE (1811–1874), German mathematician who contributed to the theory of curves and surfaces.

PROBLEM SET 9.2

1–10 INNER PRODUCT

Let $\mathbf{a} = [1, -3, 5]$, $\mathbf{b} = [4, 0, 8]$, $\mathbf{c} = [-2, 9, 1]$. Find:

1. $\mathbf{a} \cdot \mathbf{b}$, $\mathbf{b} \cdot \mathbf{a}$, $\mathbf{b} \cdot \mathbf{c}$

2. $(-3\mathbf{a} + 5\mathbf{c}) \cdot \mathbf{b}$, $15(\mathbf{a} - \mathbf{c}) \cdot \mathbf{b}$

3. $|\mathbf{a}|$, $|2\mathbf{b}|$, $|-\mathbf{c}|$

4. $|\mathbf{a} + \mathbf{b}|$, $|\mathbf{a}| + |\mathbf{b}|$

5. $|\mathbf{b} + \mathbf{c}|$, $|\mathbf{b}| + |\mathbf{c}|$

6. $|\mathbf{a} + \mathbf{c}|^2 + |\mathbf{a} - \mathbf{c}|^2 - 2(|\mathbf{a}|^2 + |\mathbf{c}|^2)$

7. $|\mathbf{a} \cdot \mathbf{c}|$, $|\mathbf{a}||\mathbf{c}|$

8. $5\mathbf{a} \cdot 13\mathbf{b}$, $65\mathbf{a} \cdot \mathbf{b}$

9. $15\mathbf{a} \cdot \mathbf{b} + 15\mathbf{a} \cdot \mathbf{c}$, $15\mathbf{a} \cdot (\mathbf{b} + \mathbf{c})$

10. $\mathbf{a} \cdot (\mathbf{b} - \mathbf{c})$, $(\mathbf{a} - \mathbf{b}) \cdot \mathbf{c}$

11–16 GENERAL PROBLEMS

11. What laws do Probs. 1 and 4–7 illustrate?

12. What does $\mathbf{u} \cdot \mathbf{v} = \mathbf{u} \cdot \mathbf{w}$ imply if $\mathbf{u} = \mathbf{0}$? If $\mathbf{u} \neq \mathbf{0}$?

13. Prove the Cauchy–Schwarz inequality.

14. Verify the Cauchy–Schwarz and triangle inequalities for the above **a** and **b**.

15. Prove the parallelogram equality. Explain its name.

16. Triangle inequality. Prove Eq. (7). *Hint.* Use Eq. (3) for $|\mathbf{a} + \mathbf{b}|$ and Eq. (6) to prove the square of Eq. (7), then take roots.

17–20 WORK

Find the work done by a force **p** acting on a body if the body is displaced along the straight segment $\overline{AB}$ from A to B. Sketch $\overline{AB}$ and **p**. Show the details.

17. $\mathbf{p} = [2, 5, 0]$, A: (1, 3, 3), B: (3, 5, 5)

18. $\mathbf{p} = [-1, -2, 4]$, A: (0, 0, 0), B: (6, 7, 5)

19. $\mathbf{p} = [0, 4, 3]$, A: (4, 5, −1), B: (1, 3, 0)

20. $\mathbf{p} = [6, -3, -3]$, A: (1, 5, 2), B: (3, 4, 1)

21. Resultant. Is the work done by the resultant of two forces in a displacement the sum of the work done by each of the forces separately? Give proof or counterexample.

22–30 ANGLE BETWEEN VECTORS

Let $\mathbf{a} = [1, 1, 0]$, $\mathbf{b} = [3, 2, 1]$, and $\mathbf{c} = [1, 0, 2]$. Find the angle between:

22. **a**, **b**

23. **b**, **c**

24. $\mathbf{a} + \mathbf{c}$, $\mathbf{b} + \mathbf{c}$

25. What will happen to the angle in Prob. 24 if we replace **c** by $n\mathbf{c}$ with larger and larger n?

26. Cosine law. Deduce the law of cosines by using vectors **a**, **b**, and $\mathbf{a} - \mathbf{b}$.

27. Addition law. $\cos(\alpha - \beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta$. Obtain this by using $\mathbf{a} = [\cos\alpha, \sin\alpha]$, $\mathbf{b} = [\cos\beta, \sin\beta]$ where $0 \leqq \alpha \leqq \beta \leqq 2\pi$.

28. Triangle. Find the angles of the triangle with vertices A: (0, 0, 2), B: (3, 0, 2), and C: (1, 1, 1). Sketch the triangle.

29. Parallelogram. Find the angles if the vertices are (0, 0), (6, 0), (8, 3), and (2, 3).

30. Distance. Find the distance of the point A: (1, 0, 2) from the plane P: $3x + y + z = 9$. Make a sketch.

31–35 ORTHOGONALITY

ORTHOGONALITY is particularly important, mainly because of orthogonal coordinates, such as ***Cartesian coordinates***, whose ***natural basis*** [Eq. (9), Sec. 9.1], consists of three orthogonal unit vectors.

31. For what values of a_1 are $[a_1, 4, 3]$ and $[3, -2, 12]$ orthogonal?

32. Planes. For what c are $3x + z = 5$ and $8x - y + cz = 9$ orthogonal?

33. Unit vectors. Find all unit vectors $\mathbf{a} = [a_1, a_2]$ in the plane orthogonal to [4, 3].

34. Corner reflector. Find the angle between a light ray and its reflection in three orthogonal plane mirrors, known as *corner reflector*.

35. Parallelogram. When will the diagonals be orthogonal? Give a proof.

36–40 COMPONENT IN THE DIRECTION OF A VECTOR

Find the component of **a** in the direction of **b**. Make a sketch.

36. $\mathbf{a} = [1, 1, 1]$, $\mathbf{b} = [2, 1, 3]$

37. $\mathbf{a} = [3, 4, 0]$, $\mathbf{b} = [4, -3, 2]$

38. $\mathbf{a} = [8, 2, 0]$, $\mathbf{b} = [-4, -1, 0]$

39. When will the component (the projection) of **a** in the direction of **b** be equal to the component (the projection) of **b** in the direction of **a**? First guess.

40. What happens to the component of **a** in the direction of **b** if you change the length of **b**?

9.3 Vector Product (Cross Product)

We shall define another form of multiplication of vectors, inspired by applications, whose result will be a *vector*. This is in contrast to the dot product of Sec. 9.2 where multiplication resulted in a *scalar*. We can construct a vector $\mathbf{v}$ that is perpendicular to two vectors $\mathbf{a}$ and $\mathbf{b}$, which are two sides of a parallelogram on a plane in space as indicated in Fig. 185, such that the length $|\mathbf{v}|$ is numerically equal to the area of that parallelogram. Here then is the new concept.

DEFINITION

Vector Product (Cross Product, Outer Product) of Vectors

The **vector product** or **cross product** $\mathbf{a} \times \mathbf{b}$ (read "$\mathbf{a}$ cross $\mathbf{b}$") of two vectors $\mathbf{a}$ and $\mathbf{b}$ is the vector $\mathbf{v}$ denoted by

$$\mathbf{v} = \mathbf{a} \times \mathbf{b}$$

I. If $\mathbf{a} = \mathbf{0}$ or $\mathbf{b} = \mathbf{0}$, then we define $\mathbf{v} = \mathbf{a} \times \mathbf{b} = \mathbf{0}$.

II. If both vectors are nonzero vectors, then vector $\mathbf{v}$ has the length

$$(1) \qquad |\mathbf{v}| = |\mathbf{a} \times \mathbf{b}| = |\mathbf{a}||\mathbf{b}| \sin \gamma,$$

where γ is the angle between $\mathbf{a}$ and $\mathbf{b}$ as in Sec. 9.2.

Furthermore, by design, $\mathbf{a}$ and $\mathbf{b}$ form the sides of a parallelogram on a plane in space. The parallelogram is shaded in blue in Fig. 185. The area of this blue parallelogram is precisely given by Eq. (1), so that the length $|\mathbf{v}|$ of the vector $\mathbf{v}$ is equal to the area of that parallelogram.

III. If $\mathbf{a}$ and $\mathbf{b}$ lie in the same straight line, i.e., $\mathbf{a}$ and $\mathbf{b}$ have the same or opposite directions, then γ is $0°$ or $180°$ so that $\sin \gamma = 0$. In that case $|\mathbf{v}| = 0$ so that $\mathbf{v} = \mathbf{a} \times \mathbf{b} = \mathbf{0}$.

IV. If cases I and III do not occur, then $\mathbf{v}$ is a nonzero vector. The direction of $\mathbf{v} = \mathbf{a} \times \mathbf{b}$ is perpendicular to both $\mathbf{a}$ and $\mathbf{b}$ such that $\mathbf{a}$, $\mathbf{b}$, $\mathbf{v}$—precisely in this order (!)—form a right-handed triple as shown in Figs. 185–187 and explained below.

Another term for vector product is outer product.

Remark. Note that I and III completely characterize the exceptional case when the cross product is equal to the zero vector, and II and IV the regular case where the cross product is perpendicular to two vectors.

Just as we did with the dot product, we would also like to express the cross product in components. Let $\mathbf{a} = [a_1, a_2, a_3]$ and $\mathbf{b} = [b_1, b_2, b_3]$. Then $\mathbf{v} = [v_1, v_2, v_3] = \mathbf{a} \times \mathbf{b}$ has the components

$$\textbf{(2)} \qquad v_1 = a_2b_3 - a_3b_2, \qquad v_2 = a_3b_1 - a_1b_3, \qquad v_3 = a_1b_2 - a_2b_1.$$

Here the Cartesian coordinate system is *right-handed*, as explained below (see also Fig. 188). (For a left-handed system, each component of $\mathbf{v}$ must be multiplied by -1. Derivation of (2) in App. 4.)

Right-Handed Triple. A triple of vectors **a**, **b**, **v** is *right-handed* if the vectors in the given order assume the same sort of orientation as the thumb, index finger, and middle finger of the right hand when these are held as in Fig. 186. We may also say that if **a** is rotated into the direction of **b** through the angle $\gamma\,(<\pi)$, then **v** advances in the same direction as a right-handed screw would if turned in the same way (Fig. 187).

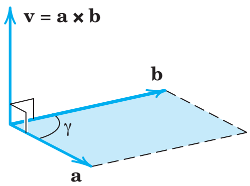


Fig. 185. Vector product

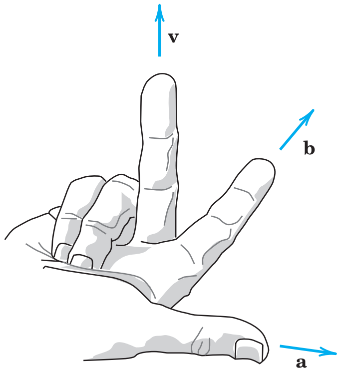


Fig. 186. Right-handed triple of vectors **a, b, v**

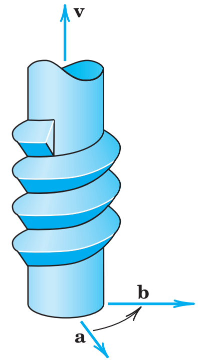
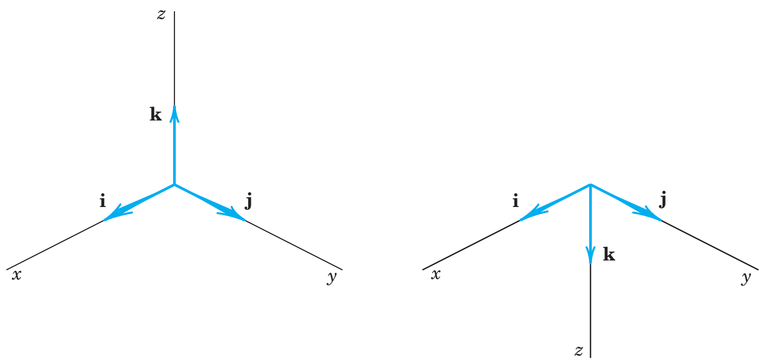


Fig. 187. Right-handed screw

Right-Handed Cartesian Coordinate System. The system is called **right-handed** if the corresponding unit vectors **i**, **j**, **k** in the positive directions of the axes (see Sec. 9.1) form a right-handed triple as in Fig. 188a. The system is called **left-handed** if the sense of **k** is reversed, as in Fig. 188b. In applications, we prefer right-handed systems.



(a) Right-handed (b) Left-handed

Fig. 188. The two types of Cartesian coordinate systems

How to Memorize (2). If you know second- and third-order determinants, you see that (2) can be written

$$(2^*)\qquad v_1 = \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}, \qquad v_2 = -\begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} = +\begin{vmatrix} a_3 & a_1 \\ b_3 & b_1 \end{vmatrix}, \qquad v_3 = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$$

and $\mathbf{v} = [v_1, v_2, v_3] = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$ is the expansion of the following symbolic determinant by its first row. (We call the determinant "symbolic" because the first row consists of vectors rather than of numbers.)

$$(2^{**}) \qquad \mathbf{v} = \mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}\mathbf{i} - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}\mathbf{j} + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}\mathbf{k}.$$

For a left-handed system the determinant has a minus sign in front.

EXAMPLE 1 Vector Product

For the vector product $\mathbf{v} = \mathbf{a} \times \mathbf{b}$ of $\mathbf{a} = [1, 1, 0]$ and $\mathbf{b} = [3, 0, 0]$ in right-handed coordinates we obtain from (2)

$$v_1 = 0, \qquad v_2 = 0, \qquad v_3 = 1 \cdot 0 - 1 \cdot 3 = -3.$$

We confirm this by (2**):

$$\mathbf{v} = \mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 1 & 0 \\ 3 & 0 & 0 \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & 0 \end{vmatrix}\mathbf{i} - \begin{vmatrix} 1 & 0 \\ 3 & 0 \end{vmatrix}\mathbf{j} + \begin{vmatrix} 1 & 1 \\ 3 & 0 \end{vmatrix}\mathbf{k} = -3\mathbf{k} = [0, 0, -3].$$

To check the result in this simple case, sketch **a**, **b**, and **v**. Can you see that two vectors in the xy-plane must always have their vector product parallel to the z-axis (or equal to the zero vector)? ■

EXAMPLE 2 Vector Products of the Standard Basis Vectors

$$(3) \qquad \begin{aligned} \mathbf{i} \times \mathbf{j} &= \mathbf{k}, & \mathbf{j} \times \mathbf{k} &= \mathbf{i}, & \mathbf{k} \times \mathbf{i} &= \mathbf{j} \\ \mathbf{j} \times \mathbf{i} &= -\mathbf{k}, & \mathbf{k} \times \mathbf{j} &= -\mathbf{i}, & \mathbf{i} \times \mathbf{k} &= -\mathbf{j}. \end{aligned}$$

We shall use this in the next proof. ■

THEOREM 1 General Properties of Vector Products

(a) *For every scalar l,*

$$(4) \qquad (l\mathbf{a}) \times \mathbf{b} = l(\mathbf{a} \times \mathbf{b}) = \mathbf{a} \times (l\mathbf{b}).$$

(b) *Cross multiplication is distributive with respect to vector addition; that is,*

$$(5) \qquad \begin{aligned} &(\alpha) \quad \mathbf{a} \times (\mathbf{b} + \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) + (\mathbf{a} \times \mathbf{c}), \\ &(\beta) \quad (\mathbf{a} + \mathbf{b}) \times \mathbf{c} = (\mathbf{a} \times \mathbf{c}) + (\mathbf{b} \times \mathbf{c}). \end{aligned}$$

(c) *Cross multiplication is* ***not commutative*** *but* ***anticommutative;*** *that is,*

$$(6) \qquad \mathbf{b} \times \mathbf{a} = -(\mathbf{a} \times \mathbf{b}) \qquad \text{(Fig. 189).}$$

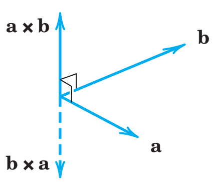


Fig. 189. Anticommutativity of cross multiplication

(d) *Cross multiplication is* ***not associative***; *that is, in general,*

(7) $$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) \neq (\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$$

so that the parentheses cannot be omitted.

PROOF Equation (4) follows directly from the definition. In (5α), formula (2*) gives for the first component on the left

$$\begin{vmatrix} a_2 & a_3 \\ b_2 + c_2 & b_3 + c_3 \end{vmatrix} = a_2(b_3 + c_3) - a_3(b_2 + c_2)$$

$$= (a_2b_3 - a_3b_2) + (a_2c_3 - a_3c_2)$$

$$= \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} + \begin{vmatrix} a_2 & a_3 \\ c_2 & c_3 \end{vmatrix}.$$

By (2*) the sum of the two determinants is the first component of $(\mathbf{a} \times \mathbf{b}) + (\mathbf{a} \times \mathbf{c})$, the right side of (5$\alpha$). For the other components in (5α) and in 5(β), equality follows by the same idea.

Anticommutativity (6) follows from (2**) by noting that the interchange of Rows 2 and 3 multiplies the determinant by -1. We can confirm this geometrically if we set $\mathbf{a} \times \mathbf{b} = \mathbf{v}$ and $\mathbf{b} \times \mathbf{a} = \mathbf{w}$; then $|\mathbf{v}| = |\mathbf{w}|$ by (1), and for $\mathbf{b}, \mathbf{a}, \mathbf{w}$ to form a *right-handed* triple, we must have $\mathbf{w} = -\mathbf{v}$.

Finally, $\mathbf{i} \times (\mathbf{i} \times \mathbf{j}) = \mathbf{i} \times \mathbf{k} = -\mathbf{j}$, whereas $(\mathbf{i} \times \mathbf{i}) \times \mathbf{j} = \mathbf{0} \times \mathbf{j} = \mathbf{0}$ (see Example 2). This proves (7). ■

Typical Applications of Vector Products

EXAMPLE 3 **Moment of a Force**

In mechanics the moment m of a force $\mathbf{p}$ about a point Q is defined as the product $m = |\mathbf{p}|d$, where d is the (perpendicular) distance between Q and the line of action L of $\mathbf{p}$ (Fig. 190). If $\mathbf{r}$ is the vector from Q to any point A on L, then $d = |\mathbf{r}| \sin \gamma$, as shown in Fig. 190, and

$$m = |\mathbf{r}||\mathbf{p}| \sin \gamma.$$

Since γ is the angle between $\mathbf{r}$ and $\mathbf{p}$, we see from (1) that $m = |\mathbf{r} \times \mathbf{p}|$. The vector

(8) $$\mathbf{m} = \mathbf{r} \times \mathbf{p}$$

is called the **moment vector** or **vector moment** of $\mathbf{p}$ about Q. Its magnitude is m. If $\mathbf{m} \neq \mathbf{0}$, its direction is that of the axis of the rotation about Q that $\mathbf{p}$ has the tendency to produce. This axis is perpendicular to both $\mathbf{r}$ and $\mathbf{p}$. ■

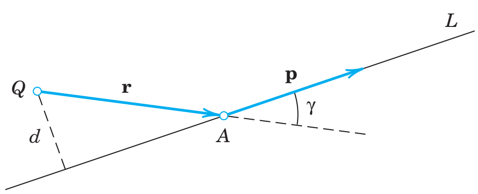


Fig. 190. Moment of a force **p**

EXAMPLE 4 **Moment of a Force**

Find the moment of the force $\mathbf{p}$ about the center Q of a wheel, as given in Fig. 191.

Solution. Introducing coordinates as shown in Fig. 191, we have

$$\mathbf{p} = [1000 \cos 30°, \quad 1000 \sin 30°, \quad 0] = [866, \quad 500, \quad 0], \qquad \mathbf{r} = [0, \quad 1.5, \quad 0].$$

(Note that the center of the wheel is at $y = -1.5$ on the y-axis.) Hence (8) and (2**) give

$$\mathbf{m} = \mathbf{r} \times \mathbf{p} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & 1.5 & 0 \\ 866 & 500 & 0 \end{vmatrix} = 0\mathbf{i} - 0\mathbf{j} + \begin{vmatrix} 0 & 1.5 \\ 866 & 500 \end{vmatrix} \mathbf{k} = [0, 0, -1299].$$

This moment vector $\mathbf{m}$ is normal, i.e., perpendicular to the plane of the wheel. Hence it has the direction of the axis of rotation about the center Q of the wheel that the force $\mathbf{p}$ has the tendency to produce. The moment $\mathbf{m}$ points in the negative z-direction, This is, the direction in which a right-handed screw would advance if turned in that way. ■

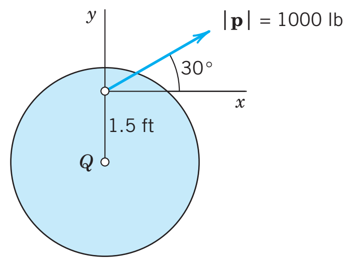


Fig. 191. Moment of a force **p**

EXAMPLE 5 **Velocity of a Rotating Body**

A rotation of a rigid body B in space can be simply and uniquely described by a vector $\mathbf{w}$ as follows. The direction of $\mathbf{w}$ is that of the axis of rotation and such that the rotation appears clockwise if one looks from the initial point of $\mathbf{w}$ to its terminal point. The length of $\mathbf{w}$ is equal to the **angular speed** $\omega(>0)$ of the rotation, that is, the linear (or tangential) speed of a point of B divided by its distance from the axis of rotation.

Let P be any point of B and d its distance from the axis. Then P has the speed ωd. Let $\mathbf{r}$ be the position vector of P referred to a coordinate system with origin 0 on the axis of rotation. Then $d = |\mathbf{r}| \sin \gamma$, where γ is the angle between $\mathbf{w}$ and $\mathbf{r}$. Therefore,

$$\omega d = |\mathbf{w}||\mathbf{r}| \sin \gamma = |\mathbf{w} \times \mathbf{r}|.$$

From this and the definition of vector product we see that the velocity vector $\mathbf{v}$ of P can be represented in the form (Fig. 192)

$$\mathbf{v} = \mathbf{w} \times \mathbf{r}. \tag{9}$$

This simple formula is useful for determining $\mathbf{v}$ at any point of B. ■

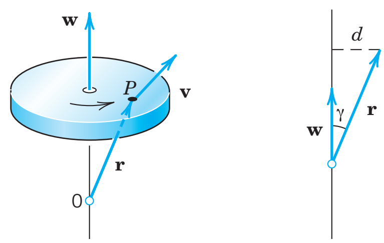


Fig. 192. Rotation of a rigid body

Scalar Triple Product

Certain products of vectors, having three or more factors, occur in applications. The most important of these products is the scalar triple product or mixed product of three vectors **a**, **b**, **c**.

$$(10^*) \qquad (\mathbf{a} \quad \mathbf{b} \quad \mathbf{c}) = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}).$$

The scalar triple product is indeed a scalar since (10*) involves a dot product, which in turn is a scalar. We want to express the scalar triple product in components and as a third-order determinant. To this end, let $\mathbf{a} = [a_1, a_2, a_3]$, $\mathbf{b} = [b_1, b_2, b_3]$, and $\mathbf{c} = [c_1, c_2, c_3]$. Also set $\mathbf{b} \times \mathbf{c} = \mathbf{v} = [v_1, v_2, v_3]$. Then from the dot product in components [formula (2) in Sec. 9.2] and from (2*) with **b** and **c** instead of **a** and **b** we first obtain

$$\begin{aligned}\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) &= \mathbf{a} \cdot \mathbf{v} = a_1v_1 + a_2v_2 + a_3v_3 \\ &= a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} + a_2 \begin{vmatrix} b_3 & b_1 \\ c_3 & c_1 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix}.\end{aligned}$$

The sum on the right is the expansion of a third-order determinant by its first row. Thus we obtain the desired formula for the scalar triple product, that is,

$$(10) \qquad (\mathbf{a} \quad \mathbf{b} \quad \mathbf{c}) = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}.$$

The most important properties of the scalar triple product are as follows.

THEOREM 2

Properties and Applications of Scalar Triple Products

(a) *In* (10) *the dot and cross can be interchanged:*

$$(11) \qquad (\mathbf{a} \quad \mathbf{b} \quad \mathbf{c}) = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}.$$

(b) Geometric interpretation. *The absolute value* $|(\mathbf{a} \quad \mathbf{b} \quad \mathbf{c})|$ *of* (10) *is the volume of the parallelepiped* (oblique box) *with* **a**, **b**, **c** *as edge vectors* (Fig. 193).

(c) Linear independence. *Three vectors in* R^3 *are linearly independent if and only if their scalar triple product is not zero.*

PROOF **(a)** Dot multiplication is commutative, so that by (10)

$$(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c} = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}) = \begin{vmatrix} c_1 & c_2 & c_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}.$$

From this we obtain the determinant in (10) by interchanging Rows 1 and 2 and in the result Rows 2 and 3. But this does not change the value of the determinant because each interchange produces a factor -1, and $(-1)(-1) = 1$. This proves (11).

(b) The volume of that box equals the height $h = |\mathbf{a}||\cos\gamma|$ (Fig. 193) times the area of the base, which is the area $|\mathbf{b} \times \mathbf{c}|$ of the parallelogram with sides **b** and **c**. Hence the volume is

$$|\mathbf{a}||\mathbf{b} \times \mathbf{c}|\,|\cos\gamma| = |\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})| \qquad \text{(Fig. 193)}$$

as given by the absolute value of (11).

(c) Three nonzero vectors, whose initial points coincide, are linearly independent if and only if the vectors do not lie in the same plane nor lie on the same straight line.

This happens if and only if the triple product in (b) is not zero, so that the independence criterion follows. (The case of one of the vectors being the zero vector is trivial.)

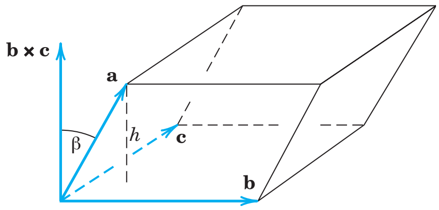


Fig. 193. Geometric interpretation of a scalar triple product

EXAMPLE 6 **Tetrahedron**

A tetrahedron is determined by three edge vectors **a**, **b**, **c**, as indicated in Fig. 194. Find the volume of the tetrahedron in Fig. 194, when $\mathbf{a} = [2, 0, 3]$, $\mathbf{b} = [0, 4, 1]$, $c = [5, 6, 0]$.

Solution. The volume V of the parallelepiped with these vectors as edge vectors is the absolute value of the scalar triple product

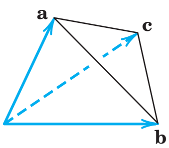


Fig. 194. Tetrahedron

$$(\mathbf{a}\quad \mathbf{b}\quad \mathbf{c}) = \begin{vmatrix} 2 & 0 & 3 \\ 0 & 4 & 1 \\ 5 & 6 & 0 \end{vmatrix} = 2\begin{vmatrix} 4 & 1 \\ 6 & 0 \end{vmatrix} + 3\begin{vmatrix} 0 & 4 \\ 5 & 6 \end{vmatrix} = -12 - 60 = -72.$$

Hence $V = 72$. The minus sign indicates that if the coordinates are right-handed, the triple **a**, **b**, **c** is left-handed. The volume of a tetrahedron is $\frac{1}{6}$ of that of the parallelepiped (can you prove it?), hence 12.

Can you sketch the tetrahedron, choosing the origin as the common initial point of the vectors? What are the coordinates of the four vertices? ■

This is the end of vector *algebra* (in space R^3 and in the plane). Vector *calculus* (differentiation) begins in the next section.

PROBLEM SET 9.3

1–10 **GENERAL PROBLEMS**

1. Give the details of the proofs of Eqs. (4) and (5).
2. What does $\mathbf{a} \times \mathbf{b} = \mathbf{a} \times \mathbf{c}$ with $\mathbf{a} \neq \mathbf{0}$ imply?
3. Give the details of the proofs of Eqs. (6) and (11).
4. **Lagrange's identity for $|\mathbf{a} \times \mathbf{b}|$.** Verify it for $\mathbf{a} = [3, 4, 2]$ and $\mathbf{b} = [1, 0, 2]$. Prove it, using $\sin^2\gamma = 1 - \cos^2\gamma$. The identity is

$$\textbf{(12)} \qquad |\mathbf{a} \times \mathbf{b}| = \sqrt{(\mathbf{a} \cdot \mathbf{a})(\mathbf{b} \cdot \mathbf{b}) - (\mathbf{a} \cdot \mathbf{b})^2}.$$

5. What happens in Example 3 of the text if you replace $\mathbf{p}$ by $-\mathbf{p}$?

6. What happens in Example 5 if you choose a P at distance $2d$ from the axis of rotation?

7. Rotation. A wheel is rotating about the y-axis with angular speed $\omega = 20 \text{ sec}^{-1}$. The rotation appears clockwise if one looks from the origin in the positive y-direction. Find the velocity and speed at the point $[8, 6, 0]$. Make a sketch.

8. Rotation. What are the velocity and speed in Prob. 7 at the point $(4, 2, -2)$ if the wheel rotates about the line $y = x$, $z = 0$ with $\omega = 10 \text{ sec}^{-1}$?

9. Scalar triple product. What does $(\mathbf{a}\ \mathbf{b}\ \mathbf{c}) = 0$ imply with respect to these vectors?

10. WRITING REPORT. Summarize the most important applications discussed in this section. Give examples. No proofs.

11–23 VECTOR AND SCALAR TRIPLE PRODUCTS

With respect to right-handed Cartesian coordinates, let $\mathbf{a} = [2, 1, 0]$, $\mathbf{b} = [-3, 2, 0]$, $\mathbf{c} = [1, 4, -2]$, and $\mathbf{d} = [5, -1, 3]$. Showing details, find:

11. $\mathbf{a} \times \mathbf{b}$, $\mathbf{b} \times \mathbf{a}$, $\mathbf{a} \cdot \mathbf{b}$

12. $3\mathbf{c} \times 5\mathbf{d}$, $15\mathbf{d} \times \mathbf{c}$, $15\mathbf{d} \cdot \mathbf{c}$, $15\mathbf{c} \cdot \mathbf{d}$

13. $\mathbf{c} \times (\mathbf{a} + \mathbf{b})$, $\mathbf{a} \times \mathbf{c} + \mathbf{b} \times \mathbf{c}$

14. $4\mathbf{b} \times 3\mathbf{c} + 12\mathbf{c} \times \mathbf{b}$

15. $(\mathbf{a} + \mathbf{d}) \times (\mathbf{d} + \mathbf{a})$

16. $(\mathbf{b} \times \mathbf{c}) \cdot \mathbf{d}$, $\mathbf{b} \cdot (\mathbf{c} \times \mathbf{d})$

17. $(\mathbf{b} \times \mathbf{c}) \times \mathbf{d}$, $\mathbf{b} \times (\mathbf{c} \times \mathbf{d})$

18. $(\mathbf{a} \times \mathbf{b}) \times \mathbf{a}$, $\mathbf{a} \times (\mathbf{b} \times \mathbf{a})$

19. $(\mathbf{i}\ \mathbf{j}\ \mathbf{k})$, $(\mathbf{i}\ \mathbf{k}\ \mathbf{j})$

20. $(\mathbf{a} \times \mathbf{b}) \times (\mathbf{c} \times \mathbf{d})$, $(\mathbf{a}\ \mathbf{b}\ \mathbf{d})\mathbf{c} - (\mathbf{a}\ \mathbf{b}\ \mathbf{c})\mathbf{d}$

21. $4\mathbf{b} \times 3\mathbf{c}$, $12|\mathbf{b} \times \mathbf{c}|$, $12|\mathbf{c} \times \mathbf{b}|$

22. $(\mathbf{a} - \mathbf{b}\ \ \mathbf{c} - \mathbf{b}\ \ \mathbf{d} - \mathbf{b})$, $(\mathbf{a}\ \mathbf{c}\ \mathbf{d})$

23. $\mathbf{b} \times \mathbf{b}$, $(\mathbf{b} - \mathbf{c}) \times (\mathbf{c} - \mathbf{b})$, $\mathbf{b} \cdot \mathbf{b}$

24. TEAM PROJECT. Useful Formulas for Three and Four Vectors. Prove (13)–(16), which are often useful in practical work, and illustrate each formula with two examples. *Hint.* For (13) choose Cartesian coordinates such that $\mathbf{d} = [d_1, 0, 0]$ and $\mathbf{c} = [c_1, c_2, 0]$. Show that each side of (13) then equals $[-b_2c_2d_1, b_1c_2d_1, 0]$, and give reasons why the two sides are then equal in any Cartesian coordinate system. For (14) and (15) use (13).

(13) $\mathbf{b} \times (\mathbf{c} \times \mathbf{d}) = (\mathbf{b} \cdot \mathbf{d})\mathbf{c} - (\mathbf{b} \cdot \mathbf{c})\mathbf{d}$

(14) $(\mathbf{a} \times \mathbf{b}) \times (\mathbf{c} \times \mathbf{d}) = (\mathbf{a}\ \mathbf{b}\ \mathbf{d})\mathbf{c} - (\mathbf{a}\ \mathbf{b}\ \mathbf{c})\mathbf{d}$

(15) $(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c})$

(16) $(\mathbf{a}\ \mathbf{b}\ \mathbf{c}) = (\mathbf{b}\ \mathbf{c}\ \mathbf{a}) = (\mathbf{c}\ \mathbf{a}\ \mathbf{b})$
$= -(\mathbf{c}\ \mathbf{b}\ \mathbf{a}) = -(\mathbf{a}\ \mathbf{c}\ \mathbf{b})$

25–35 APPLICATIONS

25. Moment m of a force p. Find the moment vector $\mathbf{m}$ and m of $\mathbf{p} = [2, 3, 0]$ about Q: $(2, 1, 0)$ acting on a line through A: $(0, 3, 0)$. Make a sketch.

26. Moment. Solve Prob. 25 if $\mathbf{p} = [1, 0, 3]$, Q: $(2, 0, 3)$, and A: $(4, 3, 5)$.

27. Parallelogram. Find the area if the vertices are $(4, 2, 0)$, $(10, 4, 0)$, $(5, 4, 0)$, and $(11, 6, 0)$. Make a sketch.

28. A remarkable parallelogram. Find the area of the quadrangle Q whose vertices are the midpoints of the sides of the quadrangle P with vertices A: $(2, 1, 0)$, B: $(5, -1. 0)$, C: $(8, 2, 0)$, and D: $(4, 3, 0)$. Verify that Q is a parallelogram.

29. Triangle. Find the area if the vertices are $(0, 0, 1)$, $(2, 0, 5)$, and $(2, 3, 4)$.

30. Plane. Find the plane through the points A: $(1, 2, \frac{1}{4})$, B: $(4, 2, -2)$, and C: $(0, 8, 4)$.

31. Plane. Find the plane through $(1, 3, 4)$, $(1, -2, 6)$, and $(4, 0, 7)$.

32. Parallelepiped. Find the volume if the edge vectors are $\mathbf{i} + \mathbf{j}$, $-2\mathbf{i} + 2\mathbf{k}$, and $-2\mathbf{i} - 3\mathbf{k}$. Make a sketch.

33. Tetrahedron. Find the volume if the vertices are $(1, 1, 1)$, $(5, -7, 3)$, $(7, 4, 8)$, and $(10, 7, 4)$.

34. Tetrahedron. Find the volume if the vertices are $(1, 3, 6)$, $(3, 7, 12)$, $(8, 8, 9)$, and $(2, 2, 8)$.

35. WRITING PROJECT. Applications of Cross Products. Summarize the most important applications we have discussed in this section and give a few simple examples. No proofs.

9.4 Vector and Scalar Functions and Their Fields. Vector Calculus: Derivatives

Our discussion of vector calculus begins with identifying the two types of functions on which it operates. Let P be any point in a domain of definition. Typical domains in applications are three-dimensional, or a surface or a curve in space. Then we define a **vector function** $\mathbf{v}$, whose values are vectors, that is,

$$\mathbf{v} = \mathbf{v}(P) = [v_1(P), v_2(P), v_3(P)]$$

5. What happens in Example 3 of the text if you replace $\mathbf{p}$ by $-\mathbf{p}$?

6. What happens in Example 5 if you choose a P at distance $2d$ from the axis of rotation?

7. Rotation. A wheel is rotating about the y-axis with angular speed $\omega = 20 \text{ sec}^{-1}$. The rotation appears clockwise if one looks from the origin in the positive y-direction. Find the velocity and speed at the point $[8, 6, 0]$. Make a sketch.

8. Rotation. What are the velocity and speed in Prob. 7 at the point $(4, 2, -2)$ if the wheel rotates about the line $y = x$, $z = 0$ with $\omega = 10 \text{ sec}^{-1}$?

9. Scalar triple product. What does $(\mathbf{a}\ \mathbf{b}\ \mathbf{c}) = 0$ imply with respect to these vectors?

10. WRITING REPORT. Summarize the most important applications discussed in this section. Give examples. No proofs.

11–23 VECTOR AND SCALAR TRIPLE PRODUCTS

With respect to right-handed Cartesian coordinates, let $\mathbf{a} = [2, 1, 0]$, $\mathbf{b} = [-3, 2, 0]$, $\mathbf{c} = [1, 4, -2]$, and $\mathbf{d} = [5, -1, 3]$. Showing details, find:

11. $\mathbf{a} \times \mathbf{b}$, $\mathbf{b} \times \mathbf{a}$, $\mathbf{a} \cdot \mathbf{b}$

12. $3\mathbf{c} \times 5\mathbf{d}$, $15\mathbf{d} \times \mathbf{c}$, $15\mathbf{d} \cdot \mathbf{c}$, $15\mathbf{c} \cdot \mathbf{d}$

13. $\mathbf{c} \times (\mathbf{a} + \mathbf{b})$, $\mathbf{a} \times \mathbf{c} + \mathbf{b} \times \mathbf{c}$

14. $4\mathbf{b} \times 3\mathbf{c} + 12\mathbf{c} \times \mathbf{b}$

15. $(\mathbf{a} + \mathbf{d}) \times (\mathbf{d} + \mathbf{a})$

16. $(\mathbf{b} \times \mathbf{c}) \cdot \mathbf{d}$, $\mathbf{b} \cdot (\mathbf{c} \times \mathbf{d})$

17. $(\mathbf{b} \times \mathbf{c}) \times \mathbf{d}$, $\mathbf{b} \times (\mathbf{c} \times \mathbf{d})$

18. $(\mathbf{a} \times \mathbf{b}) \times \mathbf{a}$, $\mathbf{a} \times (\mathbf{b} \times \mathbf{a})$

19. $(\mathbf{i}\ \mathbf{j}\ \mathbf{k})$, $(\mathbf{i}\ \mathbf{k}\ \mathbf{j})$

20. $(\mathbf{a} \times \mathbf{b}) \times (\mathbf{c} \times \mathbf{d})$, $(\mathbf{a}\ \mathbf{b}\ \mathbf{d})\mathbf{c} - (\mathbf{a}\ \mathbf{b}\ \mathbf{c})\mathbf{d}$

21. $4\mathbf{b} \times 3\mathbf{c}$, $12|\mathbf{b} \times \mathbf{c}|$, $12|\mathbf{c} \times \mathbf{b}|$

22. $(\mathbf{a} - \mathbf{b}\ \ \mathbf{c} - \mathbf{b}\ \ \mathbf{d} - \mathbf{b})$, $(\mathbf{a}\ \mathbf{c}\ \mathbf{d})$

23. $\mathbf{b} \times \mathbf{b}$, $(\mathbf{b} - \mathbf{c}) \times (\mathbf{c} - \mathbf{b})$, $\mathbf{b} \cdot \mathbf{b}$

24. TEAM PROJECT. Useful Formulas for Three and Four Vectors. Prove (13)–(16), which are often useful in practical work, and illustrate each formula with two examples. *Hint.* For (13) choose Cartesian coordinates such that $\mathbf{d} = [d_1, 0, 0]$ and $\mathbf{c} = [c_1, c_2, 0]$. Show that each side of (13) then equals $[-b_2c_2d_1, b_1c_2d_1, 0]$, and give reasons why the two sides are then equal in any Cartesian coordinate system. For (14) and (15) use (13).

(13) $\mathbf{b} \times (\mathbf{c} \times \mathbf{d}) = (\mathbf{b} \cdot \mathbf{d})\mathbf{c} - (\mathbf{b} \cdot \mathbf{c})\mathbf{d}$

(14) $(\mathbf{a} \times \mathbf{b}) \times (\mathbf{c} \times \mathbf{d}) = (\mathbf{a}\ \mathbf{b}\ \mathbf{d})\mathbf{c} - (\mathbf{a}\ \mathbf{b}\ \mathbf{c})\mathbf{d}$

(15) $(\mathbf{a} \times \mathbf{b}) \cdot (\mathbf{c} \times \mathbf{d}) = (\mathbf{a} \cdot \mathbf{c})(\mathbf{b} \cdot \mathbf{d}) - (\mathbf{a} \cdot \mathbf{d})(\mathbf{b} \cdot \mathbf{c})$

(16) $(\mathbf{a}\ \mathbf{b}\ \mathbf{c}) = (\mathbf{b}\ \mathbf{c}\ \mathbf{a}) = (\mathbf{c}\ \mathbf{a}\ \mathbf{b})$
$= -(\mathbf{c}\ \mathbf{b}\ \mathbf{a}) = -(\mathbf{a}\ \mathbf{c}\ \mathbf{b})$

25–35 APPLICATIONS

25. Moment m of a force $\mathbf{p}$. Find the moment vector $\mathbf{m}$ and m of $\mathbf{p} = [2, 3, 0]$ about Q: $(2, 1, 0)$ acting on a line through A: $(0, 3, 0)$. Make a sketch.

26. Moment. Solve Prob. 25 if $\mathbf{p} = [1, 0, 3]$, Q: $(2, 0, 3)$, and A: $(4, 3, 5)$.

27. Parallelogram. Find the area if the vertices are $(4, 2, 0)$, $(10, 4, 0)$, $(5, 4, 0)$, and $(11, 6, 0)$. Make a sketch.

28. A remarkable parallelogram. Find the area of the quadrangle Q whose vertices are the midpoints of the sides of the quadrangle P with vertices A: $(2, 1, 0)$, B: $(5, -1. 0)$, C: $(8, 2, 0)$, and D: $(4, 3, 0)$. Verify that Q is a parallelogram.

29. Triangle. Find the area if the vertices are $(0, 0, 1)$, $(2, 0, 5)$, and $(2, 3, 4)$.

30. Plane. Find the plane through the points A: $(1, 2, \frac{1}{4})$, B: $(4, 2, -2)$, and C: $(0, 8, 4)$.

31. Plane. Find the plane through $(1, 3, 4)$, $(1, -2, 6)$, and $(4, 0, 7)$.

32. Parallelepiped. Find the volume if the edge vectors are $\mathbf{i} + \mathbf{j}$, $-2\mathbf{i} + 2\mathbf{k}$, and $-2\mathbf{i} - 3\mathbf{k}$. Make a sketch.

33. Tetrahedron. Find the volume if the vertices are $(1, 1, 1)$, $(5, -7, 3)$, $(7, 4, 8)$, and $(10, 7, 4)$.

34. Tetrahedron. Find the volume if the vertices are $(1, 3, 6)$, $(3, 7, 12)$, $(8, 8, 9)$, and $(2, 2, 8)$.

35. WRITING PROJECT. Applications of Cross Products. Summarize the most important applications we have discussed in this section and give a few simple examples. No proofs.

9.4 Vector and Scalar Functions and Their Fields. Vector Calculus: Derivatives

Our discussion of vector calculus begins with identifying the two types of functions on which it operates. Let P be any point in a domain of definition. Typical domains in applications are three-dimensional, or a surface or a curve in space. Then we define a **vector function** $\mathbf{v}$, whose values are vectors, that is,

$$\mathbf{v} = \mathbf{v}(P) = [v_1(P), v_2(P), v_3(P)]$$

that depends on points P in space. We say that a vector function defines a **vector field** in a domain of definition. Typical domains were just mentioned. Examples of vector fields are the field of tangent vectors of a curve (shown in Fig. 195), normal vectors of a surface (Fig. 196), and velocity field of a rotating body (Fig. 197). Note that vector functions may also depend on time t or on some other parameters.

Similarly, we define a **scalar function** f, whose values are scalars, that is,

$$f = f(P)$$

that depends on P. We say that a scalar function defines a scalar field in that three-dimensional domain or surface or curve in space. Two representative examples of scalar fields are the temperature field of a body and the pressure field of the air in Earth's atmosphere. Note that scalar functions may also depend on some parameter such as time t.

Notation. If we introduce Cartesian coordinates x, y, z, then, instead of writing $\mathbf{v}(P)$ for the vector function, we can write

$$\mathbf{v}(x, y, z) = [v_1(x, y, z), \quad v_2(x, y, z), \quad v_3(x, y, z)].$$

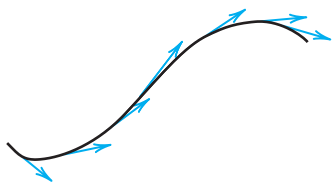


Fig. 195. Field of tangent vectors of a curve

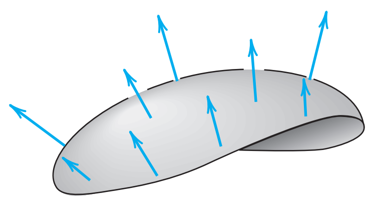


Fig. 196. Field of normal vectors of a surface

We have to keep in mind that the components depend on our choice of coordinate system, whereas a vector field that has a physical or geometric meaning should have magnitude and direction depending only on P, not on the choice of coordinate system.

Similarly, for a scalar function, we write

$$f(P) = f(x, y, z).$$

We illustrate our discussion of vector functions, scalar functions, vector fields, and scalar fields by the following three examples.

EXAMPLE 1 **Scalar Function (Euclidean Distance in Space)**

The distance $f(P)$ of any point P from a fixed point P_0 in space is a scalar function whose domain of definition is the whole space. $f(P)$ defines a scalar field in space. If we introduce a Cartesian coordinate system and P_0 has the coordinates x_0, y_0, z_0, then f is given by the well-known formula

$$f(P) = f(x, y, z) = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}$$

where x, y, z are the coordinates of P. If we replace the given Cartesian coordinate system with another such system by translating and rotating the given system, then the values of the coordinates of P and P_0 will in general change, but $f(P)$ will have the same value as before. Hence $f(P)$ is a scalar function. The direction cosines of the straight line through P and P_0 are not scalars because their values depend on the choice of the coordinate system. ■

EXAMPLE 2 **Vector Field (Velocity Field)**

At any instant the velocity vectors $\mathbf{v}(P)$ of a rotating body B constitute a vector field, called the **velocity field** of the rotation. If we introduce a Cartesian coordinate system having the origin on the axis of rotation, then (see Example 5 in Sec. 9.3)

$$\mathbf{v}(x, y, z) = \mathbf{w} \times \mathbf{r} = \mathbf{w} \times [x, y, z] = \mathbf{w} \times (x\mathbf{i} + y\mathbf{j} + z\mathbf{k}) \tag{1}$$

where x, y, z are the coordinates of any point P of B at the instant under consideration. If the coordinates are such that the z-axis is the axis of rotation and $\mathbf{w}$ points in the positive z-direction, then $\mathbf{w} = \omega\mathbf{k}$ and

$$\mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 0 & 0 & \omega \\ x & y & z \end{vmatrix} = \omega[-y, x, 0] = \omega(-y\mathbf{i} + x\mathbf{j}).$$

An example of a rotating body and the corresponding velocity field are shown in Fig. 197. ■

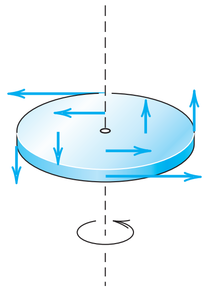


Fig. 197. Velocity field of a rotating body

EXAMPLE 3 **Vector Field (Field of Force, Gravitational Field)**

Let a particle A of mass M be fixed at a point P_0 and let a particle B of mass m be free to take up various positions P in space. Then A attracts B. According to **Newton's law of gravitation** the corresponding gravitational force $\mathbf{p}$ is directed from P to P_0, and its magnitude is proportional to $1/r^2$, where r is the distance between P and P_0, say,

$$|\mathbf{p}| = \frac{c}{r^2}, \qquad c = GMm. \tag{2}$$

Here $G = 6.67 \cdot 10^{-8}\ \mathrm{cm}^3/(\mathrm{g} \cdot \mathrm{sec}^2)$ is the gravitational constant. Hence $\mathbf{p}$ defines a vector field in space. If we introduce Cartesian coordinates such that P_0 has the coordinates x_0, y_0, z_0 and P has the coordinates x, y, z, then by the Pythagorean theorem,

$$r = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2} \qquad (\geqq 0).$$

Assuming that $r > 0$ and introducing the vector

$$\mathbf{r} = [x - x_0, \quad y - y_0, \quad z - z_0] = (x - x_0)\mathbf{i} + (y - y_0)\mathbf{j} + (z - z_0)\mathbf{k},$$

we have $|\mathbf{r}| = r$, and $(-1/r)\mathbf{r}$ is a unit vector in the direction of $\mathbf{p}$; the minus sign indicates that $\mathbf{p}$ is directed from P to P_0 (Fig. 198). From this and (2) we obtain

$$\begin{aligned} \mathbf{p} = |\mathbf{p}|\left(-\frac{1}{r}\mathbf{r}\right) &= -\frac{c}{r^3}\mathbf{r} = \left[-c\,\frac{x - x_0}{r^3}, \quad -c\,\frac{y - y_0}{r^3}, \quad -c\,\frac{z - z_0}{r^3}\right] \\ &= -c\,\frac{x - x_0}{r^3}\,\mathbf{i} - c\,\frac{y - y_0}{r^3}\,\mathbf{j} - c\,\frac{z - z_0}{r^3}\,\mathbf{k}. \end{aligned} \tag{3}$$

This vector function describes the gravitational force acting on B. ■

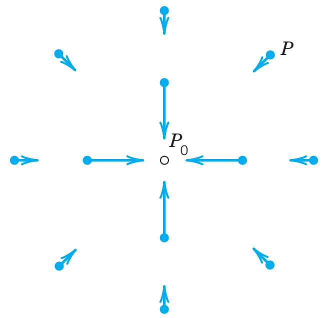


Fig. 198. Gravitational field in Example 3

Vector Calculus

The student may be pleased to learn that many of the concepts covered in (regular) calculus carry over to vector calculus. Indeed, we show how the basic concepts of convergence, continuity, and differentiability from calculus can be defined for vector functions in a simple and natural way. Most important of these is the derivative of a vector function.

Convergence. An infinite sequence of vectors $\mathbf{a}_{(n)}$, $n = 1, 2, \cdots,$ is said to **converge** if there is a vector $\mathbf{a}$ such that

$$\lim_{n\to\infty} |\mathbf{a}_{(n)} - \mathbf{a}| = 0. \tag{4}$$

$\mathbf{a}$ is called the **limit vector** of that sequence, and we write

$$\lim_{n\to\infty} \mathbf{a}_{(n)} = \mathbf{a}. \tag{5}$$

If the vectors are given in Cartesian coordinates, then this sequence of vectors converges to $\mathbf{a}$ if and only if the three sequences of components of the vectors converge to the corresponding components of $\mathbf{a}$. We leave the simple proof to the student.

Similarly, a vector function $\mathbf{v}(t)$ of a real variable t is said to have the **limit** $\boldsymbol{l}$ as t approaches t_0, if $\mathbf{v}(t)$ is defined in some neighborhood of t_0 (possibly except at t_0) and

$$\lim_{t\to t_0} |\mathbf{v}(t) - \boldsymbol{l}| = 0. \tag{6}$$

Then we write

$$\lim_{t\to t_0} \mathbf{v}(t) = \boldsymbol{l}. \tag{7}$$

Here, a *neighborhood* of t_0 is an interval (segment) on the t-axis containing t_0 as an interior point (not as an endpoint).

Continuity. A vector function $\mathbf{v}(t)$ is said to be **continuous** at $t = t_0$ if it is defined in some neighborhood of t_0 (including at t_0 itself!) and

$$\lim_{t\to t_0} \mathbf{v}(t) = \mathbf{v}(t_0). \tag{8}$$

If we introduce a Cartesian coordinate system, we may write

$$\mathbf{v}(t) = [v_1(t), v_2(t), v_3(t)] = v_1(t)\mathbf{i} + v_2(t)\mathbf{j} + v_3(t)\mathbf{k}.$$

Then $\mathbf{v}(t)$ is continuous at t_0 if and only if its three components are continuous at t_0.

We now state the most important of these definitions.

DEFINITION

Derivative of a Vector Function

A vector function $\mathbf{v}(t)$ is said to be **differentiable** at a point t if the following limit exists:

$$\mathbf{v}'(t) = \lim_{\Delta t \to 0} \frac{\mathbf{v}(t + \Delta t) - \mathbf{v}(t)}{\Delta t}. \tag{9}$$

This vector $\mathbf{v}'(t)$ is called the **derivative** of $\mathbf{v}(t)$. See Fig. 199.

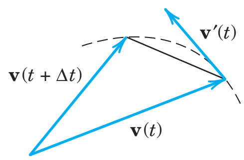


Fig. 199. Derivative of a vector function

In components with respect to a given Cartesian coordinate system,

$$\mathbf{v}'(t) = [v_1'(t), \quad v_2'(t), \quad v_3'(t)]. \tag{10}$$

Hence the derivative* $\mathbf{v}'(t)$ *is obtained by differentiating each component separately. For instance, if $\mathbf{v} = [t, t^2, 0]$, then $\mathbf{v}' = [1, 2t, 0]$.

Equation (10) follows from (9) and conversely because (9) is a "vector form" of the usual formula of calculus by which the derivative of a function of a single variable is defined. [The curve in Fig. 199 is the locus of the terminal points representing $\mathbf{v}(t)$ for values of the independent variable in some interval containing t and $t + \Delta t$ in (9)]. It follows that the familiar differentiation rules continue to hold for differentiating vector functions, for instance,

$$(c\mathbf{v})' = c\mathbf{v}' \qquad (c \text{ constant}),$$

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

and in particular

$$(\mathbf{u} \bullet \mathbf{v})' = \mathbf{u}' \bullet \mathbf{v} + \mathbf{u} \bullet \mathbf{v}' \tag{11}$$

$$(\mathbf{u} \times \mathbf{v})' = \mathbf{u}' \times \mathbf{v} + \mathbf{u} \times \mathbf{v}' \tag{12}$$

$$(\mathbf{u} \quad \mathbf{v} \quad \mathbf{w})' = (\mathbf{u}' \quad \mathbf{v} \quad \mathbf{w}) + (\mathbf{u} \quad \mathbf{v}' \quad \mathbf{w}) + (\mathbf{u} \quad \mathbf{v} \quad \mathbf{w}'). \tag{13}$$

The simple proofs are left to the student. In (12), note the order of the vectors carefully because cross multiplication is not commutative.

EXAMPLE 4 Derivative of a Vector Function of Constant Length

Let $\mathbf{v}(t)$ be a vector function whose length is constant, say, $|\mathbf{v}(t)| = c$. Then $|\mathbf{v}|^2 = \mathbf{v} \bullet \mathbf{v} = c^2$, and $(\mathbf{v} \bullet \mathbf{v})' = 2\mathbf{v} \bullet \mathbf{v}' = 0$, by differentiation [see (11)]. This yields the following result. *The derivative of a vector function* $\mathbf{v}(t)$ *of constant length is either the zero vector or is perpendicular to* $\mathbf{v}(t)$. ■

Partial Derivatives of a Vector Function

Our present discussion shows that partial differentiation of vector functions of two or more variables can be introduced as follows. Suppose that the components of a vector function

$$\mathbf{v} = [v_1, \quad v_2, \quad v_3] = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$$

are differentiable functions of n variables $t_1, \cdots, t_n$. Then the **partial derivative** of $\mathbf{v}$ with respect to t_m is denoted by $\partial\mathbf{v}/\partial t_m$ and is defined as the vector function

$$\frac{\partial \mathbf{v}}{\partial t_m} = \frac{\partial v_1}{\partial t_m}\mathbf{i} + \frac{\partial v_2}{\partial t_m}\mathbf{j} + \frac{\partial v_3}{\partial t_m}\mathbf{k}.$$

Similarly, second partial derivatives are

$$\frac{\partial^2 \mathbf{v}}{\partial t_l \partial t_m} = \frac{\partial^2 v_1}{\partial t_l \partial t_m}\mathbf{i} + \frac{\partial^2 v_2}{\partial t_l \partial t_m}\mathbf{j} + \frac{\partial^2 v_3}{\partial t_l \partial t_m}\mathbf{k},$$

and so on.

EXAMPLE 5 Partial Derivatives

Let $\mathbf{r}(t_1, t_2) = a \cos t_1\,\mathbf{i} + a \sin t_1\,\mathbf{j} + t_2\,\mathbf{k}$. Then $\dfrac{\partial \mathbf{r}}{\partial t_1} = -a \sin t_1\,\mathbf{i} + a \cos t_1\,\mathbf{j}$ and $\dfrac{\partial \mathbf{r}}{\partial t_2} = \mathbf{k}$. ■

Various physical and geometric applications of derivatives of vector functions will be discussed in the next sections as well as in Chap. 10.

PROBLEM SET 9.4

1–8 SCALAR FIELDS IN THE PLANE

Let the temperature T in a body be independent of z so that it is given by a scalar function $T = T(x, t)$. Identify the isotherms $T(x, y) = \text{const}$. Sketch some of them.

1. $T = x^2 - y^2$
2. $T = xy$
3. $T = 3x - 4y$
4. $T = \arctan(y/x)$
5. $T = y/(x^2 + y^2)$
6. $T = x/(x^2 + y^2)$
7. $T = 9x^2 + 4y^2$
8. **CAS PROJECT. Scalar Fields in the Plane.** Sketch or graph isotherms of the following fields and describe what they look like.

   (a) $x^2 - 4x - y^2$ **(b)** $x^2y - y^3/3$
   (c) $\cos x \sinh y$ **(d)** $\sin x \sinh y$
   (e) $e^x \sin y$ **(f)** $e^{2x} \cos 2y$
   (g) $x^4 - 6x^2y^2 + y^4$ **(h)** $x^2 - 2x - y^2$

9–14 SCALAR FIELDS IN SPACE

What kind of surfaces are the **level surfaces** $f(x, y, z) = \text{const}$?

9. $f = 4x - 3y + 2z$
10. $f = 9(x^2 + y^2) + z^2$
11. $f = 5x^2 + 2y^2$
12. $f = z - \sqrt{x^2 + y^2}$
13. $f = z - (x^2 + y^2)$
14. $f = x - y^2$

15–20 VECTOR FIELDS

Sketch figures similar to Fig. 198. Try to interpet the field of $\mathbf{v}$ as a velocity field.

15. $\mathbf{v} = \mathbf{i} + \mathbf{j}$

16. $\mathbf{v} = -y\mathbf{i} + x\mathbf{j}$

17. $\mathbf{v} = x\mathbf{j}$

18. $\mathbf{v} = x\mathbf{i} + y\mathbf{j}$

19. $\mathbf{v} = x\mathbf{i} - y\mathbf{j}$

20. $\mathbf{v} = y\mathbf{i} - x\mathbf{j}$

21. CAS PROJECT. Vector Fields. Plot by arrows:

(a) $\mathbf{v} = [x, x^2]$ **(b)** $\mathbf{v} = [1/y, 1/x]$

(c) $\mathbf{v} = [\cos x, \sin x]$ **(d)** $\mathbf{v} = e^{-(x^2+y^2)}[x, -y]$

22–25 DIFFERENTIATION

22. Find the first and second derivatives of $\mathbf{r} = [3 \cos 2t, 3 \sin 2t, 4t]$.

23. Prove (11)–(13). Give two typical examples for each formula.

24. Find the first partial derivatives of $\mathbf{v}_1 = [e^x \cos y, e^x \sin y]$ and $\mathbf{v}_2 = [\cos x \cosh y, -\sin x \sinh y]$.

25. WRITING PROJECT. Differentiation of Vector Functions. Summarize the essential ideas and facts and give examples of your own.

9.5 Curves. Arc Length. Curvature. Torsion

Vector calculus has important applications to curves (Sec. 9.5) and surfaces (to be covered in Sec. 10.5) in physics and geometry. The application of vector calculus to geometry is a field known as **differential geometry**. Differential geometric methods are applied to problems in mechanics, computer-aided as well as traditional engineering design, geodesy, geography, space travel, and relativity theory. For details, see [GenRef8] and [GenRef9] in App. 1.

Bodies that move in space form paths that may be represented by curves C. This and other applications show the need for **parametric representations** of C with **parameter** t, which may denote time or something else (see Fig. 200). A typical parametric representation is given by

$$\mathbf{r}(t) = [x(t), \quad y(t), \quad z(t)] = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}. \tag{1}$$

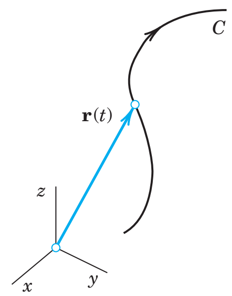


Fig. 200. Parametric representation of a curve

Here t is the parameter and x, y, z are Cartesian coordinates, that is, the usual rectangular coordinates as shown in Sec. 9.1. To each value $t = t_0$, there corresponds a point of C with position vector $\mathbf{r}(t_0)$ whose coordinates are $x(t_0), y(t_0), z(t_0)$. This is illustrated in Figs. 201 and 202.

The use of parametric representations has key advantages over other representations that involve projections into the xy-plane and xz-plane or involve a pair of equations with y or with z as independent variable. The projections look like this:

$$y = f(x), \qquad z = g(x). \tag{2}$$

Mean Value Theorems

THEOREM 2

Mean Value Theorem

Let $f(x, y, z)$ be continuous and have continuous first partial derivatives in a domain D in xyz-space. Let P_0: (x_0, y_0, z_0) and P: $(x_0 + h, y_0 + k, z_0 + l)$ be points in D such that the straight line segment P_0P joining these points lies entirely in D. Then

$$f(x_0 + h, y_0 + k, z_0 + l) - f(x_0, y_0, z_0) = h\frac{\partial f}{\partial x} + k\frac{\partial f}{\partial y} + l\frac{\partial f}{\partial z}, \tag{7}$$

the partial derivatives being evaluated at a suitable point of that segment.

Special Cases

For a function $f(x, y)$ of two variables (satisfying assumptions as in the theorem), formula (7) reduces to (Fig. 214)

$$f(x_0 + h, y_0 + k) - f(x_0, y_0) = h\frac{\partial f}{\partial x} + k\frac{\partial f}{\partial y}, \tag{8}$$

and, for a function $f(x)$ of a single variable, (7) becomes

$$f(x_0 + h) - f(x_0) = h\frac{\partial f}{\partial x}, \tag{9}$$

where in (9), the domain D is a segment of the x-axis and the derivative is taken at a suitable point between x_0 and $x_0 + h$.

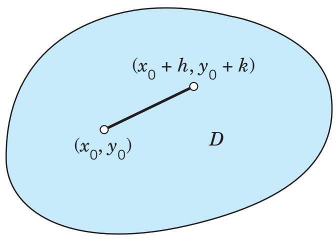


Fig. 214. Mean value theorem for a function of two variables [Formula (8)]

9.7 Gradient of a Scalar Field. Directional Derivative

We shall see that *some* of the vector fields that occur in applications—not all of them!—can be obtained from scalar fields. Using scalar fields instead of vector fields is of a considerable advantage because scalar fields are easier to use than vector fields. It is the "gradient" that allows us to obtain vector fields from scalar fields, and thus the gradient is of great practical importance to the engineer.

DEFINITION 1

Gradient

The setting is that we are given a scalar function $f(x, y, z)$ that is defined and differentiable in a domain in 3-space with Cartesian coordinates x, y, z. We denote the **gradient** of that function by grad f or ∇f (read **nabla** f). Then the qradient of $f(x, y, z)$ is defined as the vector function

$$\text{grad } f = \nabla f = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right] = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} + \frac{\partial f}{\partial z}\mathbf{k}. \tag{1}$$

Remarks. For a definition of the gradient in curvilinear coordinates, see App. 3.4. As a quick example, if $f(x, y, z) = 2y^3 + 4xz + 3x$, then grad $f = [4z + 3, 6y^2, 4x]$. Furthermore, we will show later in this section that (1) actually does define a vector.

The notation ∇f is suggested by the *differential operator* ∇ (read *nabla*) defined by

$$\nabla = \frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}. \tag{1*}$$

Gradients are useful in several ways, notably in giving the rate of change of $f(x, y, z)$ in any direction in space, in obtaining surface normal vectors, and in deriving vector fields from scalar fields, as we are going to show in this section.

Directional Derivative

From calculus we know that the partial derivatives in (1) give the rates of change of $f(x, y, z)$ in the directions of the three coordinate axes. It seems natural to extend this and ask for the rate of change of f *in an arbitrary direction* in space. This leads to the following concept.

DEFINITION 2

Directional Derivative

The directional derivative $D_{\mathbf{b}}f$ or df/ds of a function $f(x, y, z)$ at a point P in the direction of a vector $\mathbf{b}$ is defined by (see Fig. 215)

$$D_{\mathbf{b}}f = \frac{df}{ds} = \lim_{s\to 0}\frac{f(Q) - f(P)}{s}. \tag{2}$$

Here Q is a variable point on the straight line L in the direction of $\mathbf{b}$, and $|s|$ is the distance between P and Q. Also, $s > 0$ if Q lies in the direction of $\mathbf{b}$ (as in Fig. 215), $s < 0$ if Q lies in the direction of $-\mathbf{b}$, and $s = 0$ if $Q = P$.

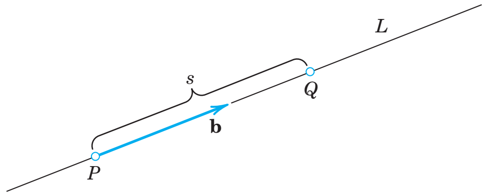


Fig. 215. Directional derivative

The next idea is to use Cartesian xyz-coordinates and for $\mathbf{b}$ a unit vector. Then the line L is given by

$$\mathbf{r}(s) = x(s)\mathbf{i} + y(s)\mathbf{j} + z(s)\mathbf{k} = \mathbf{p}_0 + s\mathbf{b} \qquad (|\mathbf{b}| = 1) \tag{3}$$

where $\mathbf{p}_0$ the position vector of P. Equation (2) now shows that $D_{\mathbf{b}}f = df/ds$ is the derivative of the function $f(x(s), y(s), z(s))$ with respect to the arc length s of L. Hence, assuming that f has continuous partial derivatives and applying the chain rule [formula (3) in the previous section], we obtain

$$D_{\mathbf{b}}f = \frac{df}{ds} = \frac{\partial f}{\partial x}x' + \frac{\partial f}{\partial y}y' + \frac{\partial f}{\partial z}z' \tag{4}$$

where primes denote derivatives with respect to s (which are taken at $s = 0$). But here, differentiating (3) gives $\mathbf{r}' = x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k} = \mathbf{b}$. Hence (4) is simply the inner product of grad f and $\mathbf{b}$ [see (2), Sec. 9.2]; that is,

$$\boxed{D_{\mathbf{b}}f = \frac{df}{ds} = \mathbf{b} \cdot \operatorname{grad} f} \qquad (|\mathbf{b}| = 1). \tag{5}$$

ATTENTION! If the direction is given by a vector $\mathbf{a}$ of any length ($\neq 0$), then

$$\boxed{D_{\mathbf{a}}f = \frac{df}{ds} = \frac{1}{|\mathbf{a}|}\mathbf{a} \cdot \operatorname{grad} f.} \tag{5*}$$

EXAMPLE 1 Gradient. Directional Derivative

Find the directional derivative of $f(x, y, z) = 2x^2 + 3y^2 + z^2$ at P: (2, 1, 3) in the direction of $\mathbf{a} = [1, 0, -2]$.

Solution. grad $f = [4x, 6y, 2z]$ gives at P the vector grad $f(P) = [8, 6, 6]$. From this and (5*) we obtain, since $|\mathbf{a}| = \sqrt{5}$,

$$D_{\mathbf{a}}f(P) = \frac{1}{\sqrt{5}}[1, 0, -2] \cdot [8, 6, 6] = \frac{1}{\sqrt{5}}(8 + 0 - 12) = -\frac{4}{\sqrt{5}} = -1.789.$$

The minus sign indicates that at P the function f is decreasing in the direction of $\mathbf{a}$. ■

Gradient Is a Vector. Maximum Increase

Here is a finer point of mathematics that concerns the consistency of our theory: grad f in (1) *looks* like a vector—after all, it has three components! But to prove that it *actually is* a vector, since it is defined in terms of components depending on the Cartesian coordinates, we must show that grad f has a length and direction independent of the choice of those coordinates. See proof of Theorem 1. In contrast, $[\partial f/\partial x, 2\partial f/\partial y, \partial f/\partial z]$ also looks like a vector but does not have a length and direction independent of the choice of Cartesian coordinates.

Incidentally, the direction makes the gradient eminently useful: grad f *points in the direction of maximum increase of* f.

THEOREM 1

Use of Gradient: Direction of Maximum Increase

Let $f(P) = f(x, y, z)$ *be a scalar function having continuous first partial derivatives in some domain* B *in space. Then* grad f *exists in* B *and is a vector, that is, its length and direction are independent of the particular choice of Cartesian coordinates. If* grad $f(P) \neq \mathbf{0}$ *at some point* P*, it has the direction of maximum increase of* f *at* P.

PROOF From (5) and the definition of inner product [(1) in Sec. 9.2] we have

$$D_{\mathbf{b}} f = |\mathbf{b}||\text{grad } f| \cos \gamma = |\text{grad } f| \cos \gamma \tag{6}$$

where γ is the angle between $\mathbf{b}$ and grad f. Now f is a scalar function. Hence its value at a point P depends on P but not on the particular choice of coordinates. The same holds for the arc length s of the line L in Fig. 215, hence also for $D_{\mathbf{b}} f$. Now (6) shows that $D_{\mathbf{b}} f$ is maximum when $\cos \gamma = 1$, $\gamma = 0$, and then $D_{\mathbf{b}} f = |\text{grad } f|$. It follows that the length and direction of grad f are independent of the choice of coordinates. Since $\gamma = 0$ if and only if $\mathbf{b}$ has the direction of grad f, the latter is the direction of maximum increase of f at P, provided grad $f \neq \mathbf{0}$ at P. Make sure that you understood the proof to get a good feel for mathematics.

Gradient as Surface Normal Vector

Gradients have an important application in connection with surfaces, namely, as surface normal vectors, as follows. Let S be a surface represented by $f(x, y, z) = c = \text{const}$, where f is differentiable. Such a surface is called a **level surface** of f, and for different c we get different level surfaces. Now let C be a curve on S through a point P of S. As a curve in space, C has a representation $\mathbf{r}(t) = [x(t), y(t), z(t)]$. For C to lie on the surface S, the components of $\mathbf{r}(t)$ must satisfy $f(x, y, z) = c$, that is,

$$f(x(t), y(t), z(t)) = c. \tag{7}$$

Now a tangent vector of C is $\mathbf{r}'(t) = [x'(t), y'(t), z'(t)]$. And the tangent vectors of all curves on S passing through P will generally form a plane, called the **tangent plane** of S at P. (Exceptions occur at edges or cusps of S, for instance, at the apex of the cone in Fig. 217.) The normal of this plane (the straight line through P perpendicular to the tangent plane) is called the **surface normal** to S at P. A vector in the direction of the surface normal is called a **surface normal vector** of S at P. We can obtain such a vector quite simply by differentiating (7) with respect to t. By the chain rule,

$$\frac{\partial f}{\partial x}x' + \frac{\partial f}{\partial y}y' + \frac{\partial f}{\partial z}z' = (\operatorname{grad} f) \bullet \mathbf{r}' = 0.$$

Hence grad f is orthogonal to all the vectors $\mathbf{r}'$ in the tangent plane, so that it is a normal vector of S at P. Our result is as follows (see Fig. 216).

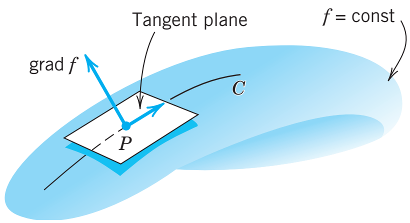


Fig. 216. Gradient as surface normal vector

THEOREM 2

Gradient as Surface Normal Vector

Let f be a differentiable scalar function in space. Let $f(x, y, z) = c = const$ represent a surface S. Then if the gradient of f at a point P of S is not the zero vector, it is a normal vector of S at P.

EXAMPLE 2

Gradient as Surface Normal Vector. Cone

Find a unit normal vector $\mathbf{n}$ of the cone of revolution $z^2 = 4(x^2 + y^2)$ at the point P: $(1, 0, 2)$.

Solution. The cone is the level surface $f = 0$ of $f(x, y, z) = 4(x^2 + y^2) - z^2$. Thus (Fig. 217)

$$\operatorname{grad} f = [8x, \quad 8y, \quad -2z], \qquad \operatorname{grad} f(P) = [8, \quad 0, \quad -4]$$

$$\mathbf{n} = \frac{1}{|\operatorname{grad} f(P)|}\operatorname{grad} f(P) = \left[\frac{2}{\sqrt{5}}, \quad 0, \quad -\frac{1}{\sqrt{5}}\right].$$

$\mathbf{n}$ points downward since it has a negative z-component. The other unit normal vector of the cone at P is $-\mathbf{n}$. ■

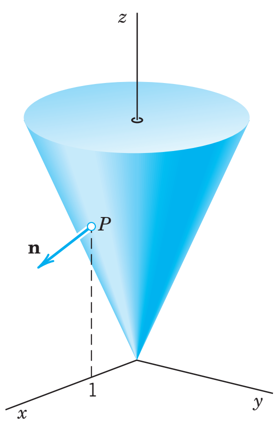


Fig. 217. Cone and unit normal vector **n**

Vector Fields That Are Gradients of Scalar Fields ("Potentials")

At the beginning of this section we mentioned that some vector fields have the advantage that they can be obtained from scalar fields, which can be worked with more easily. Such a vector field is given by a vector function $\mathbf{v}(P)$, which is obtained as the gradient of a scalar function, say, $\mathbf{v}(P) = \text{grad}\, f(P)$. The function $f(P)$ is called a *potential function* or a **potential** of $\mathbf{v}(P)$. Such a $\mathbf{v}(P)$ and the corresponding vector field are called **conservative** because in such a vector field, energy is conserved; that is, no energy is lost (or gained) in displacing a body (or a charge in the case of an electrical field) from a point P to another point in the field and back to P. We show this in Sec. 10.2.

Conservative fields play a central role in physics and engineering. A basic application concerns the gravitational force (see Example 3 in Sec. 9.4) and we show that it has a potential which satisfies Laplace's equation, the most important partial differential equation in physics and its applications.

THEOREM 3

Gravitational Field. Laplace's Equation

The force of attraction

$$\mathbf{p} = -\frac{c}{r^3}\mathbf{r} = -c\left[\frac{x - x_0}{r^3}, \frac{y - y_0}{r^3}, \frac{z - z_0}{r^3}\right] \tag{8}$$

between two particles at points P_0: (x_0, y_0, z_0) and P: (x, y, z) (as given by Newton's law of gravitation) *has the potential* $f(x, y, z) = c/r$, *where* $r\ (> 0)$ *is the distance between* P_0 *and* P.

Thus $\mathbf{p} = \text{grad}\, f = \text{grad}\,(c/r)$. *This potential* f *is a solution of* **Laplace's equation**

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = 0. \tag{9}$$

[$\nabla^2 f$ (read *nabla squared* f) is called the **Laplacian** of f.]

PROOF That distance is $r = ((x - x_0)^2 + (y - y_0)^2 + (z - z_2)^2)^{1/2}$. The key observation now is that for the components of $\mathbf{p} = [p_1, p_2, p_3]$ we obtain by partial differentiation

$$\frac{\partial}{\partial x}\left(\frac{1}{r}\right) = \frac{-2(x - x_0)}{2[(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2]^{3/2}} = -\frac{x - x_0}{r^3} \tag{10a}$$

and similarly

$$\begin{aligned} \frac{\partial}{\partial y}\left(\frac{1}{r}\right) &= -\frac{y - y_0}{r^3}, \\ \frac{\partial}{\partial z}\left(\frac{1}{r}\right) &= -\frac{z - z_0}{r^3}. \end{aligned} \tag{10b}$$

From this we see that, indeed, **p** is the gradient of the scalar function $f = c/r$. The second statement of the theorem follows by partially differentiating (10), that is,

$$\frac{\partial^2}{\partial x^2}\left(\frac{1}{r}\right) = -\frac{1}{r^3} + \frac{3(x - x_0)^2}{r^5},$$

$$\frac{\partial^2}{\partial y^2}\left(\frac{1}{r}\right) = -\frac{1}{r^3} + \frac{3(y - y_0)^2}{r^5},$$

$$\frac{\partial^2}{\partial z^2}\left(\frac{1}{r}\right) = -\frac{1}{r^3} + \frac{3(z - z_0)^2}{r^5},$$

and then adding these three expressions. Their common denominator is r^5. Hence the three terms $-1/r^3$ contribute $-3r^2$ to the numerator, and the three other terms give the sum

$$3(x - x_0)^2 + 3(y - y_0)^2 + 3(z - z_0)^2 = 3r^2,$$

so that the numerator is 0, and we obtain (9). ■

$\nabla^2 f$ is also denoted by Δf. The differential operator

$$\nabla^2 = \Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{11}$$

(read "nabla squared" or "delta") is called the **Laplace operator**. It can be shown that the field of force produced by any distribution of masses is given by a vector function that is the gradient of a scalar function f, and f satisfies (9) in any region that is free of matter.

The great importance of the Laplace equation also results from the fact that there are other laws in physics that are of the same form as Newton's law of gravitation. For instance, in electrostatics the force of attraction (or repulsion) between two particles of opposite (or like) charge Q_1 and Q_2 is

$$\mathbf{p} = \frac{k}{r^3}\mathbf{r} \qquad \textbf{(Coulomb's law}^6\textbf{)}. \tag{12}$$

Laplace's equation will be discussed in detail in Chaps. 12 and 18.

A method for finding out whether a given vector field has a potential will be explained in Sec. 9.9.

---

[6] CHARLES AUGUSTIN DE COULOMB (1736–1806), French physicist and engineer. Coulomb's law was derived by him from his own very precise measurements.

PROBLEM SET 9.7

1–6 CALCULATION OF GRADIENTS

Find grad f. Graph some level curves $f = \text{const}$. Indicate ∇f by arrows at some points of these curves.

1. $f = (x + 1)(2y - 1)$

2. $f = 9x^2 + 4y^2$

3. $f = y/x$

4. $(y + 6)^2 + (x - 4)^2$

5. $f = x^4 + y^4$

6. $f = (x^2 - y^2)/(x^2 + y^2)$

7–10 USEFUL FORMULAS FOR GRADIENT AND LAPLACIAN

Prove and illustrate by an example.

7. $\nabla(f^n) = nf^{n-1}\nabla f$

8. $\nabla(fg) = f\nabla g + g\nabla f$

9. $\nabla(f/g) = (1/g^2)(g\nabla f - f\nabla g)$

10. $\nabla^2(fg) = g\nabla^2 f + 2\nabla f \cdot \nabla g + f\nabla^2 g$

11–15 USE OF GRADIENTS. ELECTRIC FORCE

The force in an electrostatic field given by $f(x, y, z)$ has the direction of the gradient. Find ∇f and its value at P.

11. $f = xy, \quad P: (-4, 5)$

12. $f = x/(x^2 + y^2), \quad P: (1, 1)$

13. $f = \ln(x^2 + y^2), \quad P: (8, 6)$

14. $f = (x^2 + y^2 + z^2)^{-1/2} \quad P: (12, 0, 16)$

15. $f = 4x^2 + 9y^2 + z^2, \quad P: (5, -1, -11)$

16. For what points $P: (x, y, z)$ does ∇f with $f = 25x^2 + 9y^2 + 16z^2$ have the direction from P to the origin?

17. Same question as in Prob. 16 when $f = 25x^2 + 4y^2$.

18–23 VELOCITY FIELDS

Given the velocity potential f of a flow, find the velocity $\mathbf{v} = \nabla f$ of the field and its value $\mathbf{v}(P)$ at P. Sketch $\mathbf{v}(P)$ and the curve $f = \text{const}$ passing through P.

18. $f = x^2 - 6x - y^2, \quad P: (-1, 5)$

19. $f = \cos x \cosh y, \quad P: (\tfrac{1}{2}\pi, \ln 2)$

20. $f = x(1 + (x^2 + y^2)^{-1}), \quad P: (1, 1)$

21. $f = e^x \cos y, \quad P: (1, \tfrac{1}{2}\pi)$

22. At what points is the flow in Prob. 21 directed vertically upward?

23. At what points is the flow in Prob. 21 horizontal?

24–27 HEAT FLOW

Experiments show that in a temperature field, heat flows in the direction of maximum decrease of temperature T. Find this direction in general and at the given point P. Sketch that direction at P as an arrow.

24. $T = 3x^2 - 2y^2, \quad P: (2.5, 1.8)$

25. $T = z/(x^2 + y^2), \quad P: (0, 1, 2)$

26. $T = x^2 + y^2 + 4z^2, \quad P: (2, -1, 2)$

27. CAS PROJECT. Isotherms. Graph some curves of constant temperature ("isotherms") and indicate directions of heat flow by arrows when the temperature equals **(a)** $x^3 - 3xy^2$, **(b)** $\sin x \sinh y$, and **(c)** $e^x \cos y$.

28. Steepest ascent. If $z(x, y) = 3000 - x^2 - 9y^2$ [meters] gives the elevation of a mountain at sea level, what is the direction of steepest ascent at $P: (4, 1)$?

29. Gradient. What does it mean if $|\nabla f(P)| > |\nabla f(Q)|$ at two points P and Q in a scalar field?

9.8 Divergence of a Vector Field

Vector calculus owes much of its importance in engineering and physics to the gradient, divergence, and curl. From a scalar field we can obtain a vector field by the gradient (Sec. 9.7). Conversely, from a vector field we can obtain a scalar field by the divergence or another vector field by the curl (to be discussed in Sec. 9.9). These concepts were suggested by basic physical applications. This will be evident from our examples.

To begin, let $\mathbf{v}(x, y, z)$ be a differentiable vector function, where x, y, z are Cartesian coordinates, and let v_1, v_2, v_3 be the components of $\mathbf{v}$. Then the function

$$\text{div}\,\mathbf{v} = \frac{\partial v_1}{\partial x} + \frac{\partial v_2}{\partial y} + \frac{\partial v_3}{\partial z} \tag{1}$$

PROBLEM SET 9.7

1–6 CALCULATION OF GRADIENTS

Find grad f. Graph some level curves $f = \text{const}$. Indicate ∇f by arrows at some points of these curves.

1. $f = (x + 1)(2y - 1)$
2. $f = 9x^2 + 4y^2$
3. $f = y/x$
4. $(y + 6)^2 + (x - 4)^2$
5. $f = x^4 + y^4$
6. $f = (x^2 - y^2)/(x^2 + y^2)$

7–10 USEFUL FORMULAS FOR GRADIENT AND LAPLACIAN

Prove and illustrate by an example.

7. $\nabla(f^n) = nf^{n-1}\nabla f$
8. $\nabla(fg) = f\nabla g + g\nabla f$
9. $\nabla(f/g) = (1/g^2)(g\nabla f - f\nabla g)$
10. $\nabla^2(fg) = g\nabla^2 f + 2\nabla f \cdot \nabla g + f\nabla^2 g$

11–15 USE OF GRADIENTS. ELECTRIC FORCE

The force in an electrostatic field given by $f(x, y, z)$ has the direction of the gradient. Find ∇f and its value at P.

11. $f = xy$, $\quad P$: $(-4, 5)$
12. $f = x/(x^2 + y^2)$, $\quad P$: $(1, 1)$
13. $f = \ln(x^2 + y^2)$, $\quad P$: $(8, 6)$
14. $f = (x^2 + y^2 + z^2)^{-1/2}$ $\quad P$: $(12, 0, 16)$
15. $f = 4x^2 + 9y^2 + z^2$, $\quad P$: $(5, -1, -11)$
16. For what points P: (x, y, z) does ∇f with $f = 25x^2 + 9y^2 + 16z^2$ have the direction from P to the origin?
17. Same question as in Prob. 16 when $f = 25x^2 + 4y^2$.

18–23 VELOCITY FIELDS

Given the velocity potential f of a flow, find the velocity $\mathbf{v} = \nabla f$ of the field and its value $\mathbf{v}(P)$ at P. Sketch $\mathbf{v}(P)$ and the curve $f = \text{const}$ passing through P.

18. $f = x^2 - 6x - y^2$, $\quad P$: $(-1, 5)$
19. $f = \cos x \cosh y$, $\quad P$: $(\frac{1}{2}\pi, \ln 2)$
20. $f = x(1 + (x^2 + y^2)^{-1})$, $\quad P$: $(1, 1)$
21. $f = e^x \cos y$, $\quad P$: $(1, \frac{1}{2}\pi)$
22. At what points is the flow in Prob. 21 directed vertically upward?
23. At what points is the flow in Prob. 21 horizontal?

24–27 HEAT FLOW

Experiments show that in a temperature field, heat flows in the direction of maximum decrease of temperature T. Find this direction in general and at the given point P. Sketch that direction at P as an arrow.

24. $T = 3x^2 - 2y^2$, $\quad P$: $(2.5, 1.8)$
25. $T = z/(x^2 + y^2)$, $\quad P$: $(0, 1, 2)$
26. $T = x^2 + y^2 + 4z^2$, $\quad P$: $(2, -1, 2)$
27. **CAS PROJECT. Isotherms.** Graph some curves of constant temperature ("isotherms") and indicate directions of heat flow by arrows when the temperature equals **(a)** $x^3 - 3xy^2$, **(b)** $\sin x \sinh y$, and **(c)** $e^x \cos y$.
28. **Steepest ascent.** If $z(x, y) = 3000 - x^2 - 9y^2$ [meters] gives the elevation of a mountain at sea level, what is the direction of steepest ascent at P: $(4, 1)$?
29. **Gradient.** What does it mean if $|\nabla f(P)| > |\nabla f(Q)|$ at two points P and Q in a scalar field?

9.8 Divergence of a Vector Field

Vector calculus owes much of its importance in engineering and physics to the gradient, divergence, and curl. From a scalar field we can obtain a vector field by the gradient (Sec. 9.7). Conversely, from a vector field we can obtain a scalar field by the divergence or another vector field by the curl (to be discussed in Sec. 9.9). These concepts were suggested by basic physical applications. This will be evident from our examples.

To begin, let $\mathbf{v}(x, y, z)$ be a differentiable vector function, where x, y, z are Cartesian coordinates, and let v_1, v_2, v_3 be the components of $\mathbf{v}$. Then the function

$$\text{div}\,\mathbf{v} = \frac{\partial v_1}{\partial x} + \frac{\partial v_2}{\partial y} + \frac{\partial v_3}{\partial z} \tag{1}$$

is called the **divergence** of $\mathbf{v}$ or the *divergence of the vector field defined by* $\mathbf{v}$. For example, if

$$\mathbf{v} = [3xz, 2xy, -yz^2] = 3xz\mathbf{i} + 2xy\mathbf{j} - yz^2\mathbf{k}, \qquad \text{then} \qquad \operatorname{div} \mathbf{v} = 3z + 2x - 2yz.$$

Another common notation for the divergence is

$$\begin{aligned}\operatorname{div} \mathbf{v} = \nabla \bullet \mathbf{v} &= \left[\frac{\partial}{\partial x}, \frac{\partial}{\partial \mathbf{y}}, \frac{\partial}{\partial \mathbf{z}}\right] \bullet [v_1, v_2, v_3] \\ &= \left(\frac{\partial}{\partial x}\mathbf{i} + \frac{\partial}{\partial y}\mathbf{j} + \frac{\partial}{\partial z}\mathbf{k}\right) \bullet (v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}) \\ &= \frac{\partial v_1}{\partial x} + \frac{\partial v_2}{\partial y} + \frac{\partial v_3}{\partial z},\end{aligned}$$

with the understanding that the "product" $(\partial/\partial x)v_1$ in the dot product means the partial derivative $\partial v_1/\partial x$, etc. This is a convenient notation, but nothing more. Note that $\nabla \bullet \mathbf{v}$ means the scalar div $\mathbf{v}$, whereas ∇f means the vector grad f defined in Sec. 9.7.

In Example 2 we shall see that the divergence has an important physical meaning. Clearly, the values of a function that characterizes a physical or geometric property must be independent of the particular choice of coordinates. In other words, these values must be invariant with respect to coordinate transformations. Accordingly, the following theorem should hold.

THEOREM 1

Invariance of the Divergence

The divergence div $\mathbf{v}$ *is a scalar function, that is, its values depend only on the points in space (and, of course, on* $\mathbf{v}$*) but not on the choice of the coordinates in* (1)*, so that with respect to other Cartesian coordinates* x^*, y^*, z^* *and corresponding components* v_1^*, v_2^*, v_3^* *of* $\mathbf{v}$,

$$\operatorname{div} \mathbf{v} = \frac{\partial v_1^*}{\partial x^*} + \frac{\partial v_2^*}{\partial y^*} + \frac{\partial v_3^*}{\partial z^*}. \tag{2}$$

We shall prove this theorem in Sec. 10.7, using integrals.

Presently, let us turn to the more immediate practical task of gaining a feel for the significance of the divergence. Let $f(x, y, z)$ be a twice differentiable scalar function. Then its gradient exists,

$$\mathbf{v} = \operatorname{grad} f = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right] = \frac{\partial f}{\partial x}\mathbf{i} + \frac{\partial f}{\partial y}\mathbf{j} + \frac{\partial f}{\partial z}\mathbf{k}$$

and we can differentiate once more, the first component with respect to x, the second with respect to y, the third with respect to z, and then form the divergence,

$$\operatorname{div} \mathbf{v} = \operatorname{div}(\operatorname{grad} f) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}.$$

Hence we have the basic result that *the divergence of the gradient is the Laplacian* (Sec. 9.7),

$$\text{div}\,(\text{grad}\, f) = \nabla^2 f. \tag{3}$$

EXAMPLE 1 Gravitational Force. Laplace's Equation

The gravitational force $\mathbf{p}$ in Theorem 3 of the last section is the gradient of the scalar function $f(x, y, z) = c/r$, which satisfies Laplaces equation $\nabla^2 f = 0$. According to (3) this implies that div $\mathbf{p} = 0$ $(r > 0)$. ■

The following example from hydrodynamics shows the physical significance of the divergence of a vector field. We shall get back to this topic in Sec. 10.8 and add further physical details.

EXAMPLE 2 Flow of a Compressible Fluid. Physical Meaning of the Divergence

We consider the motion of a fluid in a region R having no **sources** or **sinks** in R, that is, no points at which fluid is produced or disappears. The concept of **fluid state** is meant to cover also gases and vapors. Fluids in the restricted sense, or liquids, such as water or oil, have very small compressibility, which can be neglected in many problems. In contrast, gases and vapors have high compressibility. Their density ρ (= mass per unit volume) depends on the coordinates x, y, z in space and may also depend on time t. We assume that our fluid is compressible. We consider the flow through a rectangular box B of small edges $\Delta x, \Delta y, \Delta z$ parallel to the coordinate axes as shown in Fig. 218. (Here Δ is a standard notation for small quantities and, of course, has nothing to do with the notation for the Laplacian in (11) of Sec. 9.7.) The box B has the volume $\Delta V = \Delta x\, \Delta y\, \Delta z$. Let $\mathbf{v} = [v_1, v_2, v_3] = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$ be the velocity vector of the motion. We set

$$\mathbf{u} = \rho\mathbf{v} = [u_1, u_2, u_3] = u_1\mathbf{i} + u_2\mathbf{j} + u_3\mathbf{k} \tag{4}$$

and assume that $\mathbf{u}$ and $\mathbf{v}$ are continuously differentiable vector functions of x, y, z, and t, that is, they have first partial derivatives which are continuous. Let us calculate the change in the mass included in B by considering the **flux** across the boundary, that is, the total loss of mass leaving B per unit time. Consider the flow through the left of the three faces of B that are visible in Fig. 218, whose area is $\Delta x\,\Delta z$. Since the vectors $v_1\mathbf{i}$ and $v_3\,\mathbf{k}$ are parallel to that face, the components v_1 and v_3 of $\mathbf{v}$ contribute nothing to this flow. Hence the mass of fluid entering through that face during a short time interval Δt is given approximately by

$$(\rho v_2)_y\, \Delta x\, \Delta z\, \Delta t = (u_2)_y\, \Delta x\, \Delta z\, \Delta t,$$

where the subscript y indicates that this expression refers to the left face. The mass of fluid leaving the box B through the opposite face during the same time interval is approximately $(u_2)_{y+\Delta y}\, \Delta x\, \Delta z\, \Delta t$, where the subscript $y + \Delta y$ indicates that this expression refers to the right face (which is not visible in Fig. 218). The difference

$$\Delta u_2\, \Delta x\, \Delta z\, \Delta t = \frac{\Delta u_2}{\Delta y}\, \Delta V\, \Delta t \qquad [\Delta u_2 = (u_2)_{y+\Delta y} - (u_2)_y]$$

is the approximate loss of mass. Two similar expressions are obtained by considering the other two pairs of parallel faces of B. If we add these three expressions, we find that the total loss of mass in B during the time interval Δt is approximately

$$\left(\frac{\Delta u_1}{\Delta x} + \frac{\Delta u_2}{\Delta y} + \frac{\Delta u_3}{\Delta z}\right) \Delta V\, \Delta t,$$

where

$$\Delta u_1 = (u_1)_{x+\Delta x} - (u_1)_x \qquad \text{and} \qquad \Delta u_3 = (u_3)_{z+\Delta z} - (u_3)_z.$$

This loss of mass in B is caused by the time rate of change of the density and is thus equal to

$$-\frac{\partial \rho}{\partial t}\, \Delta V\, \Delta t.$$

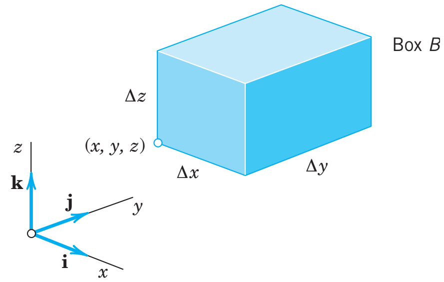


Fig. 218. Physical interpretation of the divergence

If we equate both expressions, divide the resulting equation by $\Delta V \Delta t$, and let Δx, Δy, Δz, and Δt approach zero, then we obtain

$$\operatorname{div} \mathbf{u} = \operatorname{div} (\rho \mathbf{v}) = -\frac{\partial \rho}{\partial t}$$

or

$$\frac{\partial \rho}{\partial t} + \operatorname{div} (\rho \mathbf{v}) = 0. \tag{5}$$

This important relation is called the *condition for the conservation of mass* or the **continuity equation** *of a compressible fluid flow.*

If the flow is **steady**, that is, independent of time, then $\partial \rho / \partial t = 0$ and the continuity equation is

$$\operatorname{div} (\rho \mathbf{v}) = 0. \tag{6}$$

If the density ρ is constant, so that the fluid is incompressible, then equation (6) becomes

$$\operatorname{div} \mathbf{v} = 0. \tag{7}$$

This relation is known as the **condition of incompressibility**. It expresses the fact that the balance of outflow and inflow for a given volume element is zero at any time. Clearly, the assumption that the flow has no sources or sinks in R is essential to our argument. $\mathbf{v}$ is also referred to as **solenoidal**.

From this discussion you should conclude and remember that, roughly speaking, ***the divergence measures outflow minus inflow***. ■

Comment. The **divergence theorem** of Gauss, an integral theorem involving the divergence, follows in the next chapter (Sec. 10.7).

PROBLEM SET 9.8

1–6 CALCULATION OF THE DIVERGENCE

Find div $\mathbf{v}$ and its value at P.

1. $\mathbf{v} = [x^2, 4y^2, 9z^2]$, P: $(-1, 0, \frac{1}{2})$

2. $\mathbf{v} = [0, \cos xyz, \sin xyz]$, P: $(2, \frac{1}{2}\pi, 0]$

3. $\mathbf{v} = (x^2 + y^2)^{-1}[x, y]$

4. $\mathbf{v} = [v_1(y, z), v_2(z, x), v_3(x, y)]$, P: $(3, 1, -1)]$

5. $\mathbf{v} = x^2y^2z^2[x, y, z]$, P: $(3, -1, 4)$

6. $\mathbf{v} = (x^2 + y^2 + z^2)^{-3/2}[x, y, z]$

7. For what v_3 is $\mathbf{v} = [e^x \cos y, e^x \sin y, v_3]$ solenoidal?

8. Let $\mathbf{v} = [x, y, v_3]$. Find a v_3 such that **(a)** div $\mathbf{v} > 0$ everywhere, **(b)** div $\mathbf{v} > 0$ if $|z| < 1$ and div $\mathbf{v} < 0$ if $|z| > 1$.

9. PROJECT. Useful Formulas for the Divergence. Prove

(a) $\operatorname{div}(k\mathbf{v}) = k \operatorname{div} \mathbf{v}$ (k constant)

(b) $\operatorname{div}(f\mathbf{v}) = f \operatorname{div} \mathbf{v} + \mathbf{v} \cdot \nabla f$

(c) $\operatorname{div}(f\nabla g) = f\nabla^2 g + \nabla f \cdot \nabla g$

(d) $\operatorname{div}(f\nabla g) - \operatorname{div}(g\nabla f) = f\nabla^2 g - g\nabla^2 f$

Verify (b) for $f = e^{xyz}$ and $\mathbf{v} = ax\mathbf{i} + by\mathbf{j} + cz\mathbf{k}$. Obtain the answer to Prob. 6 from (b). Verify (c) for $f = x^2 - y^2$ and $g = e^{x+y}$. Give examples of your own for which (a)–(d) are advantageous.

10. CAS EXPERIMENT. Visualizing the Divergence. Graph the given velocity field $\mathbf{v}$ of a fluid flow in a square centered at the origin with sides parallel to the coordinate axes. Recall that the divergence measures outflow minus inflow. By looking at the flow near the sides of the square, can you see whether div $\mathbf{v}$ must be positive or negative or may perhaps be zero? Then calculate div $\mathbf{v}$. First do the given flows and then do some of your own. Enjoy it.

(a) $\mathbf{v} = \mathbf{i}$

(b) $\mathbf{v} = x\mathbf{i}$

(c) $\mathbf{v} = x\mathbf{i} - y\mathbf{j}$

(d) $\mathbf{v} = x\mathbf{i} + y\mathbf{j}$

(e) $\mathbf{v} = -x\mathbf{i} - y\mathbf{j}$

(f) $\mathbf{v} = (x^2 + y^2)^{-1}(-y\mathbf{i} + x\mathbf{j})$

11. Incompressible flow. Show that the flow with velocity vector $\mathbf{v} = y\mathbf{i}$ is incompressible. Show that the particles that at time $t = 0$ are in the cube whose faces are portions of the planes $x = 0, x = 1, y = 0, y = 1, z = 0, z = 1$ occupy at $t = 1$ the volume 1.

12. Compressible flow. Consider the flow with velocity vector $\mathbf{v} = x\mathbf{i}$. Show that the individual particles have the position vectors $\mathbf{r}(t) = c_1 e^t \mathbf{i} + c_2\mathbf{j} + c_3\mathbf{k}$ with constant c_1, c_2, c_3. Show that the particles that at $t = 0$ are in the cube of Prob. 11 at $t = 1$ occupy the volume e.

13. Rotational flow. The velocity vector $\mathbf{v}(x, y, z)$ of an incompressible fluid rotating in a cylindrical vessel is of the form $\mathbf{v} = \mathbf{w} \times \mathbf{r}$, where $\mathbf{w}$ is the (constant) rotation vector; see Example 5 in Sec. 9.3. Show that div $\mathbf{v} = 0$. Is this plausible because of our present Example 2?

14. Does div $\mathbf{u}$ = div $\mathbf{v}$ imply $\mathbf{u} = \mathbf{v}$ or $\mathbf{u} = \mathbf{v} + \mathbf{k}$ ($\mathbf{k}$ constant)? Give reason.

15–20 LAPLACIAN

Calculate $\nabla^2 f$ by Eq. (3). Check by direct differentiation. Indicate when (3) is simpler. Show the details of your work.

15. $f = \cos^2 x + \sin^2 y$

16. $f = e^{xyz}$

17. $f = \ln(x^2 + y^2)$

18. $f = z - \sqrt{x^2 + y^2}$

19. $f = 1/(x^2 + y^2 + z^2)$

20. $f = e^{2x} \cosh 2y$

9.9 Curl of a Vector Field

The concepts of gradient (Sec. 9.7), divergence (Sec. 9.8), and curl are of fundamental importance in vector calculus and frequently applied in vector fields. In this section we define and discuss the concept of the curl and apply it to several engineering problems.

Let $\mathbf{v}(x, y, z) = [v_1, v_2, v_3] = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$ be a differentiable vector function of the Cartesian coordinates x, y, z. Then the **curl** *of the vector function* $\mathbf{v}$ or *of the vector field given by* $\mathbf{v}$ is defined by the “symbolic” determinant

$$\text{(1)} \qquad \operatorname{curl} \mathbf{v} = \nabla \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ v_1 & v_2 & v_3 \end{vmatrix}$$

$$= \left(\frac{\partial v_3}{\partial y} - \frac{\partial v_2}{\partial z}\right)\mathbf{i} + \left(\frac{\partial v_1}{\partial z} - \frac{\partial v_3}{\partial x}\right)\mathbf{j} + \left(\frac{\partial v_2}{\partial x} - \frac{\partial v_1}{\partial y}\right)\mathbf{k}.$$

9. PROJECT. Useful Formulas for the Divergence. Prove

(a) $\operatorname{div}(k\mathbf{v}) = k \operatorname{div} \mathbf{v}$ (k constant)

(b) $\operatorname{div}(f\mathbf{v}) = f \operatorname{div} \mathbf{v} + \mathbf{v} \cdot \nabla f$

(c) $\operatorname{div}(f\nabla g) = f\nabla^2 g + \nabla f \cdot \nabla g$

(d) $\operatorname{div}(f\nabla g) - \operatorname{div}(g\nabla f) = f\nabla^2 g - g\nabla^2 f$

Verify (b) for $f = e^{xyz}$ and $\mathbf{v} = ax\mathbf{i} + by\mathbf{j} + cz\mathbf{k}$. Obtain the answer to Prob. 6 from (b). Verify (c) for $f = x^2 - y^2$ and $g = e^{x+y}$. Give examples of your own for which (a)–(d) are advantageous.

10. CAS EXPERIMENT. Visualizing the Divergence. Graph the given velocity field $\mathbf{v}$ of a fluid flow in a square centered at the origin with sides parallel to the coordinate axes. Recall that the divergence measures outflow minus inflow. By looking at the flow near the sides of the square, can you see whether div $\mathbf{v}$ must be positive or negative or may perhaps be zero? Then calculate div $\mathbf{v}$. First do the given flows and then do some of your own. Enjoy it.

(a) $\mathbf{v} = \mathbf{i}$

(b) $\mathbf{v} = x\mathbf{i}$

(c) $\mathbf{v} = x\mathbf{i} - y\mathbf{j}$

(d) $\mathbf{v} = x\mathbf{i} + y\mathbf{j}$

(e) $\mathbf{v} = -x\mathbf{i} - y\mathbf{j}$

(f) $\mathbf{v} = (x^2 + y^2)^{-1}(-y\mathbf{i} + x\mathbf{j})$

11. Incompressible flow. Show that the flow with velocity vector $\mathbf{v} = y\mathbf{i}$ is incompressible. Show that the particles that at time $t = 0$ are in the cube whose faces are portions of the planes $x = 0, x = 1, y = 0, y = 1, z = 0, z = 1$ occupy at $t = 1$ the volume 1.

12. Compressible flow. Consider the flow with velocity vector $\mathbf{v} = x\mathbf{i}$. Show that the individual particles have the position vectors $\mathbf{r}(t) = c_1e^t\mathbf{i} + c_2\mathbf{j} + c_3\mathbf{k}$ with constant c_1, c_2, c_3. Show that the particles that at $t = 0$ are in the cube of Prob. 11 at $t = 1$ occupy the volume e.

13. Rotational flow. The velocity vector $\mathbf{v}(x, y, z)$ of an incompressible fluid rotating in a cylindrical vessel is of the form $\mathbf{v} = \mathbf{w} \times \mathbf{r}$, where $\mathbf{w}$ is the (constant) rotation vector; see Example 5 in Sec. 9.3. Show that div $\mathbf{v} = 0$. Is this plausible because of our present Example 2?

14. Does $\operatorname{div}\mathbf{u} = \operatorname{div}\mathbf{v}$ imply $\mathbf{u} = \mathbf{v}$ or $\mathbf{u} = \mathbf{v} + \mathbf{k}$ ($\mathbf{k}$ constant)? Give reason.

15–20 LAPLACIAN

Calculate $\nabla^2 f$ by Eq. (3). Check by direct differentiation. Indicate when (3) is simpler. Show the details of your work.

15. $f = \cos^2 x + \sin^2 y$

16. $f = e^{xyz}$

17. $f = \ln(x^2 + y^2)$

18. $f = z - \sqrt{x^2 + y^2}$

19. $f = 1/(x^2 + y^2 + z^2)$

20. $f = e^{2x} \cosh 2y$

9.9 Curl of a Vector Field

The concepts of gradient (Sec. 9.7), divergence (Sec. 9.8), and curl are of fundamental importance in vector calculus and frequently applied in vector fields. In this section we define and discuss the concept of the curl and apply it to several engineering problems.

Let $\mathbf{v}(x, y, z) = [v_1, v_2, v_3] = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}$ be a differentiable vector function of the Cartesian coordinates x, y, z. Then the **curl** *of the vector function* $\mathbf{v}$ or *of the vector field given by* $\mathbf{v}$ is defined by the "symbolic" determinant

$$(1) \qquad \operatorname{curl}\mathbf{v} = \nabla \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ v_1 & v_2 & v_3 \end{vmatrix}$$

$$= \left(\frac{\partial v_3}{\partial y} - \frac{\partial v_2}{\partial z}\right)\mathbf{i} + \left(\frac{\partial v_1}{\partial z} - \frac{\partial v_3}{\partial x}\right)\mathbf{j} + \left(\frac{\partial v_2}{\partial x} - \frac{\partial v_1}{\partial y}\right)\mathbf{k}.$$

This is the formula when x, y, z are *right-handed*. If they are *left-handed*, the determinant has a minus sign in front (just as in (2**) in Sec. 9.3).

Instead of curl $\mathbf{v}$ one also uses the notation rot $\mathbf{v}$. This is suggested by "rotation," an application explored in Example 2. Note that curl $\mathbf{v}$ is a vector, as shown in Theorem 3.

EXAMPLE 1 **Curl of a Vector Function**

Let $\mathbf{v} = [yz, \quad 3zx, \quad z] = yz\mathbf{i} + 3zx\mathbf{j} + z\mathbf{k}$ with right-handed x, y, z. Then (1) gives

$$\operatorname{curl} \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ yz & 3zx & z \end{vmatrix} = -3x\mathbf{i} + y\mathbf{j} + (3z - z)\mathbf{k} = -3x\mathbf{i} + y\mathbf{j} + 2z\mathbf{k}.$$

The curl has many applications. A typical example follows. More about the nature and significance of the curl will be considered in Sec. 10.9.

EXAMPLE 2 **Rotation of a Rigid Body. Relation to the Curl**

We have seen in Example 5, Sec. 9.3, that a rotation of a rigid body B about a fixed axis in space can be described by a vector $\mathbf{w}$ of magnitude ω in the direction of the axis of rotation, where ω (>0) is the angular speed of the rotation, and $\mathbf{w}$ is directed so that the rotation appears clockwise if we look in the direction of $\mathbf{w}$. According to (9), Sec. 9.3, the velocity field of the rotation can be represented in the form

$$\mathbf{v} = \mathbf{w} \times \mathbf{r}$$

where $\mathbf{r}$ is the position vector of a moving point with respect to a Cartesian coordinate system *having the origin on the axis of rotation*. Let us choose right-handed Cartesian coordinates such that the axis of rotation is the z-axis. Then (see Example 2 in Sec. 9.4)

$$\mathbf{w} = [0, \quad 0, \quad \omega] = \omega\mathbf{k}, \qquad \mathbf{v} = \mathbf{w} \times \mathbf{r} = [-\omega y, \quad \omega x, \quad 0] = -\omega y\mathbf{i} + \omega x\mathbf{j}.$$

Hence

$$\operatorname{curl} \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ -\omega y & \omega x & 0 \end{vmatrix} = [0, \quad 0, \quad 2\omega] = 2\omega\mathbf{k} = 2\mathbf{w}.$$

This proves the following theorem.

THEOREM 1 **Rotating Body and Curl**

The curl of the velocity field of a rotating rigid body has the direction of the axis of the rotation, and its magnitude equals twice the angular speed of the rotation.

Next we show how the grad, div, and curl are interrelated, thereby shedding further light on the nature of the curl.

THEOREM 2

Grad, Div, Curl

Gradient fields are ***irrotational***. *That is, if a continuously differentiable vector function is the gradient of a scalar function f, then its curl is the zero vector,*

$$\text{curl}\,(\text{grad}\, f) = \mathbf{0}. \tag{2}$$

Furthermore, the divergence of the curl of a twice continuously differentiable vector function $\mathbf{v}$ *is zero,*

$$\text{div}\,(\text{curl}\ \mathbf{v}) = 0. \tag{3}$$

PROOF Both (2) and (3) follow directly from the definitions by straightforward calculation. In the proof of (3) the six terms cancel in pairs. ■

EXAMPLE 3 **Rotational and Irrotational Fields**

The field in Example 2 is not irrotational. A similar velocity field is obtained by stirring tea or coffee in a cup. The gravitational field in Theorem 3 of Sec. 9.7 has curl $\mathbf{p} = \mathbf{0}$. It is an irrotational gradient field. ■

The term "irrotational" for curl $\mathbf{v} = \mathbf{0}$ is suggested by the use of the curl for characterizing the rotation in a field. If a gradient field occurs elsewhere, not as a velocity field, it is usually called **conservative** (see Sec. 9.7). Relation (3) is plausible because of the interpretation of the curl as a rotation and of the divergence as a flux (see Example 2 in Sec. 9.8).

Finally, since the curl is defined in terms of coordinates, we should do what we did for the gradient in Sec. 9.7, namely, to find out whether the curl is a vector. This is true, as follows.

THEOREM 3

Invariance of the Curl

curl $\mathbf{v}$ is a vector. It has a length and a direction that are independent of the particular choice of a Cartesian coordinate system in space.

PROOF The proof is quite involved and shown in App. 4.

We have completed our discussion of vector differential calculus. The companion Chap. 10 on vector integral calculus follows and makes use of many concepts covered in this chapter, including dot and cross products, parametric representation of curves C, along with grad, div, and curl.

PROBLEM SET 9.9

1. **WRITING REPORT. Grad, div, curl.** List the definitions and most important facts and formulas for grad, div, curl, and ∇^2. Use your list to write a corresponding report of 3–4 pages, with examples of your own. No proofs.
2. **(a)** What direction does curl $\mathbf{v}$ have if $\mathbf{v}$ is parallel to the yz-plane? **(b)** If, moreover, $\mathbf{v}$ is independent of x?
3. Prove Theorem 2. Give two examples for (2) and (3) each.

4–8 CALCULUTION OF CURL

Find curl $\mathbf{v}$ for $\mathbf{v}$ given with respect to right-handed Cartesian coordinates. Show the details of your work.

4. $\mathbf{v} = [2y^2, 5x, 0]$
5. $\mathbf{v} = xyz[x, y, z]$
6. $\mathbf{v} = (x^2 + y^2 + z^2)^{-3/2}[x, y, z]$
7. $\mathbf{v} = [0, 0, e^{-x} \sin y]$
8. $\mathbf{v} = [e^{-z^2}, e^{-x^2}, e^{-y^2}]$

9–13 FLUID FLOW

Let $\mathbf{v}$ be the velocity vector of a steady fluid flow. Is the flow irrotational? Incompressible? Find the streamlines (the paths of the particles). *Hint.* See the answers to Probs. 9 and 11 for a determination of a path.

9. $\mathbf{v} = [0, 3z^2, 0]$
10. $\mathbf{v} = [\sec x, \csc x, 0]$
11. $\mathbf{v} = [y, -2x, 0]$
12. $\mathbf{v} = [-y, x, \pi]$
13. $\mathbf{v} = [x, y, -z]$

14. **PROJECT. Useful Formulas for the Curl.** Assuming sufficient differentiability, show that
    (a) $\operatorname{curl}(\mathbf{u} + \mathbf{v}) = \operatorname{curl}\mathbf{u} + \operatorname{curl}\mathbf{v}$
    (b) $\operatorname{div}(\operatorname{curl}\mathbf{v}) = 0$
    (c) $\operatorname{curl}(f\mathbf{v}) = (\operatorname{grad} f) \times \mathbf{v} + f \operatorname{curl}\mathbf{v}$
    (d) $\operatorname{curl}(\operatorname{grad} f) = \mathbf{0}$
    (e) $\operatorname{div}(\mathbf{u} \times \mathbf{v}) = \mathbf{v} \cdot \operatorname{curl}\mathbf{u} - \mathbf{u} \cdot \operatorname{curl}\mathbf{v}$

15–20 DIV AND CURL

With respect to right-handed coordinates, let $\mathbf{u} = [y, z, x]$, $\mathbf{v} = [yz, zx, xy]$, $f = xyz$, and $g = x + y + z$. Find the given expressions. Check your result by a formula in Proj. 14 if applicable.

15. $\operatorname{curl}(\mathbf{u} + \mathbf{v})$, $\operatorname{curl}\mathbf{v}$
16. $\operatorname{curl}(g\mathbf{v})$
17. $\mathbf{v} \cdot \operatorname{curl}\mathbf{u}$, $\mathbf{u} \cdot \operatorname{curl}\mathbf{v}$, $\mathbf{u} \cdot \operatorname{curl}\mathbf{u}$
18. $\operatorname{div}(\mathbf{u} \times \mathbf{v})$
19. $\operatorname{curl}(g\mathbf{u} + \mathbf{v})$, $\operatorname{curl}(g\mathbf{u})$
20. $\operatorname{div}(\operatorname{grad}(fg))$

CHAPTER 9 REVIEW QUESTIONS AND PROBLEMS

1. What is a vector? A vector function? A vector field? A scalar? A scalar function? A scalar field? Give examples.
2. What is an inner product, a vector product, a scalar triple product? What applications motivate these products?
3. What are right-handed and left-handed coordinates? When is this distinction important?
4. When is a vector product the zero vector? What is orthogonality?
5. How is the derivative of a vector function defined? What is its significance in geometry and mechanics?
6. If $\mathbf{r}(t)$ represents a motion, what are $\mathbf{r}'(t)$, $|\mathbf{r}'(t)|$, $\mathbf{r}''(t)$, and $|\mathbf{r}''(t)|$?
7. Can a moving body have constant speed but variable velocity? Nonzero acceleration?
8. What do you know about directional derivatives? Their relation to the gradient?
9. Write down the definitions and explain the significance of grad, div, and curl.
10. Granted sufficient differentiability, which of the following expressions make sense? $f \operatorname{curl}\mathbf{v}$, $\mathbf{v} \operatorname{curl} f$, $\mathbf{u} \times \mathbf{v}$, $\mathbf{u} \times \mathbf{v} \times \mathbf{w}$, $f \cdot \mathbf{v}$, $f \cdot (\mathbf{v} \times \mathbf{w})$, $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$, $\mathbf{v} \times \operatorname{curl}\mathbf{v}$, $\operatorname{div}(f\mathbf{v})$, $\operatorname{curl}(f\mathbf{v})$, and $\operatorname{curl}(f \cdot \mathbf{v})$.

11–19 ALGEBRAIC OPERATIONS FOR VECTORS

Let $\mathbf{a} = [4, 7, 0]$, $\mathbf{b} = [3, -1, 5]$, $\mathbf{c} = [-6, 2, 0]$, and $\mathbf{d} = [1, -2, 8]$. Calculate the following expressions. Try to make a sketch.

11. $\mathbf{a} \cdot \mathbf{c}$, $3\mathbf{b} \cdot 8\mathbf{d}$, $24\mathbf{d} \cdot \mathbf{b}$, $\mathbf{a} \cdot \mathbf{a}$
12. $\mathbf{a} \times \mathbf{c}$, $\mathbf{b} \times \mathbf{d}$, $\mathbf{d} \times \mathbf{b}$, $\mathbf{a} \times \mathbf{a}$
13. $\mathbf{b} \times \mathbf{c}$, $\mathbf{c} \times \mathbf{b}$, $\mathbf{c} \times \mathbf{c}$, $\mathbf{c} \cdot \mathbf{c}$
14. $5(\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$, $\mathbf{a} \cdot (5\mathbf{b} \times \mathbf{c})$, $(5\mathbf{a}\ \mathbf{b}\ \mathbf{c})$, $5(\mathbf{a} \cdot \mathbf{b}) \times \mathbf{c}$
15. $6(\mathbf{a} \times \mathbf{b}) \times \mathbf{d}$, $\mathbf{a} \times 6(\mathbf{b} \times \mathbf{d})$, $2\mathbf{a} \times 3\mathbf{b} \times \mathbf{d}$
16. $(1/|\mathbf{a}|)\mathbf{a}$, $(1/|\mathbf{b}|)\mathbf{b}$, $\mathbf{a} \cdot \mathbf{b}/|\mathbf{b}|$, $\mathbf{a} \cdot \mathbf{b}/|\mathbf{a}|$
17. $(\mathbf{a}\ \mathbf{b}\ \mathbf{d})$, $(\mathbf{b}\ \mathbf{a}\ \mathbf{d})$, $(\mathbf{b}\ \mathbf{d}\ \mathbf{a})$
18. $|\mathbf{a} + \mathbf{b}|$, $|\mathbf{a}| + |\mathbf{b}|$
19. $\mathbf{a} \times \mathbf{b} - \mathbf{b} \times \mathbf{a}$, $(\mathbf{a} \times \mathbf{c}) \cdot \mathbf{c}$, $|\mathbf{a} \times \mathbf{b}|$
20. **Commutativity.** When is $\mathbf{u} \times \mathbf{v} = \mathbf{v} \times \mathbf{u}$? When is $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$?
21. **Resultant, equilibrium.** Find $\mathbf{u}$ such that $\mathbf{u}$ and $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{d}$ above and $\mathbf{u}$ are in equilibrium.
22. **Resultant.** Find the most general $\mathbf{v}$ such that the resultant of $\mathbf{v}$, $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ (see above) is parallel to the yz-plane.
23. **Angle.** Find the angle between $\mathbf{a}$ and $\mathbf{c}$. Between $\mathbf{b}$ and $\mathbf{d}$. Sketch $\mathbf{a}$ and $\mathbf{c}$.
24. **Planes.** Find the angle between the two planes P_1: $4x - y + 3z = 12$ and P_2: $x + 2y + 4z = 4$. Make a sketch.
25. **Work.** Find the work done by $\mathbf{q} = [5, 2, 0]$ in the displacement from $(1, 1, 0)$ to $(4, 3, 0)$.
26. **Component.** When is the component of a vector $\mathbf{v}$ in the direction of a vector $\mathbf{w}$ equal to the component of $\mathbf{w}$ in the direction of $\mathbf{v}$?
27. **Component.** Find the component of $\mathbf{v} = [4, 7, 0]$ in the direction of $\mathbf{w} = [2, 2, 0]$. Sketch it.

28. Moment. When is the moment of a force equal to zero?

29. Moment. A force $\mathbf{p} = [4, 2, 0]$ is acting in a line through $(2, 3, 0)$. Find its moment vector about the center $(5, 1, 0)$ of a wheel.

30. Velocity, acceleration. Find the velocity, speed, and acceleration of the motion given by $\mathbf{r}(t) = [3 \cos t,\ 3 \sin t,\ 4t]$ $(t = \text{time})$ at the point $P{:}\ (3/\sqrt{2},\ 3/\sqrt{2},\ \pi)$.

31. Tetrahedron. Find the volume if the vertices are $(0, 0, 0)$, $(3, 1, 2)$, $(2, 4, 0)$, $(5, 4, 0)$.

32–40 **GRAD, DIV, CURL, ∇^2, $D_v f$**

Let $f = xy - yz$, $\mathbf{v} = [2y, 2z, 4x + z]$, and $\mathbf{w} = [3z^2, x^2 - y^2, y^2]$. Find:

32. grad f and f grad f at P: $(2, 7, 0)$

33. div $\mathbf{v}$, div $\mathbf{w}$

34. curl $\mathbf{v}$, curl $\mathbf{w}$

35. div (grad f), $\nabla^2 f$, $\nabla^2(xyf)$

36. (curl $\mathbf{w}$) • $\mathbf{v}$ at $(4, 0, 2)$

37. grad (div $\mathbf{w}$)

38. $D_v f$ at P: $(1, 1, 2)$

39. $D_w f$ at P: $(3, 0, 2)$

40. $\mathbf{v} \cdot ((\text{curl } \mathbf{w}) \times \mathbf{v})$

SUMMARY OF CHAPTER 9

Vector Differential Calculus. Grad, Div, Curl

All vectors of the form $\mathbf{a} = [a_1, a_2, a_3] = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ constitute the **real vector space** R^3 with componentwise vector addition

$$[a_1, a_2, a_3] + [b_1, b_2, b_3] = [a_1 + b_1, a_2 + b_2, a_3 + b_3] \tag{1}$$

and componentwise scalar multiplication (c a scalar, a real number)

$$c[a_1, a_2, a_3] = [ca_1, ca_2, ca_3] \qquad \text{(Sec. 9.1).} \tag{2}$$

For instance, the *resultant* of forces $\mathbf{a}$ and $\mathbf{b}$ is the sum $\mathbf{a} + \mathbf{b}$.

The **inner product** or **dot product** of two vectors is defined by

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos \gamma = a_1b_1 + a_2b_2 + a_3b_3 \qquad \text{(Sec. 9.2)} \tag{3}$$

where γ is the angle between $\mathbf{a}$ and $\mathbf{b}$. This gives for the **norm** or **length** $|\mathbf{a}|$ of $\mathbf{a}$

$$|\mathbf{a}| = \sqrt{\mathbf{a} \cdot \mathbf{a}} = \sqrt{a_1^2 + a_2^2 + a_3^2} \tag{4}$$

as well as a formula for γ. If $\mathbf{a} \cdot \mathbf{b} = 0$, we call $\mathbf{a}$ and $\mathbf{b}$ **orthogonal**. The dot product is suggested by the *work* $W = \mathbf{p} \cdot \mathbf{d}$ done by a force $\mathbf{p}$ in a displacement $\mathbf{d}$.

The **vector product** or **cross product** $\mathbf{v} = \mathbf{a} \times \mathbf{b}$ is a vector of length

$$|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}||\mathbf{b}| \sin \gamma \qquad \text{(Sec. 9.3)} \tag{5}$$

and perpendicular to both $\mathbf{a}$ and $\mathbf{b}$ such that $\mathbf{a}$, $\mathbf{b}$, $\mathbf{v}$ form a *right-handed* triple. In terms of components with respect to right-handed coordinates,

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \qquad \text{(Sec. 9.3).} \tag{6}$$

The vector product is suggested, for instance, by moments of forces or by rotations. CAUTION! This multiplication is *anti*commutative, $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$, and is *not* associative.

An (oblique) box with edges $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ has volume equal to the absolute value of the **scalar triple product**

$$(\mathbf{a} \quad \mathbf{b} \quad \mathbf{c}) = \mathbf{a} \bullet (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) \bullet \mathbf{c}. \tag{7}$$

Sections 9.4–9.9 extend differential calculus to vector functions

$$\mathbf{v}(t) = [v_1(t), v_2(t), v_3(t)] = v_1(t)\mathbf{i} + v_2(t)\mathbf{j} + v_3(t)\mathbf{k}$$

and to vector functions of more than one variable (see below). The derivative of $\mathbf{v}(t)$ is

$$\mathbf{v}' = \frac{d\mathbf{v}}{dt} = \lim_{\Delta t \to 0} \frac{\mathbf{v}(t + \Delta t) - \mathbf{v}(t)}{\Delta t} = [v_1', v_2', v_3'] = v_1'\mathbf{i} + v_2'\mathbf{j} + v_3'\mathbf{k}. \tag{8}$$

Differentiation rules are as in calculus. They imply (Sec. 9.4)

$$(\mathbf{u} \bullet \mathbf{v})' = \mathbf{u}' \bullet \mathbf{v} + \mathbf{u} \bullet \mathbf{v}', \qquad (\mathbf{u} \times \mathbf{v})' = \mathbf{u}' \times \mathbf{v} + \mathbf{u} \times \mathbf{v}'.$$

Curves C in space represented by the position vector $\mathbf{r}(t)$ have $\mathbf{r}'(t)$ as a **tangent vector** (the **velocity** in mechanics when t is time), $\mathbf{r}'(s)$ (s arc length, Sec. 9.5) as the *unit tangent vector,* and $|\mathbf{r}''(s)| = \kappa$ as the *curvature* (the *acceleration* in mechanics).

Vector functions $\mathbf{v}(x, y, z) = [v_1(x, y, z), v_2(x, y, z), v_3(x, y, z)]$ represent vector fields in space. Partial derivatives with respect to the Cartesian coordinates x, y, z are obtained componentwise, for instance,

$$\frac{\partial \mathbf{v}}{\partial x} = \left[\frac{\partial v_1}{\partial x}, \frac{\partial v_2}{\partial x}, \frac{\partial v_3}{\partial x}\right] = \frac{\partial v_1}{\partial x}\mathbf{i} + \frac{\partial v_2}{\partial x}\mathbf{j} + \frac{\partial v_3}{\partial x}\mathbf{k} \qquad \text{(Sec. 9.6).}$$

The **gradient** of a scalar function f is

$$\operatorname{grad} f = \nabla f = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right] \qquad \text{(Sec. 9.7).} \tag{9}$$

The **directional derivative** of f in the direction of a vector $\mathbf{a}$ is

$$D_{\mathbf{a}} f = \frac{df}{ds} = \frac{1}{|\mathbf{a}|} \mathbf{a} \bullet \nabla f \qquad \text{(Sec. 9.7).} \tag{10}$$

The **divergence** of a vector function $\mathbf{v}$ is

$$\operatorname{div} \mathbf{v} = \nabla \bullet \mathbf{v} = \frac{\partial v_1}{\partial x} + \frac{\partial v_2}{\partial y} + \frac{\partial v_3}{\partial z}. \qquad \text{(Sec. 9.8).} \tag{11}$$

The **curl** of **v** is

(12) $$\operatorname{curl} \mathbf{v} = \nabla \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ v_1 & v_2 & v_3 \end{vmatrix}$$ (Sec. 9.9)

or minus the determinant if the coordinates are left-handed.

Some basic formulas for grad, div, curl are (Secs. 9.7–9.9)

(13) $$\nabla(fg) = f\nabla g + g\nabla f$$
$$\nabla(f/g) = (1/g^2)(g\nabla f - f\nabla g)$$

(14) $$\operatorname{div}(f\mathbf{v}) = f \operatorname{div} \mathbf{v} + \mathbf{v} \cdot \nabla f$$
$$\operatorname{div}(f\nabla g) = f\nabla^2 g + \nabla f \cdot \nabla g$$

(15) $$\nabla^2 f = \operatorname{div}(\nabla f)$$
$$\nabla^2(fg) = g\nabla^2 f + 2\nabla f \cdot \nabla g + f\nabla^2 g$$

(16) $$\operatorname{curl}(f\mathbf{v}) = \nabla f \times \mathbf{v} + f \operatorname{curl} \mathbf{v}$$
$$\operatorname{div}(\mathbf{u} \times \mathbf{v}) = \mathbf{v} \cdot \operatorname{curl} \mathbf{u} - \mathbf{u} \cdot \operatorname{curl} \mathbf{v}$$

(17) $$\operatorname{curl}(\nabla f) = \mathbf{0}$$
$$\operatorname{div}(\operatorname{curl} \mathbf{v}) = 0.$$

For grad, div, curl, and ∇^2 in **curvilinear coordinates** see App. A3.4.