Chapter 8

Systems of Linear Algebraic Equations; Gauss Elimination

8.1 Introduction

There are many applications in science and engineering where application of the relevant physical law(s) immediately produces a set of linear algebraic equations. For instance, the application of Kirchoff’s laws to a DC electrical circuit containing any number of resistors, batteries, and current loops immediately produces such a set of equations on the unknown currents. In other cases, the problem is stated in some other form such as one or more ordinary or partial differential equations, but the solution method eventually leads us to a system of linear algebraic equations. For instance, to find a particular solution to the differential equation

\[ y'' - y'' = 3x^2 + 5 \sin x \]  

(1)

by the method of undetermined coefficients (Section 3.7.2), we seek it in the form

\[ y_p(x) = Ax^4 + Bx^3 + Cx^2 + Dx + E \cos x. \]  

(2)

Putting (2) into (1) and equating coefficients of like terms on both sides of the equation gives five linear algebraic equations on the unknown coefficients \( A, B, \ldots, E \). Or, solving the so-called Laplace partial differential equation

\[ \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \]  

(3)

on the rectangle \( 0 < x < 1, 0 < y < 1 \) by the method of finite differences (which is studied in Section 20.5), using a mesh size \( \Delta x = \Delta y = 0.05 \), gives 19\(^2 = 361\) linear algebraic equations on the unknown values of \( u \) at the 361 nodal points of the mesh. Our point here is not to get ahead of ourselves by plunging into partial differential equations, but to say that the solution of practical problems of interest in science and engineering often leads us to systems of linear algebraic equations.
Such systems often involve a great many unknowns. Thus, the question of existence (Does a solution exist?), which often sounds "too theoretical" to the practicing engineer, takes on great practical importance because a considerable computational effort is at stake.

The subject of linear algebra and matrices encompasses a great deal more than the theory and solution of systems of linear algebraic equations, but the latter is indeed a central topic and is foundational to others. Thus, we begin this sequence of five chapters (8–12) on linear algebra with an introduction to the theory of systems of linear algebraic equations, and their solution by the method of Gauss elimination. Results obtained here are used, and built upon, in Chapters 9–12.

Chapters 9 and 10 take us from vectors in 3-space to vectors in \( n \)-space and generalized vector space, to matrices and determinants. Linear systems of algebraic equations are considered again, in the second half of Chapter 10, in terms of rank, inverse matrix, LU decomposition, Cramer's rule, and linear transformation. Chapter 11 introduces the eigenvalue problem, diagonalization, and quadratic forms; areas of application include systems of ordinary differential equations, vibration theory, chemical kinetics, and buckling. Chapter 12 is optional and brief and provides an extension of results in Chapters 9–11 to complex spaces.

## 8.2 Preliminary Ideas and Geometrical Approach

The problem of finding solutions of equations of the form

\[
f(x) = 0
\]  

(1)

occupies a place of both practical and historical importance. Equation (1) is said to be algebraic, or polynomial, if \( f(x) \) is expressible in the form \( a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 \), where \( a_n \neq 0 \) for definiteness [i.e., if \( f(x) \) is a polynomial of finite degree \( n \)], and it is said to be transcendental otherwise.

**EXAMPLE 1.** The equations \( 6x - 5 = 0 \) and \( 3x^4 - x^3 + 2x + 1 = 0 \) are algebraic, whereas \( x^3 + 2 \sin x = 0 \) and \( e^x - 3 = 0 \) are transcendental since \( \sin x \) and \( e^x \) cannot be expressed as polynomials of finite degree.

Besides the algebraic versus transcendental distinction, we classify (1) as linear if \( f(x) \) is a first-degree polynomial,

\[
a_1 x + a_0 = 0,
\]  

(2)

and nonlinear otherwise. Thus, the first equation in Example 1 is linear, and the other three are nonlinear.

While (1) is one equation in one unknown, we often encounter problems involving more than one equation and/or more than one unknown — that is, a system
of equations consisting of \( m \) equations in \( n \) unknowns, where \( m \geq 1 \) and \( n \geq 1 \),

\[
\begin{align*}
f_1(x_1, \ldots, x_n) &= 0, \\
f_2(x_1, \ldots, x_n) &= 0, \\
& \vdots \\
f_m(x_1, \ldots, x_n) &= 0
\end{align*}
\]

such as

\[
\begin{align*}
x_1 - \sin (x_1 + 7x_2) &= 0, \\
x_2^2 + x_2 - 5x_1 + 6 &= 0.
\end{align*}
\]

In (4) it happens that \( m = n \) (namely, \( m = n = 2 \)) so that there are as many equations as unknowns. In general, however, \( m \) may be less than, equal to, or greater than \( n \) so we allow for \( m \neq n \) in this discussion even though \( m = n \) is the most important case.

In this chapter we consider only the case where (3) is linear, of the form

\[
\begin{align*}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= c_1, \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= c_2, \\
& \vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= c_m,
\end{align*}
\]

and we restrict \( m \) and \( n \) to be finite, and the \( a_{ij} \)'s and \( c_j \)'s to be real numbers. If all the \( c_j \)'s are zero then (5) is homogeneous; if they are not all zero then (5) is nonhomogeneous.

The subscript notation adopted in (5) is not essential but is helpful in holding the nomenclature to a minimum, in rendering inherent patterns more visible, and in permitting a natural transition to matrix notation. The first subscript in \( a_{ij} \) indicates the equation, and the second indicates the \( x_j \) variable that it multiplies. For instance, \( a_{21} \) appears in the second equation and multiplies the \( x_1 \) variable. To avoid ambiguity we should write \( a_{2,1} \) rather than \( a_{21} \) so that one does not mistakenly read the subscripts as twenty-one, but we will omit commas except when such ambiguity is not easily resolved from the context.

We say that a sequence of numbers \( s_1, s_2, \ldots, s_n \) is a solution of (5) if and only if each of the \( m \) equations is satisfied numerically when we substitute \( s_1 \) for \( x_1 \), \( s_2 \) for \( x_2 \), and so on. If there exist one or more solutions to (5), we say that the system is consistent; if there is precisely one solution, that solution is unique; and if there is more than one, the solution is nonunique. If, on the other hand, there are no solutions to (5), the system is said to be inconsistent. The collection of all solutions to (5) is called its solution set so, by “solving (5)” we mean finding its solution set.

Let us begin with the simple case, where \( m = n = 1 \):

\[
a_{11}x_1 = c_1.
\]
In the generic case, \( a_{11} \neq 0 \) and (6) admits the unique solution \( x_1 = c_1/a_{11} \), but if \( a_{11} = 0 \) there are two possibilities: if \( c_1 \neq 0 \) then there are no values of \( x_1 \) such that \( 0x_1 = c_1 \) and (6) is inconsistent, but if \( c_1 = 0 \) then (6) becomes \( 0x_1 = 0 \), and \( x_1 = 0 \) is a solution for any value of \( a_1 \); that is, the solution is nonunique.

Far from being too simple to be of interest, the case where \( m = n = 1 \) establishes a pattern that will hold in general, for any values of \( m \) and \( n \). Specifically, the system (5) will admit a unique solution, no solution, or an infinity of solutions. For instance, it will never admit 4 solutions, 12 solutions, or 137 solutions.

Next, consider the case where \( m = n = 2 \):

\[
\begin{align*}
a_{11}x_1 + a_{12}x_2 &= c_1, \quad (\text{eq.1}) \\
a_{21}x_1 + a_{22}x_2 &= c_2. \quad (\text{eq.2})
\end{align*}
\]

If \( a_{11} \) and \( a_{12} \) are not both zero, then (eq.1) defines a straight line, say \( L_1 \), in a Cartesian \( x_1, x_2 \) plane; that is the solution set of (eq.1) is the set of all points on that line. Similarly, if \( a_{21} \) and \( a_{22} \) are not both zero then the solution set of (eq.2) is the set of all points on a straight line \( L_2 \). There exist exactly three possibilities, and these are illustrated in Fig. 1. First, the lines may intersect at a point, say \( P \), in which case (7) admits the unique solution given by the coordinate pair \( x_1, x_2 \) of the point \( P \) (Fig. 1a). That is, any solution pair \( x_1, x_2 \) of (7) needs to be in the solution set of (eq.1) and in the solution set of (eq.2) hence at an intersection of \( L_1 \) and \( L_2 \). This is the generic case, and it occurs (Exercise 2) as long as

\[
a_{11}a_{22} - a_{12}a_{21} \neq 0; \quad (8)
\]

is the analog of the \( a_{11} \neq 0 \) condition for the \( m = n = 1 \) case discussed above.

Second, the lines may be parallel and nonintersecting (Fig. 1b), in which case there is no solution. Then (7) is inconsistent, the solution set is empty.

Third, the lines may coincide (Fig. 1c), in which case the coordinate pair of each point on the line is a solution. Then (7) is consistent and there are an infinite number of solutions.

**EXAMPLE 2.**

\[
\begin{align*}
2x_1 - x_2 &= 5, \\
x_1 + 3x_2 &= 1, \\
x_1 + 3x_2 &= 1, \\
x_1 + 3x_2 &= 1, \\
x_1 + 3x_2 &= 0, \\
x_1 + 3x_2 &= 0, \\
2x_1 + 6x_2 &= 2,
\end{align*}
\]

illustrate these three cases, respectively. 

Below (7) we said "If \( a_{11} \) and \( a_{12} \) are not both zero …" What if they are both zero? Then if \( c_1 \neq 0 \) there is no solution of (7a), and hence there is no solution to the system (7). But if \( c_1 = 0 \), then (7a) reduces to \( 0x_1 = 0 \) and can be discarded, leaving just (7b). If \( a_{21} \) and \( a_{22} \) are not both zero, then (7b) gives a line of solutions, but if they are both zero then everything hinges on \( c_2 \). If \( c_2 \neq 0 \) there is no solution and (7) is inconsistent, but if \( c_2 = 0 \), so both (7a) and (7b) are simply \( 0x_1 = 0 \), then both \( x_1 \) and \( x_2 \) are arbitrary, and every point in the plane is a solution.
Next, consider the case where \( m = n = 3 \):

\[
\begin{align*}
    a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= c_1, \quad (eq.1) \\
    a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= c_2, \quad (eq.2) \\
    a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= c_3, \quad (eq.3)
\end{align*}
\]

Continuing the geometric approach exemplified by Fig. 1, observe that if \( a_{11}, a_{12}, a_{13} \) are not all zero then (eq.1) defines a plane, say \( P_1 \), in Cartesian \( x_1, x_2, x_3 \) space, and similarly for (eq.2) and (eq.3). In the generic case, \( P_1 \) and \( P_2 \) intersect along a line \( L \), and \( L \) pierces \( P_3 \) at a point \( P \). Then the \( x_1, x_2, x_3 \) coordinates of \( P \) give the unique solution of (9).

In the nongeneric case we can have no solution or an infinity of solutions in the following ways. There will be no solution if \( L \) is parallel to \( P_3 \) and hence fails to pierce it, or if any two of the planes are parallel and not coincident. There will be an infinity of solutions if \( L \) lies in \( P_3 \) (i.e., a line of solutions), if two planes are coincident and intersect the third (again, a line of solutions), or if all three planes are coincident (this time an entire plane of solutions).

The case where all of the \( a_{ij} \) coefficients are zero in one or more of equations (9) is left for the exercises.

An abstract extension of such geometrical reasoning can be continued even if \( m = n \geq 4 \). For instance, one speaks of \( a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + a_{14}x_4 = c_1 \) as defining a hyperplane in an abstract four-dimensional space. In fact, perhaps we should mention that even the familiar \( x_1, x_2 \) plane and \( x_1, x_2, x_3 \) space discussed here could be abstract as well. For instance, if \( x_1 \) and \( x_2 \) are unknown currents in two loops of an electrical circuit, then what physical meaning is there to an \( x_1, x_2 \) plane? None, but we can introduce it, create it, to assist our reasoning.

**Closure.** Most of this section is devoted to a geometrical discussion of the system (5) of linear algebraic equations. A great advantage of geometrical reasoning is that it brings our visual system into play. It is estimated that at least a third of the neurons in our brain are devoted to vision, hence our visual sense is extremely sophisticated. No wonder we say "Now I see what you mean; now I get the picture." The more geometry, pictures, visual images to aid our thinking, the better! We have not yet aimed at theorems, and have been content to lay the groundwork for the ideas of existence and uniqueness of solutions. In considering the cases where \( m = n = 1, m = n = 2, \) and \( m = n = 3 \), we have not meant to imply that we need to have \( m = n; all \) possibilities are considered in the next section. To proceed further, we need to consider the process of finding solutions, and that we do, in Section 8.3, by the method of Gauss elimination.
Chapter 8. Systems of Linear Algebraic Equations; Gauss Elimination

EXERCISES 8.2

1. True or false? If false, give a counterexample.
   (a) An algebraic equation is necessarily linear.
   (b) An algebraic equation is necessarily nonlinear.
   (c) A transcendental equation is necessarily linear.
   (d) A transcendental equation is necessarily nonlinear.
   (e) A linear equation is necessarily algebraic.
   (f) A nonlinear equation is necessarily algebraic.
   (g) A linear equation is necessarily transcendental.
   (h) A nonlinear equation is necessarily transcendental.

2. Derive the condition (8) as the necessary and sufficient condition for (7) to admit a unique solution.

3. (a) Discuss all possibilities of the existence and uniqueness of solutions of (9) from a geometrical point of view, in the event that \( a_{11} = a_{12} = a_{13} = 0 \), but \( a_{21}, a_{22}, a_{23} \) and \( a_{31}, a_{32}, a_{33} \) are not all zero.
   (b) Same as (a), but with \( a_{21} = a_{22} = a_{23} = 0 \) as well.
   (c) Same as (a), but with \( a_{21} = a_{22} = a_{23} = a_{31} = a_{32} = a_{33} = 0 \) as well.

8.3 Solution by Gauss Elimination

8.3.1. Motivation. In this section we continue to consider the system of \( m \) linear algebraic equations

\[
\begin{align*}
    a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= c_1, \\
    a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= c_2, \\
    & \vdots \\
    a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= c_m,
\end{align*}
\]

in the \( n \) unknowns \( x_1, \ldots, x_n \), and develop the solution technique known as Gauss elimination. To motivate the ideas, we begin with an example.

EXAMPLE 1. Determine the solution set of the system

\[
\begin{align*}
    x_1 + x_2 - x_3 &= 1, \\
    3x_1 + x_2 + x_3 &= 9, \\
    x_1 - x_2 + 4x_3 &= 8.
\end{align*}
\]

Keep the first equation intact, and add \(-3\) times the first equation to the second (as a replacement for the second equation), and add \(-1\) times the first equation to the third (as a replacement for the third equation). These steps yield the new "Indented" system

\[
\begin{align*}
    x_1 + x_2 - x_3 &= 1, \\
    -2x_2 + 4x_3 &= 6, \\
    -2x_2 + 5x_3 &= 7.
\end{align*}
\]

Next, keep the first two of these intact, and add \(-1\) times the second equation to the third, and obtain

\[
\begin{align*}
    x_1 + x_2 - x_3 &= 1, \\
    -2x_2 + 4x_3 &= 6, \\
    x_3 &= 1.
\end{align*}
\]
Finally, multiplying the second of these by $-1/2$ to normalize the leading coefficient (to unity), gives

\[
\begin{align*}
x_1 + x_2 - x_3 &= 1, \quad (eq.1) \\
x_2 - 2x_3 &= -3, \quad (eq.2) \\
x_3 &= 1. \quad (eq.3)
\end{align*}
\] (5)

It is helpful to think of the original system (2) as a tangle of string that we wish to unravel. The first step is to find a loose end and that is, in effect, what the foregoing process of successive indentations has done for us. Specifically, (eq.3) in (5) is the “loose end,” and with that in hand we may unravel (5) just as we would unravel a tangle: putting $x_3 = 1$ into (eq.2) gives $x_2 = -1$, and then putting $x_3 = 1$ and $x_2 = -1$ into (eq.1) gives $x_1 = 3$. Thus, we obtain the unique solution

\[
x_3 = 1, \quad x_2 = -1, \quad x_1 = 3.
\] (6)

COMMENT 1. From a mathematical point of view, the system (2) was a “tangle” because the equations were coupled: that is, each equation contained more than one unknown. Actually, the final system (5) is coupled too, since (eq.1) contains all three unknowns and (eq.2) contains two of them. However, the coupling in (5) is not as debilitating because (5) is in what we call triangular form. Thus, we were able to solve (eq.3) for $x_3$, put that value into (eq.2) and solve for $x_2$, and then put these values into (eq.1) and solve for $x_1$, which steps are known as back substitution.

COMMENT 2. However, the process begs this question: Is it obvious that the systems (2)–(5) all have the same solution sets so that when we solve (5) we are actually solving (2)? That is, is it not conceivable that in applying the arithmetic steps that carried us from (2) to (5) we might, inadvertently, have altered the solution set? For example, $x - 1 = 4$ has the unique solution $x = 5$, but if we innocently square both sides, the resulting equation $(x - 1)^2 = 16$ admits the two solutions $x = 5$ and $x = -3$.

The question just raised applies to linear systems in general. It is answered in Theorem 8.3.1 that follows, but first we define two terms: “equivalent systems” and “elementary equation operations.”

Two linear systems in $n$ unknowns, $x_1$ through $x_n$, are said to be equivalent if their solution sets are identical.

The following operations on linear systems are known as elementary equation operations:

1. Addition of a multiple of one equation to another
   Symbolically: $(eq.j) \to (eq.j) + \alpha (eq.k)$

2. Multiplication of an equation by a nonzero constant
   Symbolically: $(eq.j) \to \alpha (eq.j)$

3. Interchange of two equations
   Symbolically: $(eq.j) \leftrightarrow (eq.k)$

Then we can state the following result.
THEOREM 8.3.1 Equivalent Systems
If one linear system is obtained from another by a finite number of elementary
equation operations, then the two systems are equivalent.

Outline of Proof: The truth of this claim for elementary equation operations of
types 2 and 3 should be evident, so we confine our remarks to operations of type
1. It suffices to look at the effect of one such operation. Thus, suppose that a given
linear system $A$ is altered by replacing its $j$th equation by its $j$th plus $\alpha$ times its
$k$th, its other equations being kept intact. Let us call the new system $A'$. Surely,
every solution of $A$ will also be a solution $A'$ since we have merely added equal
quantities to equal quantities. That is, if $A'$ results from $A$ by the application of an
elementary equation operation of type 1, then every solution of $A$ is also a solution
of $A'$. Further, we can convert $A'$ back to $A$ by an elementary equation operation of
type 1, namely, by replacing the $j$th equation of $A'$ by the $j$th equation of $A'$ plus
$-\alpha$ times the $k$th equation of $A'$. Consequently, it follows from the italicized result
two sentences back) that every solution of $A'$ is also a solution of $A$. Then $A$ and
$A'$ are equivalent, as claimed. ■

In Example 1, we saw that each step is an elementary equation operation:
Three elementary equation operations of type 1 took us from (2) to (4), and one of
type 2 took us from (4) to (5); finally, the back substitution amounted to several
operations of type 1. Thus, according to Theorem 8.3.1, equivalence was maintained
throughout so we can be sure that (6) is the solution set of the original system (2)
(as can be verified by direct substitution).

The system in Example 1 admitted a unique solution. To see how the method
of successive elimination works out when there is no solution, or a nonunique
solution, let us work two more examples.

EXAMPLE 2. Inconsistent System. Consider the system

$$
\begin{align*}
2x_1 + 3x_2 - 2x_3 &= 4, \\
x_1 - 2x_2 + x_3 &= 3, \\
7x_1 &- x_3 = 2.
\end{align*}
$$

(7)

Keep the first equation intact, add $-\frac{1}{2}$ times the first equation to the second (eq.2 \rightarrow eq.2
$-\frac{1}{2}$ eq.1), and add $-\frac{5}{2}$ times the first to the third (eq.3 \rightarrow eq.3 $-\frac{5}{2}$ eq.1):

$$
\begin{align*}
2x_1 + 3x_2 - 2x_3 &= 4, \\
\quad - \frac{7}{2}x_2 + 2x_3 &= 1, \\
\quad - \frac{21}{2}x_2 + 6x_3 &= -12.
\end{align*}
$$

(8)

Keep the first two equations intact, and add $-3$ times the second equation to the third (eq.3

→ eq.3 - eq.2):

\[
\begin{align*}
2x_1 + 3x_2 - 2x_3 &= 4, \\
-\frac{7}{2}x_2 + 2x_3 &= 1, \\
0 &= -15.
\end{align*}
\]  

(9)

Any solution \(x_1, x_2, x_3\) of (9) must satisfy each of the three equations, but there are no values of \(x_1, x_2, x_3\) that can satisfy \(0 = -15\). Thus, (9) is inconsistent (has no solution), and therefore (7) is as well.

COMMENT. The source of the inconsistency is the fact that whereas the left-hand side of the third equation is 2 times the left-hand side of the first equation plus 3 times the left-hand side of the second, the right-hand sides do not bear that relationship: \(2(4) + 3(3) = 17 \neq 2\). [While that built-in contradiction is not obvious from (7), it eventually comes to light in the third equation in (9).] If we modify the system (7) by changing the final 2 in (7) to 17, then the final -12 in (8) becomes a 3, and the final -15 in (9) becomes a zero:

\[
\begin{align*}
2x_1 + 3x_2 - 2x_3 &= 4, \\
-\frac{7}{2}x_2 + 2x_3 &= 1, \\
0 &= 0
\end{align*}
\]  

(10)

or, multiplying the first by \(\frac{1}{4}\) and the second by \(-\frac{3}{7}\),

\[
\begin{align*}
x_1 + \frac{3}{2}x_2 - x_3 &= 2, \\
x_2 - \frac{4}{7}x_3 &= -\frac{2}{7},
\end{align*}
\]  

(11a,b)

where we have discarded the identity \(0 = 0\). Thus, by changing the \(c_i\)'s so as to be "compatible," the system now admits an infinity of solutions rather than none. Specifically, we can let \(x_3\) (or \(x_2\), it doesn't matter which) in (11b) be any value, say \(\alpha\), where \(\alpha\) is arbitrary. Then (11b) gives \(x_2 = -\frac{2}{7} + \frac{4}{7}\alpha\), and putting these into (11a), \(x_1 = \frac{17}{7} + \frac{1}{7}\alpha\). Thus, we have the infinity of solutions

\[
x_3 = \alpha, \quad x_2 = -\frac{2}{7} + \frac{4}{7}\alpha, \quad x_1 = \frac{17}{7} + \frac{1}{7}\alpha
\]  

(12)

for any \(\alpha\). Evidently, two of the three planes intersect, giving a line that lies in the third plane, and equations (12) are parametric equations of that line!  

EXAMPLE 3. *Nonunique Solution.* Consider the system of four equations in six unknowns \((m = 4, n = 6)\)

\[
\begin{align*}
2x_2 + x_3 + 4x_4 + 3x_5 + x_6 &= 2, \\
x_1 - x_2 + x_3 + 2x_6 &= 0, \\
x_1 + x_2 + 2x_3 + 4x_4 + x_5 + 2x_6 &= 3, \\
x_1 - 3x_2 - 4x_4 - 2x_5 + x_6 &= 0.
\end{align*}
\]  

(13)

Wanting the top equation to begin with \(x_1\) and subsequent equations to indent at the left,
let us first move the top equation to the bottom (eq.1 $\leftrightarrow$ eq.4):

$$
\begin{align*}
x_1 - 3x_2 & \quad - 4x_4 - 2x_5 + x_6 = 0, \\
x_1 - x_2 + x_3 & \quad + 2x_6 = 0, \\
x_1 + x_2 + 2x_3 + 4x_4 + x_5 + 2x_6 = 3, \\
2x_2 + x_3 + 4x_4 + 3x_5 + x_6 = 2.
\end{align*}
$$

(14)

Add $-1$ times the first equation to the second (eq.2 $\rightarrow$ eq.2 $-1$ eq.1) and third (eq.3 $\rightarrow$ eq.3 $-1$ eq.1) equations:

$$
\begin{align*}
x_1 - 3x_2 & \quad - 4x_4 - 2x_5 + x_6 = 0, \\
2x_2 + x_3 + 4x_4 + 2x_5 + x_6 = 0, \\
4x_2 + 2x_3 + 8x_4 + 3x_5 + x_6 = 3, \\
2x_2 + x_3 + 4x_4 + 3x_5 + x_6 = 2.
\end{align*}
$$

(15)

Add $-2$ times the second to the third (eq.3 $\rightarrow$ eq.3 $-2$ eq.2) and $-1$ times the second to the fourth (eq.4 $\rightarrow$ eq.4 $-1$ eq.2):

$$
\begin{align*}
x_1 - 3x_2 & \quad - 4x_4 - 2x_5 + x_6 = 0, \\
2x_2 + x_3 + 4x_4 + 2x_5 + x_6 = 0, \\
-x_5 - x_6 = 3, \\
x_5 = 2.
\end{align*}
$$

(16)

Add the third to the fourth (eq.4 $\rightarrow$ eq.4 + eq.3):

$$
\begin{align*}
x_1 - 3x_2 & \quad - 4x_4 - 2x_5 + x_6 = 0, \\
2x_2 + x_3 + 4x_4 + 2x_5 + x_6 = 0, \\
-x_5 - x_6 = 3, \\
-x_5 = 5.
\end{align*}
$$

(17)

Finally, multiply the second, third, and fourth by $\frac{1}{2}$, $-1$, and $-1$, respectively, to normalize the leading coefficients (eq.2 $\rightarrow$ $\frac{1}{2}$ eq.2, eq.3 $\rightarrow$ $-1$ eq.3, eq.4 $\rightarrow$ $-1$ eq.4):

$$
\begin{align*}
x_1 - 3x_2 & \quad - 4x_4 - 2x_5 + x_6 = 0, \\
x_2 + \frac{1}{2}x_3 + 2x_4 + x_5 + \frac{1}{2}x_6 = 0, \\
x_5 + \frac{1}{2}x_6 = -3, \\
x_6 = -5.
\end{align*}
$$

(18)

The last two equations give $x_6 = -5$ and $x_5 = 2$, and these values can be substituted back into the second equation. In that equation we can let $x_4$ be arbitrary, say $\alpha_1$, and we can also let $x_3$ be arbitrary, say $\alpha_2$. Then that equation gives $x_2$ and, again by back substitution, the first equation gives $x_1$. The result is the infinity of solutions

$$
\begin{align*}
x_6 = -5, \quad x_5 = 2, \quad x_4 = \alpha_1, \quad x_3 = \alpha_2, \\
x_2 = \frac{1}{2} - 2\alpha_1 - \frac{1}{2}\alpha_2, \quad x_1 = \frac{21}{2} - 2\alpha_1 - \frac{3}{2}\alpha_2,
\end{align*}
$$

(19)

where $\alpha_1$ and $\alpha_2$ are arbitrary.

If a solution set contains $p$ independent arbitrary parameters ($\alpha_1, \ldots, \alpha_p$), we call it (in this text) a $p$-parameter family of solutions. Thus, (12) and (19) are
one- and two-parameter families of solutions, respectively. Each choice of values for $\alpha_1, \ldots, \alpha_p$ yields a particular solution. In (19), for instance, the choice $\alpha_1 = 1$ and $\alpha_2 = 0$ yields the particular solution $x_1 = \frac{17}{2}, x_2 = -\frac{3}{2}, x_3 = 0, x_4 = 1, x_5 = 2$, and $x_6 = -5$.

8.3.2. Gauss elimination. The method of Gauss elimination,* illustrated in Examples 1–3, can be applied to any linear system (1), whether or not the system is consistent, and whether or not the solution is unique. Though hard to tell from the foregoing hard calculations, the method is efficient and is commonly available in computer systems.

Observe that the end result of the Gauss elimination process enables us to determine, merely from the pattern of the final equations, whether or not a solution exists and is unique. For instance, we can see from the pattern of (5) that there is a unique solution, from the bottom equation in (9) that there no solution, and from the extra double indentation in (18) that there is a two-parameter family of solutions.

As representative of the case where $m < n$, let $m = 3$ and $n = 5$. There are four possible final patterns, and these are shown schematically in Fig. 1. For instance, the third equation in Fig. 1a could be $x_3 - 6x_4 + 2x_5 = 0$ or $x_3 + 2x_4 + 0x_5 = 4$, and the given third equation in Fig. 1b could be $0 = 0$ or $0 = 0$. It may seem foolish to include the case shown in Fig. 1d because there are no $x_j$’s (all of the $a_{ij}$ coefficients being zero), but it is possible so we have included it. From these patterns we draw these conclusions: (a) there exists a two-parameter family of solutions; (b) there is no solution (the system is inconsistent) if the right-hand member of the third equation is nonzero, and a three-parameter family of solutions if the latter is zero; (c) there is no solution if either of the right-hand members of the second and third equations is nonzero, and a four-parameter family of solutions if each of them is zero; (d) there is no solution if any of the right-hand members is nonzero, and a five-parameter family of solutions if each of them is zero.

It may appear that Fig. 1 does not cover all possible cases. For instance, what about the case shown in Fig. 2? That case can be converted to the case shown in Fig. 1a simply by renaming the unknowns: let $x_3$ become $x_2$ and let $x_5$ become $x_3$. Specifically, let $x_1 \rightarrow x_1, x_3 \rightarrow x_2, x_5 \rightarrow x_3, x_4 \rightarrow x_4$, and $x_2 \rightarrow x_5$.

The case where $m \geq n$ can be studied in a similar manner, and we can draw the following general conclusions.

---

**THEOREM 8.3.2** Existence/Uniqueness for Linear Systems

If $m < n$, the system (1) can be consistent or inconsistent. If it is consistent it cannot have a unique solution; it will have a $p$-parameter family of solutions, where $n - m \leq p \leq n$. If $m \geq n$, (1) can be consistent or inconsistent. If it is

---

*The method is attributed to Karl Friedrich Gauss (1777–1855), who is generally regarded as the foremost mathematician of the nineteenth century and often referred to as the "prince of mathematicians."
consistent it can have a unique solution or a $p$-parameter family of solutions, where $1 \leq p \leq n$.

The next theorem follows immediately from Theorem 8.3.2, but we state it separately for emphasis.

**THEOREM 8.3.3** Existence/Uniqueness for Linear Systems
Every system (1) necessarily admits no solution, a unique solution, or an infinity of solutions.

Observe that a system (1) is inconsistent only if, in its Gauss-eliminated form, one or more of the equations is of the form zero equal to a nonzero number. But that can never happen if every $c_j$ in (1) is zero, that is, if (1) is homogeneous.

**THEOREM 8.3.4** Existence/Uniqueness for Homogeneous Systems
Every homogeneous linear system of $m$ equations in $n$ unknowns is consistent. Either it admits the unique trivial solution or else it admits an infinity of nontrivial solutions in addition to the trivial solution. If $m < n$, then there is an infinity of nontrivial solutions in addition to the trivial solution.

In summary, not only did the method of Gauss elimination provide us with an efficient and systematic solution procedure, it also led us to important results regarding the existence and uniqueness of solutions.

**8.3.3. Matrix notation.** In applying Gauss elimination, we quickly discover that writing the variables $x_1, \ldots, x_n$ over and over is inefficient, and even tends to upstage the more central role of the $a_{ij}$'s and $c_j$'s. It is therefore preferable to omit the $x_j$'s altogether and to work directly with the rectangular array

\[
\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} & c_1 \\
a_{21} & a_{22} & \cdots & a_{2n} & c_2 \\
& & \vdots & \vdots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn} & c_m
\end{bmatrix},
\]

(20)

known as the augmented matrix of the system (1), that is, the coefficient matrix

\[
\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
& & \vdots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix},
\]

(21)
augmented by the column of \( c_j \)'s. By matrix we simply mean a rectangular array of numbers, called elements; it is customary to enclose the elements between parentheses to emphasize that the entire matrix is regarded as a single entity. A horizontal line of elements is called a row, and a vertical line is called a column. Counting rows from the top, and columns from the left,

\[
\begin{pmatrix}
a_{21} & a_{22} & \cdots & a_{2n} & c_2 \\
c_1 \\
\vdots \\
c_m
\end{pmatrix}
\]

say, are the second row and \((n+1)\)th column, respectively, of the augmented matrix (20).

In terms of the abbreviated matrix notation, the calculation in Example 1 would look like this.

Original system:

\[
\begin{bmatrix}
1 & 1 & -1 & 1 \\
3 & 1 & 1 & 9 \\
1 & -1 & 4 & 8
\end{bmatrix}
\]

Add \(-3\) times first row to second row, and add \(-1\) times first row to third row:

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

Add \(-1\) times second row to third row, and multiply second row by \(-\frac{1}{2}\):

\[
\begin{bmatrix}
1 & 1 & -1 & 1 \\
0 & 1 & -2 & -3 \\
0 & 0 & 1 & 1
\end{bmatrix}
\]

Thus, corresponding to the so-called elementary equation operations on members of a system of linear equations there are elementary row operations on the augmented matrix, as follows:

1. Addition of a multiple of one row to another:
   \textit{Symbolically:} \((j\)th row\) \rightarrow \((j\)th row\) + \(\alpha(k\)th row\)

2. Multiplication of a row by a nonzero constant:
   \textit{Symbolically:} \((j\)th row\) \rightarrow \(\alpha(j\)th row\)

3. Interchange of two rows:
   \textit{Symbolically:} \((j\)th row\) \leftrightarrow \((k\)th row\)
And we say that two matrices are row equivalent if one can be obtained from the other by finitely many elementary row operations.

8.3.4. Gauss–Jordan reduction. With the Gauss elimination completed, the remaining steps consist of back substitution. In fact, those steps are elementary row operations as well. The difference is that whereas in the Gauss elimination we proceed from the top down, in the back substitution we proceed from the bottom up.

EXAMPLE 4. To illustrate, let us return to Example 1 and pick up at the end of the Gauss elimination, with (5), and complete the back substitution steps using elementary row operations. In matrix format, we begin with

\[
\begin{bmatrix}
1 & 1 & -1 & 1 \\
0 & 1 & -2 & -3 \\
0 & 0 & 1 & 1 \\
\end{bmatrix}
\]  \hspace{1cm} (23)

Keeping the bottom row intact, add 2 times that row to the second, and add 1 times that row to the first:

\[
\begin{bmatrix}
1 & 1 & 0 & 2 \\
0 & 1 & 0 & -1 \\
0 & 0 & 1 & 1 \\
\end{bmatrix}
\]  \hspace{1cm} (24)

Now keeping the bottom two rows intact, add −1 times the second row to the first:

\[
\begin{bmatrix}
1 & 0 & 0 & 3 \\
0 & 1 & 0 & -1 \\
0 & 0 & 1 & 1 \\
\end{bmatrix}
\]  \hspace{1cm} (25)

which is the solution: \(x_1 = 3, x_2 = -1, x_3 = 1\) as obtained in Example 1.  \(\blacksquare\)

The entire process, of Gauss elimination plus back substitution, is known as Gauss–Jordan reduction, after Gauss and Wilhelm Jordan (1842–1899). The final result is an augmented matrix in **reduced row-echelon form**. That is:

1. In each row not made up entirely of zeros, the first nonzero element is a 1, a so-called leading 1.
2. In any two consecutive rows not made up entirely of zeros, the leading 1 in the lower row is to the right of the leading 1 in the upper row.
3. If a column contains a leading 1, every other element in that column is a zero.
4. All rows made up entirely of zeros are grouped together at the bottom of the matrix.
8.3. Solution by Gauss Elimination

For instance, (25) is in reduced row-echelon form, as is the final matrix in the next example.

EXAMPLE 5. Let us return to Example 3 and finish the Gauss–Jordan reduction, beginning with (18):

\[
\begin{bmatrix}
1 & -3 & 0 & -4 & -2 & 1 & 0 \\
0 & 1 & \frac{1}{2} & 2 & 1 & \frac{1}{2} & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & -3 \\
0 & 0 & 0 & 0 & 0 & 1 & -5
\end{bmatrix}
\rightarrow
\begin{bmatrix}
1 & 0 & \frac{3}{4} & 2 & 1 & \frac{5}{4} & 0 \\
0 & 1 & \frac{1}{2} & 2 & 1 & \frac{1}{2} & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & -3 \\
0 & 0 & 0 & 0 & 0 & 1 & -5
\end{bmatrix}
\rightarrow
\begin{bmatrix}
1 & 0 & \frac{3}{4} & 2 & 0 & 0 & \frac{21}{4} \\
0 & 1 & \frac{1}{2} & 2 & 0 & 0 & \frac{1}{2} \\
0 & 0 & 0 & 0 & 1 & 0 & 2 \\
0 & 0 & 0 & 0 & 0 & 1 & -5
\end{bmatrix}
\]

The last augmented matrix is in reduced row-echelon form. The four leading 1's are displayed in bold type, and we see that, as a result of the back substitution steps, only 0's are to be found above each leading 1. The final augmented matrix once again gives the solution (19).

8.3.5. Pivoting. Recall that the first step in the Gauss elimination of the system

\[
\begin{align*}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= c_1, \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= c_2, \\
\vdots
\end{align*}
\]

is to subtract \(a_{21}/a_{11}\) times the first equation from the second, \(a_{31}/a_{11}\) times the first equation from the third, and so on, while keeping the first equation intact. The first equation is called the pivot equation (or, the first row is the pivot row if one is using the matrix format), and \(a_{11}\) is called the pivot. That step produces an indented system of the form

\[
\begin{align*}
a_{11}'x_1 + a_{12}'x_2 + \cdots + a_{1n}'x_n &= c_1, \\
a_{22}'x_2 + \cdots + a_{2n}'x_n &= c_2', \\
\vdots
\end{align*}
\]

Next, we keep the first two equations intact and use the second equation as the new pivot equation to indent the third through \(m\)th equations, and so on.

Naturally, we need each pivot to be nonzero. For instance, we need \(a_{11} \neq 0\) for \(a_{21}/a_{11}, a_{31}/a_{11}, \ldots\) to be defined. If a pivot is zero, interchange that equation
with any one below it, such as the next equation or last equation (as we did in Example 3), until a nonzero pivot is available. Such interchange of equations is called partial pivoting. If a pivot is zero we have no choice but to use partial partial pivoting, but in practice even a nonzero pivot should be rejected if it is "very small," since the smaller it is the more susceptible is the calculation to the adverse effect of machine roundoff error (see Exercise 13). To be as safe as possible, one can choose the pivot equation as the one with the largest leading coefficient (relative to the other coefficients in the equation).

Closure. Beginning with a system of coupled linear algebraic equations, one can use a sequence of elementary operations to minimize the coupling between the equations while leaving the solution set intact. Besides putting forward the important method of Gauss elimination, which is used heavily in the following chapters, we used that method to establish several important theoretical results regarding the existence and uniqueness of solutions.

The Gauss elimination and Gauss–Jordan reduction discussions lead naturally to a convenient, and equivalent, formulation in matrix notation. We will return to the concept of matrices in Chapter 10, and develop it in detail.

Computer software. Chapters 8–12 cover the domain known as linear algebra. A great many calculations in linear algebra can be carried out using computer algebra systems. In Maple, for instance, a great many commands ("functions") are contained within the linalg package. A listing of these commands can be obtained by entering ?linalg. That list includes the linsolve command, which can be used to solve a system of \( m \) linear algebraic equations in \( n \) unknowns. To access linsolve (or any other command within the linalg package), first enter with(linalg). Then, linsolve(A,b) solves (1) for \( x_1, \ldots, x_n \), where \( A \) is the coefficient matrix and \( b \) is the column of \( c_j \)’s. For instance, the system

\[
\begin{align*}
x_1 - x_2 + 2x_3 - 3x_4 &= 4, \\
x_1 + 2x_2 - x_3 + 3x_4 &= 1
\end{align*}
\]

admits the two-parameter family of solutions

\[
x_4 = \alpha_1, \quad x_3 = \alpha_2, \quad x_2 = -1 - 2\alpha_1 + \alpha_2, \quad x_1 = 3 + \alpha_1 - \alpha_2,
\]

where \( \alpha_1, \alpha_2 \) are arbitrary. To solve (28) using Maple, enter

with(linalg):

then return and enter

\[
\text{linsolve(array([[1, -1, 2, -3], [1, 2, -1, 3]]), array([4, 1]))};
\]

and return. The output is

\[
[ -t_1 + -t_2 + 3, \ -t_1 - 2t_2 - 1, \ -t_1, \ -t_2] \]
where the entries are $x_1, \ldots, x_4$ and where $-t_1$ and $-t_2$ are arbitrary constants.
With $-t_1 = \alpha_2$ and $-t_2 = \alpha_1$, this result is the same as (29). If you prefer, you could use the sequence

\[
\text{with(linalg):} \\
A := \text{array}([[1, -1, 2, -3], [1, 2, -1, 3]]); \\
b := \text{array}([4, 1]); \\
\text{linsolve}(A, b);
\]

instead. If the system is inconsistent, then either the output will be NULL, or there will be no output.

**EXERCISES 8.3**

1. Derive the solution set for each of the following systems using Gauss elimination and augmented matrix format. Document each step (e.g., 2nd row $\rightarrow$ 2nd row + 5 times 1st row), and classify the result (e.g., unique solution, the system is inconsistent, 3-parameter family of solutions, etc.).

   (a) \[ 2x - 3y = 1 \]
       \[ 5x + y = 2 \]
   (b) \[ 2x + y = 0 \]
       \[ 3x - 2y = 0 \]
   (c) \[ x + 2y = 4 \]
   (d) \[ x - y + z = 1 \]
       \[ 2x - y - z = 8 \]
   (e) \[ 2x_1 - x_2 - x_3 - 5x_4 = 6 \]
   (f) \[ 2x_1 - x_2 - x_3 - 3x_4 = 0 \]
       \[ x_1 - x_2 + 4x_4 = 2 \]
   (g) \[ x + 2y + 3z = 4 \]
       \[ 5x + 6y + 7z = 8 \]
       \[ 9x + 10y + 11z = 12 \]
   (h) \[ x_1 + x_2 - 2x_3 = 3 \]
       \[ x_1 - x_2 - 3x_3 = 1 \]
       \[ x_1 - 3x_2 - 4x_3 = -1 \]
   (i) \[ 2x_1 - x_2 = 6 \]
       \[ 3x_1 + 2x_2 = 4 \]
       \[ x_1 + 10x_2 = -12 \]
       \[ 6x_1 + 11x_2 = -2 \]
   (j) \[ x_1 - x_2 + 2x_3 + x_4 = -1 \]
       \[ 2x_1 + x_2 + x_3 - x_4 = 4 \]
       \[ x_1 + 2x_2 - x_3 - 2x_4 = 5 \]
       \[ x_1 + x_3 = 1 \]
   (k) \[ x_1 + x_3 = 1 \]
       \[ x_1 + 2x_2 - x_3 - 2x_4 = 5 \]
       \[ x_1 - x_2 + 2x_3 + x_4 = 0 \]
       \[ 2x_1 + x_2 + x_3 - x_4 = 4 \]
   (l) \[ x_3 + x_4 = 2 \]
       \[ 4x_2 - x_3 + x_4 = 0 \]
   (m) \[ x + 2y + 3z = 5 \]
       \[ 2x + 3y + 4z = 8 \]
       \[ 3x + 4y + 5z = c \]
       \[ x + y = 2 \]
   (n) \[ 2x + y + z = 10 \]
       \[ 3x + y - z = 6 \]
       \[ x - 2y - 4z = -10 \]
   (o) \[ 2x_1 + x_2 = 1 \]
       \[ x_1 + 2x_2 + x_3 = 1 \]
       \[ x_2 + 2x_3 + x_4 = 1 \]
       \[ x_3 + 2x_4 = 1 \]
   (p) \[ 2x_1 + x_2 = 0 \]
       \[ x_1 + 2x_2 + x_3 = -1 \]
       \[ x_2 + 2x_3 = -4 \]
   (q) \[ 2x_1 + x_2 + x_3 + 2x_5 = 0 \]
   (r) \[ x_1 + x_2 + x_3 = 0 \]
   (s) \[ x_1 + 3x_2 + 4x_3 + 2x_5 = 0 \]
   (t) \[ x_2 + 2x_3 = 0 \]

2. (a)–(q) Same as Exercise 1 but using Gauss–Jordan reduction instead of Gauss elimination.

3. (a)–(q) Same as Exercise 1 but using computer software such as the *Maple* linsolve command.

5. Let
\[ a_1x_1 + a_2x_2 + a_3x_3 = 0, \]
\[ b_1x_1 + b_2x_2 + b_3x_3 = 0 \]
represent any two planes through the origin in a Cartesian space. For the case where the planes intersect along a line, show whether or not that line necessarily passes through the origin.

6. If possible, adapt the methods of this section to solve the following nonlinear systems. If it is not possible, say so.
(a) \[ x^2 + 2x^2 - x^2 = 29 \]
\[ x_1^2 + x^2 + x^3 = 19 \]
\[ 3x^2 + 4x^3 = 67 \]
(b) \[ x + 3y = 13 \]
\[ \sin x + 2y = 5 \]
(c) \[ \sin x + \sin y = 1 \]
\[ \sin x - \sin y + 4 \cos z = 1.2 \]
\[ \sin x + \sin y + 2 \cos z = 1.6 \]
where \(-\pi/2 \leq x \leq \pi/2, -\pi/2 \leq y \leq \pi/2, \) and \(0 \leq z \leq 2\pi.\)

7. For what values of the parameter \(\lambda\) do the following homogeneous (do you agree that they are homogeneous?) systems admit nontrivial solutions? Find the nontrivial solutions corresponding to each such \(\lambda\).

(a) \[ 2x + y = \lambda x \]
\[ 2x + y = \lambda y \]
\[ x + 2y = \lambda y \]
\[ x = \lambda y \]
\[ 4x - 8y = \lambda y \]
\[ z = \lambda z \]
\[ x + y + z = \lambda z \]
\[ x + y + z = \lambda z \]
\[ y + z = \lambda y \]
\[ 2z = \lambda z \]
\[ x + y + 2z = \lambda z \]

8. Evaluate these excerpts from examination papers.

(g) “Given the system
\[ x_1 - 2x_2 = 0, \]
\[ 2x_1 - 4x_2 = 0, \]
add \(-2\) times the first equation to the second and add \(-\frac{1}{2}\) times the second equation to the first. By these Gauss elimination steps we obtain the equivalent system \(0 = 0\) and \(0 = 0,\) and hence the two-parameter family of solutions \(x_1 = \alpha_1\) (arbitrary), \(x_2 = \alpha_2\) (arbitrary).”

(b) “Given the system
\[ x_1 + x_2 - 4x_3 = 0, \]
\[ 2x_1 - x_2 + x_3 = 0, \]
since both left-hand sides equal zero, they must equal each other. Hence we have the equation
\[ x_1 + x_2 - 4x_3 = 2x_1 - x_2 + x_3, \]
which equation is equivalent to the original system.”

9. Make up an example of an inconsistent linear algebraic system of equations, with
(a) \(m = 2, n = 4\)
(b) \(m = 1, n = 4\)

10. (Physical example of nonexistence and nonuniqueness: DC circuit) Kirchoff’s current and voltage laws were given in Section 2.3.1. If we apply those laws to the DC circuit shown,

we obtain the equations

\[ i_1 - i_2 - i_3 = 0, \]
\[ i_1 - i_2 - i_3 = 0, \]
\[ R_2i_2 - R_3i_3 = 0, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
\[ R_1i_1 + R_3i_3 = E, \]
\[ R_1i_1 + R_2i_2 = E, \]
11. (Physical example of nonexistence and nonuniqueness: statically indeterminate structures) (a) Consider the static equilibrium of the system shown, consisting of two weightless cables connected at $P$, at which point a vertical load $F$ is applied. Requiring an equilibrium of vertical force components, and horizontal force components too, derive two linear algebraic equations on the unknown tensions $T_1$ and $T_2$. Are there any combinations of angles $\theta_1$ and $\theta_2$ (where $0 \leq \theta_1 \leq \frac{\pi}{2}$ and $0 \leq \theta_2 \leq \frac{\pi}{2}$) such that there is either no solution or a nonunique solution? Explain.

(b) This time let there be three cables at angles of $45^\circ$, $60^\circ$, and $30^\circ$ as shown. Again, requiring an equilibrium of vertical and horizontal forces at $P$, derive two linear algebraic equations on the unknown tensions $T_1$, $T_2$, $T_3$. Show that the equations are consistent so there is a nonunique solution. NOTE: We say that such a structure is statically indeterminate because the forces in it cannot be determined from the laws of statics alone. What information needs to be added if we are to complete the evaluation of $T_1$, $T_2$, $T_3$? What is needed is information about the relative stiffness of the cables. We pursue this to a conclusion in (c), below.

(c) [Completion of part (b)] Before the load $F$ is applied, locate an $x$, $y$ Cartesian coordinate system at $P$. Let $P$ be 1 foot below the “ceiling” so the coordinates of $A$, $B$, $C$ are $(-1, 1)$, $(1/\sqrt{3}, 1)$, and $(\sqrt{3}, 1)$, respectively. Now apply the load $F$.

The point $P$ will move to a point $(x, y)$, and we assume that the cables are stiff enough so that $x$ and $y$ are very small: $|x| \ll 1$ and $|y| \ll 1$. Let the cables obey Hooke’s law: $T_1 = k_1 \delta_1$, $T_2 = k_2 \delta_2$, and $T_3 = k_3 \delta_3$, where $\delta_1$ is the increase in length of the $j$th cable due to the tension $T_j$. Since $P$ moves to $(x, y)$, it follows that

\[
\delta_1 = \sqrt{(x + 1)^2 + (y - 1)^2} - \sqrt{2} \\
= \sqrt{2 + 2(x - y) + (x^2 + y^2)} - \sqrt{2} \\
\approx \sqrt{2 + 2(x - y) - \sqrt{2}} \\
= \sqrt{2} [1 + (x - y)]^{1/2} - \sqrt{2} \\
\approx \sqrt{2} \left[1 + \frac{1}{2}(x - y)\right] - \sqrt{2} = \frac{1}{\sqrt{2}}(x - y).
\]

(11.1)

Explain each step in (11.1), and show, similarly, that

\[
\delta_2 \approx -\frac{1}{2}x - \frac{\sqrt{3}}{2}y,
\]

(11.2)

\[
\delta_3 \approx -\frac{\sqrt{3}}{2}x - \frac{1}{2}y.
\]

(11.3)

Thus,

\[
T_1 = k_1 \delta_1 \approx \frac{k_1}{\sqrt{2}}(x - y),
\]

(11.4)

\[
T_2 = k_2 \delta_2 \approx -\frac{k_2}{2}(x + \sqrt{3}y),
\]

\[
T_3 = k_3 \delta_3 \approx -\frac{k_3}{2}(\sqrt{3}x + y).
\]

Putting (11.4) into the two equilibrium equations obtained in (b) then gives two equations in the unknown displacements $x$, $y$. Show that that system can be solved uniquely for $x$ and $y$, and thus complete the solution for $T_1$, $T_2$, $T_3$.

12. (Roundoff error difficulty due to small pivots) To illustrate how small pivots can accentuate the effects of roundoff error, consider the system

\[
0.005x_1 + 1.47x_2 = 1.49,
0.975x_1 + 2.32x_2 = 6.22
\]

(12.1)

with exact solution $x_1 = 4$ and $x_2 = 1$. Suppose that our computer carries three significant figures and then rounds off. Using the first equation as our pivot equation, Gauss elimination gives

\[
\begin{bmatrix}
0.005 & 1.47 & 1.49 \\
0.975 & 2.32 & 6.22
\end{bmatrix} \rightarrow
\begin{bmatrix}
0.005 & 1.47 & 1.49 \\
0 & -285 & -284
\end{bmatrix}
\]
so \( x_2 = \frac{284}{285} = 0.996 \) and \( x_1 = \frac{1.49 - (1.47)(0.996)}{0.005} = \frac{1.49 - 1.46}{0.005} = 6 \). Show that if we use partial pivoting and then use the first equation of the system

\[
\begin{align*}
0.975x_1 + 2.32x_2 &= 6.22, \\
0.005x_1 + 1.47x_2 &= 1.49
\end{align*}
\]  

(12.2)

as our pivot equation, we obtain the result \( x_1 = 4.00 \) and \( x_2 = 1.00 \) (which happens to be exactly correct).

13. (Ill-conditioned systems) Practically speaking, our numerical calculations are normally carried out on computers, be they hand-held calculators or large digital computers. Such machines carry only a finite number of significant figures and thus introduce roundoff error into most calculations. One might expect (or hope) that such slight deviations will lead to answers that are only slightly in error. For example, the solution of

\[
\begin{align*}
x + y &= 2, \\
x - 1.014y &= 0
\end{align*}
\]  

(13.1)

is \( x \approx 1.007, y \approx 0.993 \), whereas the solution of the rounded-off version

\[
\begin{align*}
x + y &= 2, \\
x - 1.01y &= 0
\end{align*}
\]

is very much the same, namely \( x \approx 1.005, y \approx 0.995 \). In sharp contrast, the solutions of

\[
\begin{align*}
x + y &= 2, \\
x + 1.014y &= 0
\end{align*}
\]  

(13.2)

and the rounded-off version

\[
\begin{align*}
x + y &= 2, \\
x + 1.01y &= 0
\end{align*}
\]

\( x \approx 144.9, y \approx -142.9 \) and \( x = 202, y = -200 \), respectively; (13.2) is an example of a so-called ill-conditioned system (ill-conditioned in the sense that small changes in the coefficients lead to large changes in the solution). Here, we ask the following: Explain why (13.2) is much more sensitive to roundoff than (13.1) by exploring the two cases graphically, that is, in the \( x, y \) plane.

---

**Chapter 8 Review**

This chapter deals with systems of linear algebraic equations, \( m \) equations in \( n \) unknowns, insofar as the existence and uniqueness of solutions and solution technique. We find that there are three possibilities: a unique solution, no solution, and an infinity of solutions. If one or more solutions exist then the system is said to be consistent, if there are no solutions then it is inconsistent.

The key, in assessing existence/uniqueness as well as in finding solutions, is provided by elementary operations because they enable us to manipulate the system so as to reduce the coupling to a minimum, while at the same time keeping the solution set intact.

The method of Gauss elimination is introduced, as a systematic solution procedure based upon the three elementary operations, and it is shown that the subsequent back substitution steps amount to elementary operations as well. The entire process, Gauss elimination followed by the back substitution, is known as Gauss–Jordan reduction. Realize that the latter is a solution method, or algorithm, not a formula for the solution. Explicit solution formulas are developed, but not until Chapter 10.

We find that the process of Gauss elimination and Gauss–Jordan reduction are expressed most conveniently in matrix notation, although that notation is not essential to the method. In subsequent chapters the matrix approach is developed
more fully.

The most important results of this chapter are contained in Theorems 8.3.1–3. Finally, we also stress the value of geometrical and visual reasoning, and suggest that you keep that idea in mind as we proceed.
Chapter 9

Vector Space

9.1 Introduction

Normally, one meets vectors for the first time within some physical context — in studying mechanics, electric and magnetic fields, and so on. There, the vectors exist within two- or three-dimensional space and correspond to force, velocity, position, magnetic field, and so on. They have both magnitude and direction; they can be scaled by multiplicative factors, added according to the parallelogram law; dot and cross product operations are defined between vectors; the angle between two vectors is defined; vectors can be expanded as linear combinations of base vectors; and so on.

Alternatively, there exists a highly formalized axiomatic approach to vectors known as linear vector space or abstract vector space. Although this generalized vector concept is essentially an outgrowth of the more primitive system of “arrow vectors” in 2-space and 3-space, described above, it extends well beyond that system in scope and applicability.

For pedagogical reasons, we break the transition from 2-space and 3-space to abstract vector space into two steps: in Sections 9.4 and 9.5 we introduce a generalization to “n-space,” and in Section 9.6 we complete the extension to general vector space, including function spaces where the vectors are functions! However, we do not return to function spaces until Chapter 17, in connection with Fourier series and the Sturm–Liouville theory; in Chapters 9–12 our chief interest is in n-space.

9.2 Vectors; Geometrical Representation

Some quantities that we encounter may be completely defined by a single real number, or magnitude; the mass or kinetic energy of a given particle, and the temperature or salinity at some point in the ocean, are examples. Others are not defined solely by a magnitude but rather by a magnitude and a direction, examples being force, velocity, momentum, and acceleration. Such quantities are called
9.2. Vectors; Geometrical Representation

Vectors.

The defining features of a vector being magnitude and direction suggests the geometric representation of a vector as a directed line segment, or “arrow,” where the length of the arrow is scaled according to the magnitude of the vector. For example, if the wind is blowing at 8 meters/sec from the northeast, that defines a wind-velocity vector \( \mathbf{v} \), where we adopt **boldface type** to signify that the quantity is a vector; alternative notations include the use of an overhead arrow as in \( \vec{v} \). Choosing, according to convenience, a scale of 5 meters/sec per centimeter, say, the geometric representation of \( \mathbf{v} \) is as shown in Fig. 1. Denoting the magnitude, or **norm**, of any vector \( \mathbf{v} \) as \( \| \mathbf{v} \| \), we have \( \| \mathbf{v} \| = 8 \) for the \( \mathbf{v} \) vector in Fig. 1.

Observe that the **location** of a vector is not specified, only its magnitude and direction. Thus, the two unlabeled arrows in Fig. 1 are equally valid alternative representations of \( \mathbf{v} \). That is not to say that the physical effect of the vector will be entirely independent of its position. For example, it should be apparent that the motion of the body \( \mathbf{B} \) induced by a force \( \mathbf{F} \) (Fig. 2) will certainly depend on the point of application of \( \mathbf{F} \) as will the stress field induced in \( \mathbf{B} \). Nevertheless, the two vectors in Fig. 2 are still regarded as equal, as are the three in Fig. 1.

Like numbers, vectors do not become useful until we introduce rules for their manipulation, that is, a vector algebra. Having elected the arrow representation of vectors, the vector algebra that we now introduce will, likewise, be geometric.

First, we say that two vectors are **equal** if and only if their lengths are identical and if their directions are identical as well.

Next, we define a process of **addition** between any two vectors \( \mathbf{u} \) and \( \mathbf{v} \). The first step is to move \( \mathbf{v} \) (if necessary), parallel to itself, so that its tail coincides with the head of \( \mathbf{u} \). Then the sum, or **resultant**, \( \mathbf{u} + \mathbf{v} \) is defined as the arrow from the tail of \( \mathbf{u} \) to the head of \( \mathbf{v} \), as in Fig. 3a. Reversing the order, \( \mathbf{v} + \mathbf{u} \) is as shown in Fig. 3b. Equivalently, we may place \( \mathbf{u} \) and \( \mathbf{v} \) tail to tail, as in Fig. 3c. Comparing Fig. 3c with Fig. 3a and b, we see that the diagonal of the parallelogram (Fig. 3c), is both \( \mathbf{u} + \mathbf{v} \) and \( \mathbf{v} + \mathbf{u} \). Thus,

\[
\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u},
\]

so addition is commutative. One may show (Exercise 3) that it is associative as well,

\[
(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w}).
\]

Next, we define any vector of zero length to be a **zero vector**, denoted as \( \mathbf{0} \). Its length being zero, its direction is immaterial; any direction may be assigned if desired. From the definition of addition above, it follows that

\[
\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}
\]

for each vector \( \mathbf{u} \).

Corresponding to \( \mathbf{u} \) we define a **negative inverse** \( -\mathbf{u} \) such that if \( \mathbf{u} \) is any nonzero vector, then \( -\mathbf{u} \) is determined uniquely, as shown in Fig. 4a; that is, it is

---

\*Students of mechanics know that the point of application of \( \mathbf{F} \) affects the **rotational** part of the motion but not the **translational** part.
414 Chapter 9. Vector Space

of the same length as \( u \) but is directed in the opposite direction (again, \( u \) and \( -u \) have the same length, the length of \( -u \) is not negative). For the zero vector we have \(-0 = 0\). We denote \( u + (-v) \) as \( u - v \) ("\( u \) minus \( v \") but emphasize that it is really the addition of \( u \) and \(-v \), as in Fig. 4b.

Finally, we introduce another operation, called scalar multiplication, between any vector \( u \) and any scalar (i.e., a real number) \( \alpha \): If \( \alpha \neq 0 \) and \( u \neq 0 \), then \( \alpha u \) is a vector whose length is \( |\alpha| \) times the length of \( u \) and whose direction is the same as that of \( u \) if \( \alpha > 0 \), and the opposite if \( \alpha < 0 \); if \( \alpha = 0 \) and/or \( u = 0 \), then \( \alpha u = 0 \). This definition is illustrated in Fig. 5. It follows from this definition that scalar multiplication has the following algebraic properties:

\[
\begin{align*}
\alpha(\beta u) &= (\alpha\beta)u, \\
(\alpha + \beta)u &= \alpha u + \beta u, \\
\alpha(u + v) &= \alpha u + \alpha v, \\
1u &= u,
\end{align*}
\]

where \( \alpha, \beta \) are any scalars and \( u, v \) are any vectors.

Observe that the parallelogram rule of vector addition is a definition so it does not need to be proved. Nevertheless, definitions are not necessarily fruitful so it is worthwhile to reflect for a moment on why the parallelogram rule has proved important and useful. Basically, if we say that "the sum of \( u \) and \( v \) is \( w \)," and thereby pass from the two vectors \( u, v \) to the single vector \( w \), it seems fair to expect some sort of equivalence to exist between the action of \( w \) and the joint action of \( u \) and \( v \). For example, if \( F_1 \) and \( F_2 \) are two forces acting on a body \( B \), as shown in Fig. 6, it is known from fundamental principles of mechanics that their combined effect will be the same as that due to the single force \( F \), so it seems reasonable and natural to say that \( F \) is the sum of \( F_1 \) and \( F_2 \). This concept goes back at least as far as Aristotle (384–322 B.C.). Thus, while the algebra of vectors is developed here as an essentially mathematical matter, it is important to appreciate the role of physics and physical motivation.

In closing this section, let us remark that our foregoing discussion should not be construed to imply that objects of physical interest are necessarily vectors (as are force and velocity) or scalars (as are temperature, mass, and speed (i.e., the magnitude of the velocity vector)). For example, in the study of mechanics one finds that more than a magnitude and a direction are needed to fully define the state of stress at a point; in fact, a "second-order tensor" is needed – a quantity that is more exotic than a vector in much the same way that a vector is more exotic than a scalar.\(^*\)

\(^*\)For an introduction to tensors, we recommend to the interested reader the 68-page book Tensor Analysis by H. D. Block (Columbus, OH: Charles E. Merrill, 1962).
EXERCISES 9.2

1. Trace the vectors \( \mathbf{A}, \mathbf{B}, \mathbf{C} \), shown where \( \mathbf{A} \) is twice as long as \( \mathbf{B} \). Then determine each of the following by graphical means.
   (a) \( \mathbf{A} + \mathbf{B} + \mathbf{C} \)
   (b) \( \mathbf{B} - \mathbf{A} \)
   (c) \( \mathbf{A} - \mathbf{C} + 3\mathbf{B} \)
   (d) \( 2(\mathbf{B} - \mathbf{A}) + 60 \)
   (e) \( \mathbf{A} + (4\mathbf{B} - \mathbf{C}) \)
   (f) \( \mathbf{A} + 2\mathbf{B} - 2\mathbf{C} \)

2. In each case, \( \mathbf{C} \) can be expressed as a linear combination of \( \mathbf{A} \) and \( \mathbf{B} \), that is, as \( \mathbf{C} = \alpha \mathbf{A} + \beta \mathbf{B} \). Trace the three vectors and by graphical means determine \( \alpha \) and \( \beta \).

3. Show that the associative property (2) follows from the graphical definition of vector addition.

4. Derive the following from the definitions of vector addition and scalar multiplication:
   (a) property (4a)
   (b) property (4b)
   (c) property (4c)
   (d) property (4d)

5. (a) If \( \|\mathbf{A}\| = 1 \), \( \|\mathbf{B}\| = 2 \), and \( \|\mathbf{C}\| = 5 \), can \( \mathbf{A} + \mathbf{B} + \mathbf{C} = 0 \)?

HINT: Use the law of cosines \( s^2 = q^2 + r^2 - 2qr \cos \theta \) (see the accompanying figure) or the Euclidean proposition that the length of any one side of a triangle cannot exceed the sum of the lengths of the other two sides.

(b) Repeat part (a), with \( \|\mathbf{A}\| = 1 \) changed to \( \|\mathbf{A}\| = 4 \).

6. Use the definitions and properties given in the reading to show that \( \mathbf{A} + \mathbf{B} = \mathbf{C} \) implies that \( \mathbf{A} = \mathbf{C} - \mathbf{B} \).

7. (a) Show that if \( \mathbf{A} + \mathbf{B} = 0 \) and \( \mathbf{A} \) and \( \mathbf{B} \) are not parallel, then each of \( \mathbf{A} \) and \( \mathbf{B} \) must be 0.
   (b) Vectors are often of help in deriving geometrical relationships. For example, to show that the diagonals of a parallelogram bisect each other one may proceed as follows. From the accompanying figure \( \mathbf{A} + \mathbf{B} = \mathbf{C} \), \( \mathbf{A} - \alpha \mathbf{D} = \beta \mathbf{C} \), and \( \mathbf{A} = \mathbf{B} + \mathbf{D} \). Eliminating \( \mathbf{A} \) and \( \mathbf{B} \), we obtain \( (2\beta - 1)\mathbf{C} = (1 - 2\alpha)\mathbf{D} \), and since \( \mathbf{C} \) and \( \mathbf{D} \) are not parallel, it must be true [per part (a)] that \( 2\beta - 1 = 1 - 2\alpha = 0 \) (i.e., \( \alpha = \beta = \frac{1}{2} \)), which completes the proof. We now state the problem: Use this sort of procedure to show that a line from one vertex of a parallelogram to the midpoint of a nonadjacent side trisects a diagonal.

8. If (see the accompanying figure) the vector \( \mathbf{A} + \alpha \mathbf{B} \) is placed with its tail at point \( P \), show the line generated by its head as \( \alpha \) varies between \(-\infty \) and \(+\infty \).

9. If (see the accompanying figure) \( \|\mathbf{AB}\| / \|\mathbf{AC}\| = \alpha \), show that \( \mathbf{OB} = \alpha \mathbf{OC} + (1 - \alpha)\mathbf{OA} \).
10. One may express linear displacement as a vector: If a particle moves from point A to point B, the displacement vector is the directed line segment, say \( \mathbf{u} \), from A to B. For example, observe that a displacement \( \mathbf{u} \) from A to B, followed by a displacement \( \mathbf{v} \) from B to C, is equivalent to a single displacement \( \mathbf{w} \) from A to C: \( \mathbf{u} + \mathbf{v} = \mathbf{w} \) [part (a) in the accompanying figure]. Reversing the order, displacements \( \mathbf{v} \) and then \( \mathbf{u} \) also carry us from A to C: \( \mathbf{v} + \mathbf{u} = \mathbf{w} \) [part (b) in the figure]. Thus, \( \mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u} \) so that the commutativity axiom (1) is indeed satisfied. How about angular displacements? Suppose that we express the angular displacement of a rigid body about an axis as \( \theta \), where the magnitude of \( \theta \) is equal to the angle of rotation, and the orientation of \( \theta \) is along the axis of rotation, in the direction specified by the “right-hand rule.” That is, if we curl the fingers of our right hand about the axis of rotation, in the direction of rotation, then the direction \( \theta \) along the axis of rotation is the direction in which our thumb points. The problem is to show that \( \theta \), defined in this way, is not a proper vector quantity. HINT: Considering the unit cube shown below, say, show (by keeping track of the coordinates of the corner A) that the orientation that results from a rotation of \( \pi/2 \) about the \( x \) axis, followed by a rotation of \( \pi/2 \) about the \( y \) axis, is not the same as that which results when the order of the rotations is reversed. NOTE: If you have encountered angular velocity vectors (usually denoted as \( \omega \) or \( \Omega \)), in mechanics, it may seem strange to you that finite rotations (assigned a vector direction by the right-hand rule) are not true vectors. The idea is that angular velocity involves infinitesimal rotations, and infinitesimal rotations (assigned a vector direction by the right-hand rule) are true vectors. This subtle point is discussed in many sources (e.g., Robert R. Long, Engineering Science Mechanics, Englewood Cliffs, NJ: Prentice Hall, 1963, pp. 31–36).

9.3 Introduction of Angle and Dot Product

Continuing our discussion, we define here the angle between two vectors and a “dot product” operation between two vectors. The angle \( \theta \) between two nonzero vectors \( \mathbf{u} \) and \( \mathbf{v} \) will be understood to mean the ordinary angle between the two vectors when they are arranged tail to tail as in Fig. 1. (We will not attempt to define \( \theta \) if one or both of the vectors is \( \mathbf{0} \).) Of course, this definition of \( \theta \) is ambiguous in that there are two such angles, an interior angle (\( \leq \pi \)) and an exterior angle (\( \geq \pi \)); for definiteness, we choose \( \theta \) to be the interior angle,

\[
0 \leq \theta \leq \pi,
\]

as in Fig. 1. Unless explicitly stated otherwise, angular measure will be understood to be in radians.
Next, we define the so-called dot product, \( \mathbf{u} \cdot \mathbf{v} \), between two vectors \( \mathbf{u} \) and \( \mathbf{v} \) as

\[
\mathbf{u} \cdot \mathbf{v} \equiv \begin{cases} 
\|\mathbf{u}\| \|\mathbf{v}\| \cos \theta & \text{if } \mathbf{u}, \mathbf{v} \neq \mathbf{0}, \\
0 & \text{if } \mathbf{u} = \mathbf{0} \text{ or } \mathbf{v} = \mathbf{0};
\end{cases}
\]  

(2a,b)

\( \|\mathbf{u}\|, \|\mathbf{v}\|, \) and \( \cos \theta \) are scalars so \( \mathbf{u} \cdot \mathbf{v} \) is a scalar, too.*

By way of geometrical interpretation, observe (Fig. 2a) that \( \|\mathbf{u}\| \cos \theta \) is the length of the orthogonal projection of \( \mathbf{u} \) on the line of action of \( \mathbf{v} \) so that \( \mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta = (\|\mathbf{v}\|) (\|\mathbf{u}\| \cos \theta) \) is the length of \( \mathbf{v} \) times the length of the orthogonal projection of \( \mathbf{u} \) on the line of action of \( \mathbf{v} \).† Actually, that statement holds if \( 0 \leq \theta \leq \pi/2 \); if \( \pi/2 < \theta < \pi \), the cosine is negative, and \( \mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta \) is the negative of the length of \( \mathbf{v} \) times the length of the orthogonal projection of \( \mathbf{u} \) on the line of action of \( \mathbf{v} \).

**EXAMPLE 1.** *Work Done by a Force.* In mechanics the work \( W \) done when a body undergoes a linear displacement from an initial point \( A \) to a final point \( B \), under the action of a constant force \( \mathbf{F} \) (Fig. 3), is defined as the length of the orthogonal projection of \( \mathbf{F} \) on the line of displacement, positive if \( \mathbf{F} \) is "assisting" the motion (i.e., if \( 0 \leq \theta < \pi/2 \), as in Fig. 3a) and negative if \( \mathbf{F} \) is "opposing" the motion (i.e., if \( \pi/2 < \theta < \pi \), as in Fig. 3b), times the displacement. By the displacement we mean the length of the vector \( \mathbf{AB} \) with head at \( B \) and tail at \( A \). But that product is precisely the dot product of \( \mathbf{F} \) with \( \mathbf{AB} \).

\[ W = \mathbf{F} \cdot \mathbf{AB}. \]  

(3)

An important special case of the dot product occurs when \( \theta = \pi/2 \). Then \( \mathbf{u} \) and \( \mathbf{v} \) are perpendicular, and

\[ \mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos \frac{\pi}{2} = 0. \]  

(4)

Also of importance is the case where \( \mathbf{u} = \mathbf{v} \). Then, according to (2),

\[ \mathbf{u} \cdot \mathbf{u} \equiv \begin{cases} 
\|\mathbf{u}\| \|\mathbf{u}\| \cos 0 = \|\mathbf{u}\|^2 & \text{if } \mathbf{u} \neq \mathbf{0}, \\
0 & \text{if } \mathbf{u} = \mathbf{0}
\end{cases} \]  

(5)

so that we have

\[ \|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}} \]  

(6)

---

*You may wonder why (2b) is needed since if \( \mathbf{u} = \mathbf{0} \), say, then \( \|\mathbf{u}\| = 0 \) and \( \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta \) is apparently zero anyway. But the point is that if \( \mathbf{u} \) and/or \( \mathbf{v} \) are \( \mathbf{0} \), then \( \theta \) is undefined; hence \( \cos \theta \) (and even zero times \( \cos \theta \)) is undefined, too.

†Alternatively, we could decompose \( \mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta = (\|\mathbf{u}\|) (\|\mathbf{v}\| \cos \theta) \); that is, as the length of \( \mathbf{u} \) times the length of the orthogonal projection of \( \mathbf{v} \) on the line of action of \( \mathbf{u} \).
whether \( u \neq 0 \) or not. The relationship (6) between the dot product and the norm will be useful in subsequent sections.

**EXERCISES 9.3**

1. Evaluate \( u \cdot v \) in each case. In (a) \( \|u\| = 5 \), in (b) \( \|u\| = 3 \), and in (c) \( \|u\| = 6 \).

   ![Diagram](image)

   (a) \( u = (\pi, 0, 3) \), (b) \( u = (2, 0, 1) \), (c) \( u = (-1, 0, -2) \).

2. (Properties of the dot product) Prove each of the following properties of the dot product, where \( \alpha, \beta \) are any scalars, and \( u, v, w \) are any vectors.

   (a) \( u \cdot v = v \cdot u \) (commutativity)
   (b) \( u \cdot u > 0 \) for all \( u \neq 0 \) (nonnegativeness)
   (c) \( (\alpha u + \beta v) \cdot w = \alpha(u \cdot w) + \beta(v \cdot w) \) (linearity)

   HINT: In proving part (c), you may wish to show, first, that part (c) is equivalent to the two conditions \( (u + v) \cdot w = u \cdot w + v \cdot w \) and \( (\alpha u) \cdot v = \alpha(u \cdot v) \).

3. Using the properties given in Exercise 2, show that

   \[
   (u + v) \cdot (w + x) = u \cdot w + u \cdot x + v \cdot w + v \cdot x. \tag{3.1}
   \]

4. Consider the unit cube shown, where \( P \) is the midpoint of the right-hand face. Evaluate each of the following using the definition (2), and (3.1) in Exercise 3. HINT: To evaluate \( AC \cdot OP \), for instance, write \( AC \cdot OP = (AD + DC) \cdot (OD + DP) \) and then use (3.1).

5. Referring to the figure in Exercise 4, use the dot product to compute the following angles. (See the hint in Exercise 4.) You may use (2), (6), and (3.1).

   (a) \( \angle APO \) (b) \( \angle APB \) (c) \( \angle APC \) (d) \( \angle AOP \) (e) \( \angle ABP \) (f) \( \angle ACP \) (g) \( \angle BPO \) (h) \( \angle BPC \) (i) \( \angle BPD \) (j) \( \angle BOP \) (k) \( \angle CPO \) (l) \( \angle DPO \)

6. If \( u \) and \( v \) are nonzero, show that \( w = \|v\| u + \|u\| v \) bisects the angle between \( u \) and \( v \). (You may use any of the properties given in Exercise 2.)

---

**9.4 n-Space**

Here we move away from our dependence on the arrow representation of vectors, in 2-space and 3-space, by introducing an alternative representation in terms of 2-tuples and 3-tuples. This step will lead us to a more general notion of vectors in "\( n \)-space."
The idea is simple and is based on the familiar representation of points in Cartesian 1-, 2-, and 3-space as 1-, 2-, and 3-tuples of real numbers. For example, the 2-tuple \((a_1, a_2)\) denotes the point \(P\) indicated in Fig. 1a, where \(a_1, a_2\) are the \(x, y\) coordinates, respectively. But it can also serve to denote the vector \(\mathbf{OP}\) in Fig. 1b or, indeed, any equivalent vector \(\mathbf{QR}\).

Thus the vector is now represented as the 2-tuple \((a_1, a_2)\) rather than as an arrow, and while pictures may still be drawn, as in Fig. 1b, they are no longer essential and can be discarded if we wish — at least once the algebra of 2-tuples is established (in the next paragraph). The set of all such real 2-tuple vectors will be called 2-space and will be denoted by the symbol \(\mathbb{R}^2\); that is,

\[
\mathbb{R}^2 = \{ (a_1, a_2) \mid a_1, a_2 \text{ real numbers} \}.
\]

Vectors \(\mathbf{u} = (u_1, u_2)\) and \(\mathbf{v} = (v_1, v_2)\) in \(\mathbb{R}^2\) are defined to be equal if \(u_1 = v_1\) and \(u_2 = v_2\); their sum is defined as

\[
\mathbf{u} + \mathbf{v} \equiv (u_1 + v_1, u_2 + v_2)
\]

as can be seen from Fig. 2; the scalar multiple \(\alpha \mathbf{u}\) is defined, for any scalar \(\alpha\), as

\[
\alpha \mathbf{u} \equiv (\alpha u_1, \alpha u_2); \tag{3}
\]

the zero vector is \(\mathbf{0} \equiv (0, 0);\)

and the negative of \(\mathbf{u}\) is \(-\mathbf{u} \equiv (-u_1, -u_2).\)

Similarly, for \(\mathbb{R}^3\):

\[
\mathbb{R}^3 = \{ (a_1, a_2, a_3) \mid a_1, a_2, a_3 \text{ real numbers} \},
\]

\[
\mathbf{u} + \mathbf{v} \equiv (u_1 + v_1, u_2 + v_2, u_3 + v_3), \tag{7}
\]

and so on.†

It may not be evident that we have gained much since the arrow and \(n\)-tuple representations are essentially equivalent. But, in fact, the \(n\)-tuple format begins to "open doors." For example, the instantaneous state of the electrical circuit (consisting of a battery and two resistors) shown in Fig. 3 may be defined by the two currents \(i_1\) and \(i_2\) or, equivalently, by the single 2-tuple vector \((i_1, i_2)\). Thus, even though "magnitudes," "directions," and "arrow vectors" may not leap to mind in describing the system shown in Fig. 3, a vector representation is quite natural within the \(n\)-tuple framework, and that puts us in a position, in dealing with that electrical system, to make use of whatever vector theorems and techniques are available, as developed in subsequent sections and chapters.

---

†We use the \(\equiv\) equal sign to mean equal to by definition.

†The space \(\mathbb{R}^1\) of 1-tuples will not be of interest here.
Indeed, why stop at 3-tuples? One may introduce the set of all ordered real \( n \)-tuple vectors, even if \( n \) is greater than 3. We call this \( n \)-space, and denote it as \( \mathbb{R}^n \), that is,

\[
\mathbb{R}^n = \{(a_1, \ldots, a_n) \mid a_1, \ldots, a_n \text{ real numbers}\}.
\]

Consider two vectors, \( \mathbf{u} = (u_1, \ldots, u_n) \) and \( \mathbf{v} = (v_1, \ldots, v_n) \), in \( \mathbb{R}^n \). The scalars \( u_1, \ldots, u_n \) and \( v_1, \ldots, v_n \) are called the components of \( \mathbf{u} \) and \( \mathbf{v} \). As you may well expect, based on our foregoing discussion of \( \mathbb{R}^2 \) and \( \mathbb{R}^3 \), \( \mathbf{u} \) and \( \mathbf{v} \) are said to be equal if \( u_1 = v_1, \ldots, u_n = v_n \), and we define

\[
\begin{align*}
\mathbf{u} + \mathbf{v} &\equiv (u_1 + v_1, \ldots, u_n + v_n), & \text{(addition)} \\
\alpha \mathbf{u} &\equiv (\alpha u_1, \ldots, \alpha u_n), & \text{(scalar multiplication)} \\
\mathbf{0} &\equiv (0, \ldots, 0), & \text{(zero vector)} \\
-\mathbf{u} &\equiv (-1) \mathbf{u}, & \text{(negative inverse)} \\
\mathbf{u} - \mathbf{v} &\equiv \mathbf{u} + (-\mathbf{v}). & \text{(9e)}
\end{align*}
\]

From these definitions we may deduce the following properties:

\[
\begin{align*}
\mathbf{u} + \mathbf{v} &= \mathbf{v} + \mathbf{u}, & \text{(commutativity)} \\
(\mathbf{u} + \mathbf{v}) + \mathbf{w} &= \mathbf{u} + (\mathbf{v} + \mathbf{w}), & \text{(associativity)} \\
\mathbf{u} + \mathbf{0} &= \mathbf{u}, & \text{(10c)} \\
\mathbf{u} + (-\mathbf{u}) &= \mathbf{0}, & \text{(10d)} \\
\alpha(\beta \mathbf{u}) &= (\alpha \beta) \mathbf{u}, & \text{(associativity)} \\
(\alpha + \beta) \mathbf{u} &= \alpha \mathbf{u} + \beta \mathbf{u}, & \text{(10f)} \\
\alpha(\mathbf{u} + \mathbf{v}) &= \alpha \mathbf{u} + \alpha \mathbf{v}, & \text{(10g)} \\
1 \mathbf{u} &= \mathbf{u}, & \text{(10h)} \\
0 \mathbf{u} &= \mathbf{0}, & \text{(10i)} \\
(-1) \mathbf{u} &= -\mathbf{u}, & \text{(10j)} \\
\alpha \mathbf{0} &= \mathbf{0}. & \text{(10k)}
\end{align*}
\]

To illustrate how such \( n \)-tuples might arise, observe that the state of the electrical system shown in Fig. 4, may be defined at any instant by the four currents \( i_1, i_2, i_3, i_4 \), and that these may be regarded as the components of a single vector \( \mathbf{i} = (i_1, i_2, i_3, i_4) \) in \( \mathbb{R}^4 \).

Of course, the notation of \( (u_1, \ldots, u_n) \) as a point or arrow in an \( n \)-dimensional space" can be realized graphically only if \( n \leq 3 \); if \( n > 3 \), the interpretation is valid only in an abstract, or schematic, sense. However, our inability to carry out traditional Cartesian graphical constructions for \( n > 3 \) will be no hindrance. Indeed, part of the idea here is to move away from a dependence on graphical constructions.

Having extended the vector concept to \( \mathbb{R}^n \), you may well wonder if further extension is possible. Such extension is not only possible, it constitutes an important step in modern mathematics; more about this in Section 9.6.

![Figure 4. Another electrical circuit.](image)
EXERCISES 9.4

1. If \( t = (5,0,1,2) \), \( u = (2,-1,3,4) \), \( v = (4,-5,1) \), \( w = (-1,-2,5,6) \), evaluate each of the following (as a single vector); if the operation is undefined (i.e., has not been defined here), state that. At each step cite the equation number of the definition or property being used.

(a) \( 2t + 7u \)  
(b) \( 3t - 5u \)  
(c) \( 4[u + 5(w - 2u)] \)  
(d) \( 4tu + w \)  
(e) \( -w + t \)  
(f) \( 2t/u \)  
(g) \( t + 2u + 3w \)  
(h) \( t - 2u - 4v \)  
(i) \( u(3t + w) \)  
(j) \( u^2 + 2t \)  
(k) \( 2t + 7u - 4 \)  
(l) \( u + wt \)  
(m) \( \sin u \)  
(n) \( w + t - 2u \)

2. Let \( u = (1,3,0,-2) \), \( v = (2,0,-5,0) \), and \( w = (4,3,2,-1) \).

(a) \( 3u - x = 4(v + 2x) \), solve for \( x \). (i.e., find its components.)
(b) If \( x + u + v + w = 0 \), solve for \( x \).

3. Let \( u \), \( v \), and \( w \) be as given in Exercise 2. Citing the definition or property used, at each step, solve each of the following for \( x \). NOTE: Besides the definitions and properties stated in this section, it should be clear that if \( x = y \), then \( x + z = y + z \) for any \( z \), and \( \alpha x = \alpha y \) for any \( \alpha \) (adding and multiplying equals by equals.)

(a) \( 3x + 2(u - 5v) = w \)  
(b) \( 3x = 40 + (1,0,0,0) \)  
(c) \( u - 4x = 0 \)  
(d) \( u + v - 2x = w \)

4. If \( t = (2,1,3) \), \( u = (1,2,-4) \), \( v = (0,1,1) \), \( w = (-2,1,-1) \), solve each of the following for the scalars \( \alpha_1, \alpha_2, \alpha_3 \). If no such scalars exist, state that.

(a) \( \alpha_1 t + \alpha_2 u + \alpha_3 v = 0 \)  
(b) \( \alpha_1 t + \alpha_2 v + \alpha_3 w = 0 \)  
(c) \( \alpha_1 t + \alpha_2 u + \alpha_3 w = (1,3,2) \)  
(d) \( \alpha_1 t + \alpha_2 v + \alpha_3 w = (2,0,-1) \)  
(e) \( \alpha_1 u + \alpha_2 v = 0 \)  
(f) \( \alpha_1 u + \alpha_2 v = \alpha_3 w - (2,0,0) \)

5. (a) If \( u \) and \( v \) are given 4-tuples and \( 0 = (0,0,0,0) \), does the vector equation \( \alpha_1 u + \alpha_2 v = 0 \) necessarily have nontrivial solutions for the scalars \( \alpha_1 \) and \( \alpha_2 \)? Explain. (If the answer is "no," a counterexample will suffice.)
(b) Repeat part (a), but where \( u \), \( v \) are 3-tuples and \( 0 = (0,0,0) \).
(c) Repeat part (a), but where \( u \), \( v \) are 2-tuples and \( 0 = (0,0) \).

9.5 Dot Product, Norm, and Angle for \( n \)-Space

9.5.1. Dot product, norm, and angle. We wish to define the norm of an \( n \)-tuple vector, and the dot product and angle between two \( n \)-tuple vectors, just as we did for "arrow vectors." These definitions should be expressed in terms of the components of the \( n \)-tuples since the graphical and geometrical arguments used for arrow vectors will not be possible here for \( n > 3 \). Thus, if \( u = (u_1, \ldots, u_n) \), we wish to define the norm or "length" of \( u \), denoted as \( ||u|| \), in terms of the components \( u_1, \ldots, u_n \) of \( u \); and given another vector \( v = (v_1, \ldots, v_n) \), we wish to define the angle \( \theta \) between \( u \) and \( v \), and the dot product \( u \cdot v \), in terms of \( u_1, \ldots, u_n \) and \( v_1, \ldots, v_n \).

Furthermore, we would like these definitions to reduce to the definitions given in Sections 9.2 and 9.3 in the event that \( n = 2 \) or 3.

Let us begin with the dot product. Our plan is to return to the arrow vector formula

\[
u \cdot v = ||u|| ||v|| \cos \theta,
\]

(1)
to re-express it in terms of vector components for $\mathbb{R}^2$ and $\mathbb{R}^3$, and then to generalize those forms to $\mathbb{R}^n$.

If $u$ and $v$ are vectors in $\mathbb{R}^2$ as shown in Fig. 1, formula (1) may be expressed in terms of the components of $u$ and $v$ as follows:

$$u \cdot v = \|u\| \|v\| \cos \theta$$
$$= \|u\| \|v\| \cos (\beta - \alpha)$$
$$= \|u\| \|v\| (\cos \beta \cos \alpha + \sin \beta \sin \alpha)$$
$$= (\|u\| \cos \alpha ) (\|v\| \cos \beta) + (\|u\| \sin \alpha)(\|v\| \sin \beta)$$
$$= u_1v_1 + u_2v_2.$$  

(2)

We state, without derivation, that the analogous result for $\mathbb{R}^3$ is

$$u \cdot v = u_1v_1 + u_2v_2 + u_3v_3.$$  

(3)

Generalizing (2) and (3) to $\mathbb{R}^n$, it is eminently reasonable to define the (scalar-valued) dot product of two $n$-tuple vectors $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n)$ as

$$u \cdot v \equiv u_1v_1 + u_2v_2 + \cdots + u_nv_n = \sum_{j=1}^{n} u_jv_j.$$  

(4)

Observe that we have not proved (4); it is a definition.

Defining the dot product is the key, for now $\|u\|$ and $\theta$ follow readily. Specifically, we define

$$\|u\| \equiv \sqrt{u \cdot u} = \sqrt{\sum_{j=1}^{n} u_j^2}.$$  

(5)

in accordance with equation (6) in Section 9.3, and

$$\theta \equiv \cos^{-1} \left( \frac{u \cdot v}{\|u\| \|v\|} \right).$$  

(6)

from (1), where the inverse cosine is understood to be in the interval $[0, \pi]$. Notice the generalized Pythagorean-theorem nature of (5).

Other dot products and norms are sometimes defined for $n$-space, but we choose to use (4) and (5), which are known as the Euclidean dot product and Euclidian norm, respectively. To signify that the Euclidean dot product and norm have been adopted, we henceforth refer to the space as Euclidean $n$-space, rather

---

*By the “interval $[a, b]$ on a real $x$ axis,” we mean the points $a \leq x \leq b$. Such an interval is said to be closed since it includes the two endpoints. To denote the open interval $a < x < b$, we write $(a, b)$. Similarly, $[a, b)$ means $a \leq x < b$, and $(a, b]$ means $a < x \leq b$. Implicit in the closed-interval notation $[a, b]$ is the finiteness of $a$ and $b$. /
than just \( n \)-space. We will still denote it by the symbol \( \mathbb{R}^n \) (although some authors prefer the notation \( \mathbb{R}^n \)).

**EXAMPLE 1.** Let \( u = (1, 0) \) and \( v = (2, -2) \). Then
\[
\begin{align*}
\mathbf{u} \cdot \mathbf{v} &= (1)(2) + (0)(-2) = 2, \\
||\mathbf{u}|| &= \sqrt{(1)^2 + (0)^2} = 1, \\
||\mathbf{v}|| &= \sqrt{(2)^2 + (-2)^2} = 2\sqrt{2}, \\
\theta &= \cos^{-1} \left( \frac{2}{2\sqrt{2}} \right) = \frac{\pi}{4} \text{ (or } 45^\circ) ,
\end{align*}
\]
as is readily verified if we sketch \( \mathbf{u} \) and \( \mathbf{v} \) as arrow vectors in a Cartesian plane. \( \blacksquare \)

**EXAMPLE 2.** Let \( u = (2, -2, 4, -1) \) and \( v = (5, 9, -1, 0) \). Then,
\[
\begin{align*}
\mathbf{u} \cdot \mathbf{v} &= (2)(5) + (-2)(9) + (4)(-1) + (-1)(0) = -12, \\
||\mathbf{u}|| &= \sqrt{(2)^2 + (-2)^2 + (4)^2 + (-1)^2} = 5, \\
||\mathbf{v}|| &= \sqrt{(5)^2 + (9)^2 + (-1)^2 + (0)^2} = \sqrt{107}, \\
\theta &= \cos^{-1} \left( \frac{-12}{5\sqrt{107}} \right) \approx \cos^{-1} (-0.232) \approx 1.805 \text{ (or } 103.4^\circ) .
\end{align*}
\]

In this case, \( n (= 4) \) is greater than 3 so (7) through (10) are not to be understood in any physical or graphical sense, but merely in terms of the definitions (4) to (6).

**COMMENT.** The dot product of \( \mathbf{u} = (2, -2, 4) \) and \( \mathbf{v} = (5, 9, -1, 0) \), on the other hand, is **not defined** since here \( \mathbf{u} \) and \( \mathbf{v} \) are members of different spaces, \( \mathbb{R}^3 \) and \( \mathbb{R}^4 \), respectively. It is not legitimate to augment \( \mathbf{u} \) to the form \( (2, -2, 4, 0) \) on the grounds that “surely adding a zero can’t hurt.” \( \blacksquare \)

There is one catch that you may have noticed: (6) serves to define a (real) \( \theta \) only if the argument of the inverse cosine is less than or equal to unity in magnitude. That this is indeed true is not so obvious. Nevertheless, that
\[
-1 \leq \frac{\mathbf{u} \cdot \mathbf{v}}{||\mathbf{u}|| ||\mathbf{v}||} \leq 1 \quad \text{or} \quad ||\mathbf{u} \cdot \mathbf{v}|| \leq ||\mathbf{u}|| ||\mathbf{v}|| \tag{11}
\]
does necessarily hold will be proved in a moment. Whereas double braces denote vector norm, the single braces in (11) denote the absolute value of the scalar \( \mathbf{u} \cdot \mathbf{v} \).

**9.5.2. Properties of the dot product.** The dot product defined by (4) possesses the following important properties:

- **Commutative:** \( \mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u} \), \( \tag{12a} \)
- **Nonnegative:** \( \mathbf{u} \cdot \mathbf{u} > 0 \) for all \( \mathbf{u} \neq 0 \) \( = 0 \) for \( \mathbf{u} = 0 \), \( \tag{12b} \)
- **Linear:** \( (\alpha \mathbf{u} + \beta \mathbf{v}) \cdot \mathbf{w} = \alpha (\mathbf{u} \cdot \mathbf{w}) + \beta (\mathbf{v} \cdot \mathbf{w}) \), \( \tag{12c} \)
for any scalars $\alpha, \beta$ and any vectors $u, v, w$. The linearity condition (12c) is equivalent to the two conditions $(u + v) \cdot w = (u \cdot w) + (v \cdot w)$ and $(\alpha u) \cdot v = \alpha (u \cdot v)$. Verification of these claims is left for the exercises.

**EXAMPLE 3.** Expand the dot product $(6t - 2u) \cdot (v + 4w)$. Using (12), we obtain

\[
(6t - 2u) \cdot (v + 4w) = 6(t \cdot (v + 4w)) - 2(u \cdot (v + 4w))
\]

by (12c)

\[
= 6[(v + 4w) \cdot t] - 2[(v + 4w) \cdot u]
\]

by (12a)

\[
= 6(v \cdot t) + 24(w \cdot t) - 2(v \cdot u) - 8(w \cdot u)
\]

by (12c)

in much the same way that we obtain $(a - b)(e + d) = ae + ad - be - bd$ in scalar arithmetic.

As a consequence of (12) we are in a position to prove the promised inequality (11), namely, the **Schwarz inequality**

\[
|u \cdot v| \leq \|u\| \|v\|
\]

(13)

To derive this result, we start with the inequality

\[
(u + \alpha v) \cdot (u + \alpha v) \geq 0,
\]

(14)

which is guaranteed by (12b), for any scalar $\alpha$ and any vectors $u$ and $v$. Expanding the left-hand side and noting that $u \cdot u = \|u\|^2$ and $v \cdot v = \|v\|^2$, (14) becomes

\[
\|u\|^2 + 2au \cdot v + \alpha^2 \|v\|^2 \geq 0.
\]

(15)

Regarding $u$ and $v$ as fixed and $\alpha$ as variable, the left-hand side is then a quadratic function of $\alpha$. If we choose $\alpha$ so as to minimize the left-hand side, then (15) will be as close to an equality as possible and hence as informative as possible. Thus, setting $d$ (left-hand side)/$d\alpha = 0$, we obtain

\[
2u \cdot v + 2\alpha \|v\|^2 = 0 \quad \text{or} \quad \alpha = -\frac{u \cdot v}{\|v\|^2}.
\]

Putting this optimal value of $\alpha$ back into (15) gives us

\[
\|u\|^2 - 2\frac{(u \cdot v)^2}{\|v\|^2} + \frac{(u \cdot v)^2}{\|v\|^2} \geq 0,
\]

\[
\|u\|^2 \|v\|^2 - 2(u \cdot v)^2 + (u \cdot v)^2 \geq 0,
\]

\[
\|u\|^2 \|v\|^2 \geq (u \cdot v)^2,
\]

*After Hermann Amandus Schwarz (1843–1921). The names Cauchy and Bunyakovsky are also associated with this well-known inequality.

\[\text{Does a term such as } \alpha u \cdot v \text{ in (15) mean } (\alpha u) \cdot v \text{ or } (u \cdot v)? \text{ It does not matter; by virtue of (12c) (with } \beta = 0 \text{ and } w \text{ changed to } v), (\alpha u) \cdot v = \alpha (u \cdot v), \text{ so the parentheses are not needed.}**
and taking square roots of both sides yields the Schwarz inequality (13). Thus, it was not merely a matter of luck that the arguments of the inverse cosines were smaller than unity in magnitude in Examples 1 and 2, it was guaranteed in advance by the Schwarz inequality (13).

9.5.3. **Properties of the norm.** Since the norm is related to the dot product according to

\[ \|u\| = \sqrt{u \cdot u}, \]  

(16)

the properties (12) of the dot product should imply certain corresponding properties of the norm. These properties are as follows:

- **Scaling:** \[ \|\alpha u\| = |\alpha| \|u\|, \]  
  (17a)

- **Nonnegative:** \[ \|u\| \geq 0 \text{ for all } u \neq 0 \]  
  (17b)

\[ \|u\| = 0 \text{ for } u = 0, \]

- **Triangle Inequality:** \[ \|u + v\| \leq \|u\| + \|v\|. \]  
  (17c)

Equation (17a) simply says that \(\alpha u\) is \(|\alpha|\) times as long as \(u\), and for arrow representations of 2-tuples or 3-tuples the triangle inequality (17c) amounts to the Euclidean proposition that the length of any one side of a triangle cannot exceed the sum of the lengths of the other two sides (Fig. 2). Less obvious, however, is the fact that (17c) holds for \(n\)-tuples for \(n's > 3\).

Let us prove only (17a) and (17c) since (17b) follows readily from (16) and (12b). First, (17a):

\[ \|\alpha u\| = \sqrt{(\alpha u) \cdot (\alpha u)} \]  

by (16)

\[ = \sqrt{\alpha u \cdot (\alpha u)} \]  

by (12c) with \(\beta = 0\) and \(w = \alpha u\)

\[ = \sqrt{\alpha (\alpha u) \cdot u} \]  

by (12a)

\[ = \sqrt{\alpha^2 u \cdot u} \]  

by (12c) with \(\beta = 0\) and \(w = u\)

\[ = |\alpha| \sqrt{u \cdot u} = |\alpha| \|u\|. \]

Turning to (17c), we find that

\[ \|u + v\|^2 = (u + v) \cdot (u + v) \]  

by (16)

\[ = u \cdot u + v \cdot u + u \cdot v + v \cdot v \]

\[ = \|u\|^2 + 2u \cdot v + \|v\|^2 \]

\[ \leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 \]

\[ \leq \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 \]  

by (13)

\[ = (\|u\| + \|v\|)^2 \]

\[ * \text{That the choice } \alpha = -u \cdot v/\|v\|^2 \text{ minimizes the left-hand side of (15) follows from the fact that } d^2(\text{left-hand side})/d\alpha^2 = 2\|v\|^2 > 0. \]
so that
\[ \|u + v\| \leq \|u\| + \|v\|, \]
as claimed. A key step was the use of the Schwarz inequality (13), but we also used
the simple inequality \( u \cdot v \leq |u \cdot v| \), which holds since \( u \cdot v \) is a (positive, zero, or
negative) real number; that is, if \( u \cdot v \) is negative, then the \(<\) holds, and if \( u \cdot v \) is
zero or positive, then the \(=\) holds.

**EXAMPLE 4.** Let us verify the triangle inequality for a specific example, say the vectors
\( u = (2, 1, 3, -1) \) and \( v = (0, 4, 2, 1) \). Then \( u + v = (2, 5, 5, 0) \) so (17c) becomes
\[ \sqrt{54} \leq \sqrt{15} + \sqrt{21} \]
or \( 7.348 \leq 3.873 + 4.583 \), which is indeed true. \( \blacksquare \)

9.5.4. **Orthogonality.** If \( u \) and \( v \) are nonzero vectors such that \( u \cdot v = 0 \), then
\[ \theta = \cos^{-1} \left( \frac{u \cdot v}{\|u\| \|v\|} \right) = \cos^{-1} \left( \frac{0}{\|u\| \|v\|} \right) = \cos^{-1} (0) = \frac{\pi}{2}, \quad (18) \]
and we say that \( u \) and \( v \) are **perpendicular**. \([\text{Here we have used the nonzeroness of } u \text{ and } v \text{ in the third equality in } (18); \text{if } u \text{ and/or } v \text{ were } 0, \text{ we would have had } \cos^{-1} (0/\|u\| \|v\|) = \cos^{-1} (0/0), \text{ which is not defined.}]\]

But to equate the condition \( u \cdot v = 0 \) to perpendicularity (\( \theta = \pi/2 \)) would not
be correct since \( u \cdot v \) will also be zero in the event that \( u \) and/or \( v \) are \( 0 \), in which
case \( \theta \) is not defined. Let us therefore make a distinction between perpendicularity and
“orthogonality.” We will say that \( u \) and \( v \) are **orthogonal** if
\[ u \cdot v = 0. \quad (19) \]

Only if \( u \) and \( v \) are both nonzero does their orthogonality imply their perpendicularity (i.e., \( \theta = \pi/2 \)). With this definition, we see that the zero vector \( 0 \) is
orthogonal to every vector including itself (Exercise 14).

Finally, we say that a set of vectors, say \( \{u_1, \ldots, u_k\} \), is an **orthogonal set** if
every vector in the set is orthogonal to every other one:
\[ u_i \cdot u_j = 0 \quad \text{if } i \neq j. \quad (20) \]

**EXAMPLE 5.** \( u_1 = (2, 3, -1, 0), \ u_2 = (1, 2, 8, 3), \ u_3 = (9, -6, 0, 1) \) is an orthogonal
set because \( u_1 \cdot u_2 = u_1 \cdot u_3 = u_2 \cdot u_3 = 0. \) \( \blacksquare \)

**EXAMPLE 6.** \( u_1 = (1, 3), \ u_2 = (0, 0) \) is an orthogonal set because \( u_1 \cdot u_2 = 0. \) \( \blacksquare \)
9.5. Dot Product, Norm, and Angle for n-Space

9.5.5. Normalization. Any nonzero vector \( \mathbf{u} \) can be scaled to have unit length by multiplying it by \( 1/\|\mathbf{u}\| \) so we say that the vector

\[
\hat{\mathbf{u}} = \frac{1}{\|\mathbf{u}\|} \mathbf{u}
\]

has been "normalized." That \( \hat{\mathbf{u}} \) has unit length is readily verified:

\[
\|\hat{\mathbf{u}}\| = \left| \frac{1}{\|\mathbf{u}\|} \mathbf{u} \right| = \frac{1}{\|\mathbf{u}\|} \|\mathbf{u}\| \quad \text{by (17a)}
\]

\[
= \frac{1}{\|\mathbf{u}\|} \|\mathbf{u}\| \quad \text{by (17b)}
\]

\[
= 1.
\]

A vector of unit length is called a unit vector. We will often use the caret notation \( \hat{\mathbf{u}} \) for unit vectors.

**EXAMPLE 7.** Normalize \( \mathbf{u} = (1, -1, 0, 2) \). Since \( \|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}} = \sqrt{6} \), we have

\[
\hat{\mathbf{u}} = \frac{1}{\|\mathbf{u}\|} \mathbf{u} = \frac{1}{\sqrt{6}} (1, -1, 0, 2) = \left( \frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, 0, \frac{2}{\sqrt{6}} \right).
\]

A set of vectors is said to be orthonormal if it is orthogonal and if each vector is normalized (i.e., is a unit vector). We will use that term so frequently that it will be useful to abbreviate it as ON, but be aware that that abbreviation is not standard. Thus, \( \{\mathbf{u}_1, \ldots, \mathbf{u}_k\} \) is ON if and only if \( \mathbf{u}_i \cdot \mathbf{u}_j = 0 \) whenever \( i \neq j \) (for orthogonality), and \( \mathbf{u}_j \cdot \mathbf{u}_j = 1 \) for each \( j \) (so \( \|\mathbf{u}_j\| = 1 \), so the set is normalized).

The symbol

\[
\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}
\]

will be useful, and is known as the Kronecker delta, after Leopold Kronecker (1823–1891). Thus, \( \{\mathbf{u}_1, \ldots, \mathbf{u}_k\} \) is ON if and only if

\[
\mathbf{u}_i \cdot \mathbf{u}_j = \delta_{ij}
\]

for \( i = 1, 2, \ldots, k \) and \( j = 1, 2, \ldots, k \).

**EXAMPLE 8.** Let

\[
\mathbf{u}_1 = (1, 0, 0, 0), \quad \mathbf{u}_2 = \left( 0, \frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}} \right), \quad \mathbf{u}_3 = \left( 0, \frac{1}{\sqrt{2}}, 0, -\frac{1}{\sqrt{2}} \right).
\]
Then \( \{u_1, u_2, u_3\} \) is ON because \( \|u_1\| = \|u_2\| = \|u_3\| = 1 \) and \( u_1 \cdot u_2 = u_1 \cdot u_3 = u_2 \cdot u_3 = 0. \)

**Closure.** In this section we introduce a dot product \( u \cdot v \), a norm \( \|u\| \), and an angle \( \theta \) between two vectors for \( n \)-space. Their introduction is not a matter of derivation but, rather, a matter of definition. The definitions are designed as extensions of the definitions for the familiar “arrow” vectors of 2- and 3-space, somewhat as the upper floors of a home are built upon the foundation rather than being placed on an adjacent lot. Those extensions become apparent once we express the dot product, norm, and angle for arrow vectors in \( n \)-tuple notation. The key is the definition

\[
u \cdot v = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n,
\]

because then the arrow vector formula \( \|u\| = \sqrt{u \cdot u} \) gives us a norm, and the arrow vector formula \( \theta = \cos^{-1} \left( \frac{u \cdot v}{\|u\| \|v\|} \right) \) gives us an angle \( \theta \) between \( u \) and \( v \).

From these definitions we then derive the properties (12a,b,c) of the dot product and (17a,b,c) of the norm. In 2- and 3-space the triangle inequality amounts to the familiar Euclidean proposition that the length of any one side of a triangle cannot exceed the sum of the lengths of the other two sides, but in \( n \)-space, for \( n > 3 \), it amounts to an abstract generalization of that notion and does not have such a realizable physical or geometrical interpretation.

**EXERCISES 9.5**

1. Given the following vectors \( u \) and \( v \), determine \( \|u\|, \|v\| \) and \( \theta \) (in radians and degrees). If \( u \) and \( v \) are orthogonal, state that.

   \[\begin{align*}
   &\text{(a) } u = (4, 3), \quad v = (2, -1) \\
   &\text{(b) } u = (1, 2, 3, 4), \quad v = (-4, -3, -2, -1) \\
   &\text{(c) } u = (3, 0, 1), \quad v = (-2, 3, 6) \\
   &\text{(d) } u = (2, 2, 2), \quad v = (-4, -5, -6) \\
   &\text{(e) } u = (2, 5), \quad v = (10, -4) \\
   &\text{(f) } u = (1, 2, 3, 4), \quad v = (4, 3, 2, 1) \\
   &\text{(g) } u = (3, 2, 0, -1, 1), \quad v = (-5, 0, 0, 2, 4)
   \end{align*}\]

2. State whether or not each of the following expressions is defined.

   \[\begin{align*}
   &\text{(a) } \|u\| u \\
   &\text{(b) } u \cdot (v \cdot w) \\
   &\text{(c) } (u + v) \cdot (u - v) \\
   &\text{(d) } (u + v) \cdot w \\
   &\text{(e) } \cos^{-1}(2u + v) \\
   &\text{(f) } u + 6(v \cdot w) \\
   &\text{(g) } \cos^{-1}(u + v) \\
   &\text{(h) } \frac{u}{\|u\|^2} \\
   &\text{(i) } \frac{7u}{\cdot 2v} \\
   &\text{(j) } \|u + 3u^2\|
   \end{align*}\]

3. Let us denote, as points in 2- and 3-space, \( A = (2, 0), B = (3, -1), C = (5, 0), D = (4, 2), E = (2, 2), F = (1, 3, -2), G = (2, 0, 4), H = (5, 4, 3), I = (-3, -1, 0), J = (0, 0, 0) \). Determine, by vector methods, all interior angles and their sum, in degrees, for each of the following polygons.

   \[\begin{align*}
   &\text{(a) } ABCA \\
   &\text{(b) } ABCDA \\
   &\text{(c) } ABCDEA \\
   &\text{(d) } BCDB \\
   &\text{(e) } BCDEB \\
   &\text{(f) } FGHF \\
   &\text{(g) } FGIF \\
   &\text{(h) } GHIG \\
   &\text{(i) } FGIE \\
   &\text{(j) } GHJG \\
   &\text{(k) } H1JH \\
   &\text{(l) } F1JF
   \end{align*}\]

4. (a)–(g) Normalize each pair of \( u, v \) vectors in Exercise 1; that is, obtain \( \hat{u} \) and \( \hat{v} \).

5. If vectors \( A, B, C \) represented as arrows, form a triangle such that \( A = B + C \), derive the law of cosines \( C^2 = A^2 + B^2 - 2AB \cos \alpha \), where \( \alpha \) is the interior angle between \( A \) and \( B \), and where \( A, B, C \) are the lengths of \( A, B, C \), respectively, by starting with the identity \( C \cdot C = (A - B) \cdot (A - B) \).

6. (Orthogonalization) In each of following, find scalars \( \alpha, \beta, \gamma \) and vectors \( u_1, u_2, u_3 \) such that \( u_1 = u, u_2 = u + \alpha v, u_3 = u + \beta v + \gamma w \) is a nonzero orthogonal set, that is,
9.5. Dot Product, Norm, and Angle for n-Space 429

(a) Show that $u_1$ and $u_2$ can be found, in terms of $u$ and $v$, as
$$u_1 = (u \cdot \bar{v}) \bar{v} \quad \text{where} \quad \bar{v} = v / \|v\|,$$
$$u_2 = u - u_1.$$

Is (10.1) valid only for 2- and 3-space, or does it hold, without modification, for n-space as well? Explain.

(b) Use (10.1) to carry out the separation for the cases where $u = (1, 3)$ and $v = (2, 4)$, and where $u = (1, 3)$ and $v = (-1, -2)$. Interpret your results graphically for each of these cases.

(c) Use (10.1) to carry out the separation for $u = (2, 3, 1)$ and $v = (0, 2, 3)$.

(d) Repeat part (c), for $u = (1, 2, -1), v = (3, -1, 1)$.

(e) Repeat part (c), for $u = (3, 0, 5, 6), v = (1, -2, 0, 4)$.

(f) Repeat part (c), for $u = (2, 1, 0, 0, 3), v = (0, 0, 1, -2, 1)$.

11. (a) Prove the associative property $(uv) \cdot v = \alpha(u \cdot v)$.
(b) Prove the distributive property $(u + v) \cdot w = u \cdot w + v \cdot w$.
(c) Prove that the linearity property (12c) is equivalent to the two properties given in parts (a) and (b).

12. (Direction cosines) The direction cosines of a vector $u = (u_1, u_2, u_3)$ in 3-space are defined as $l_1 \equiv \cos \alpha$, $l_2 \equiv \cos \beta$, $l_3 \equiv \cos \gamma$, where $\alpha, \beta, \gamma$ are the angles between $u$ and the positive coordinate axes, as shown.

(a) Obtain general expressions for $l_1, l_2, l_3$ in terms of the components $u_1, u_2, u_3$.
(b) Evaluate $l_1, l_2, l_3$ for $u = (2, -1, 5)$.
(c) Evaluate $l_1, l_2, l_3$ for $u = (2, -4, 1)$.
(d) Evaluate $l_1, l_2, l_3$ for $u = (4, 0, -3)$.
(e) Show that $l_1^2 + l_2^2 + l_3^2 = 1$.

13. If $u \cdot v = 0$ and $v \cdot w = 0$, does that imply that $u \cdot w = 0$? Prove or disprove. HINT: If a claim is true, it needs to be proved in general, that is, for all possible cases. But if it is false, it can be disproved merely by putting forward a single counterexample.
14. Determine whether or not each set of vectors is orthogonal.
   (a) \( (1,3), (-6,2), (0,0) \)
   (b) \( (2,3,0), (-3,2,1), (1,1,1), (1,-3,1) \)
   (c) \( (1,0,0,0), (0,1,0,0), (0,0,1,0), (0,0,0,1) \)
   (d) \( (1,1,1,1), (1,-1,1,-1), (0,1,0,-1), (2,0,-2,0) \)
   (e) \( (2,1,-1,1), (1,1,3,0), (1,-1,0,-1), (2,1,1,1) \)

15. Determine a unit vector along the line of intersection of the following two planes in \( \mathbb{R}^3 \). NOTE: Do not use “cross products” since this topic has not yet been discussed here.
   (g) \( x_1 + 2x_2 - x_3 = 8 \)
   \( x_1 - x_2 + x_3 = 0 \)
   (h) \( x_1 + x_2 = 0 \)
   \( x_1 - x_2 + 2x_3 = 0 \)
   (i) \( x_1 - x_2 - 5x_3 = 0 \)
   \( x_2 + 4x_3 = 6 \)
   (j) \( 2x_1 + x_3 = 1 \)
   \( x_1 + 4x_2 = 1 \)

(c) \( x_1 - 5x_2 + x_3 = 4 \)
\( 2x_1 - x_2 - x_3 = 3 \)

(f) \( x_1 + x_2 + 12x_3 = 0 \)
\( x_1 + 2x_2 + 12x_3 = 5 \)

(g) \( x_1 - x_2 - x_3 = 2 \)
\( x_1 - x_2 - 2x_3 = 5 \)

16. (Schwarz inequality) To make (15) as close to an equality as possible, and hence as informative as possible, we minimized the left-hand side by setting \( d(\text{left-hand side})/d\alpha = 0 \). That step gave \( \alpha = -u \cdot v/\|v\|^2 \), and putting that result back into (15) gave the Schwarz inequality. That proof is valid for \( \mathbb{R}^n \) for any \( n \geq 1 \). For the special case of \( \mathbb{R}^2 \), show that the optimal \( \alpha \) is \( -u \cdot v/\|v\|^2 \) by using a graphical approach; that is, using a suitable sketch. HINT: Given \( u \) and \( v \), make \( u + \alpha v \) as short as possible.

9.6 Generalized Vector Space

9.6.1. Vector space. In Section 9.5 we generalize our vector concept from the familiar arrow vectors of 2- and 3-space to n-tuple vectors in abstract n-space, and it is n-space that is used in the remainder of this chapter and in Chapters 10–12. Yet, it is interesting to wonder if further generalization is possible. The answer is yes, and we will complete that story in this section. Far from being just a mathematical curiosity, the results will be essential in later chapters, when we study Fourier series, Sturm–Liouville theory, and partial differential equations.

The idea is as follows. In preceding sections we introduced the vectors and arithmetic rules for their manipulation, and then derived the various properties, such as \( u + v = v + u \), \( u + 0 = u \), \( \alpha(\beta u) = (\alpha \beta)u \), and so on. In generalizing, the essential idea is to reverse the cart and the horse. Specifically, we elevate the derived properties to axioms, or requirements, and regard the vectors as “objects,” the nature of which is not restricted in advance. They may be chosen to be n-tuples or whatever; all that we ask is that a plus (+) operation, a zero vector, a negative inverse, and scalar multiplication be defined such that all of the vector space axioms are satisfied. Thus:

**DEFINITION 9.6.1 Vector Space**
We call a (nonempty) set \( S \) of “objects,” which are denoted by **boldface type** and referred to as **vectors**, a **vector space** if the following requirements are met:
(i) An operation, which will be called vector *addition* and denoted as $+$, is defined between any two vectors in $S$ in such a way that if $u$ and $v$ are in $S$, then $u + v$ is too (i.e., $S$ is *closed under addition*). Furthermore,

$$u + v = v + u,$$

(commutative) \hspace{1cm} (1)

$$(u + v) + w = u + (v + w),$$

(associative) \hspace{1cm} (2)

(ii) $S$ contains a unique zero vector $0$ such that

$$u + 0 = u$$

for each $u$ in $S$.

(iii) For each $u$ in $S$ there is a unique vector $-u$ in $S$, called the *negative inverse* of $u$, such that

$$u + (-u) = 0.$$  \hspace{1cm} (4)

We denote $u + (-v)$ as $u - v$ for brevity, but emphasize that it is actually the $+$ operation between $u$ and $-v$.

(iv) Another operation, called *scalar multiplication*, is defined such that if $u$ is any vector in $S$ and $\alpha$ is any scalar, then the scalar multiple $\alpha u$ is in $S$, too (i.e., $S$ is *closed under scalar multiplication*). Further, we require that

$$\alpha(\beta u) = (\alpha\beta)u,$$  \hspace{1cm} (associative) \hspace{1cm} (5)

$$(\alpha + \beta)u = \alpha u + \beta u,$$  \hspace{1cm} (associative) \hspace{1cm} (6)

$$\alpha(u + v) = \alpha u + \alpha v,$$  \hspace{1cm} (distributive) \hspace{1cm} (7)

$$1u = u,$$  \hspace{1cm} (8)

if the vectors $u$, $v$ are in $S$, and $\alpha$, $\beta$ are scalars.

Observe that if we write $u + v + w$, it is not clear whether we mean $(u + v) + w$ (i.e., first add $u$ and $v$, and then add the result to $w$) or $u + (v + w)$. However, the associative property (2) guarantees that it does not matter, so the parentheses can be omitted without ambiguity. Similarly, $\alpha \beta u$ is unambiguous by virtue of (5).

**EXAMPLE 1.** $\mathbb{R}^n$-Space. Surely, the $n$-space $\mathbb{R}^n$, defined earlier, does constitute a vector space; after all, the axioms listed in Definition 9.6.1 come from the properties of $\mathbb{R}^n$ listed in Section 9.4. Thus, there is no need to check to see if those axioms are satisfied.

Instead, and for heuristic purposes, let us modify our addition operation from

$$u + v \equiv (u_1 + v_1, \ldots, u_n + v_n)$$  \hspace{1cm} (9)

We continue to restrict all scalars to be (finite) real numbers. Hence, we call the vector space a *real vector space*. 
to

\[ u + v \equiv (u_1 + 2v_1, \ldots, u_n + 2v_n), \]  

(10)
and see if (10) works; that is, let us see if the vector space axioms listed under (i) in Definition 9.6.1 are still satisfied if we use (10) as our addition operation instead of (9). According to (10),

\[ v + u \equiv (v_1 + 2u_1, \ldots, v_n + 2u_n) \]  

(11)
so a comparison of (10) and (11) shows that the commutativity axiom (1) is satisfied only if \( u_j + 2v_j = v_j + 2u_j \) (\( j = 1, \ldots, n \)), hence only if \( v_j = u_j \), hence only if \( v = u \). Since (1) does not hold for any chosen vectors \( u \) and \( v \), but only for vectors \( u \) and \( v \) that are equal, we conclude that if \( u + v \) is defined by (10), then we do not have a vector space. Of course, it is possible that (10) violates other axioms besides (1), but one failure is sufficient to show that the set is not a legitimate vector space.

COMMENT. Observe that we have not shown that \( u + v \) must be defined as in (9); conceivably,

\[ u + v \equiv (u_1^2 + v_1, \ldots, u_n^2 + v_n) \]  

(12)
or

\[ u + v \equiv (u_1 - v_1, \ldots, u_n - v_n) \]  

(13)
might work; that is, might satisfy the requirements listed under (i). Thus, understand that the plus signs on the left- and right-hand sides of (9) are not the same. The ones on the right denote the usual addition of real numbers (e.g., \( 2 + 5 = 7 \)), whereas the one on the left is more exotic; it denotes a certain operation between vectors \( u \) and \( v \), which is being defined by (9), or (10), or (12), or (13). To emphasize that point we could use a different notation such as \( u * v \), in place of \( u + v \), as some authors do. However, having made that point let us continue to use \( u + v \). \[ \]

\( \mathbb{R}^n \) is but one example of a vector space. Many other useful spaces can be introduced by using objects other than \( n \)-tuples as the vectors. For example, the vectors may be functions, matrices, or whatever, provided that vector addition, a zero vector, a negative inverse, and scalar multiplication are defined such that all of the vector space axioms are satisfied. For nowhere in Definition 9.6.1 is the nature of the vectors specified or in any way restricted.

**EXAMPLE 2.** A Function Space. This time, let the vectors be functions. Specifically, let \( u = u(x) \) be any continuous function defined on \( 0 \leq x \leq 1 \), say. For the addition operation let

\[ u + v \equiv u(x) + v(x); \]

that is, let \( u + v \) be the function whose values are the ordinary sum \( u(x) + v(x) \). For scalar multiplication let

\[ \alpha u \equiv \alpha u(x); \]

for the zero vector choose the zero function

\[ 0 \equiv 0; \]  

(14c)
and for the negative of \( u \) define

\[
-\mathbf{v} = -u(x); \tag{14d}
\]

that is, the function whose values are \(-u(x)\).

With these definitions, we can verify that all of the vector space requirements are satisfied, so that the set \( S \) of such vectors is a bona fide vector space. For instance, if \( \mathbf{u} = u(x) \) and \( \mathbf{v} = v(x) \) are continuous on \( 0 \leq x \leq 1 \), then so is \( \mathbf{u} + \mathbf{v} = u(x) + v(x) \) so \( S \) is closed under addition. Further, \( \mathbf{v} + \mathbf{u} = v(x) + u(x) = u(x) + v(x) = \mathbf{u} + \mathbf{v} \) so addition satisfies the commutative property (1), and so on.

This \( S \) is but one example of a function space, a space in which the vectors are functions. \( \blacksquare \)

The following theorem is useful, and its proof illustrates the axiomatic approach.

### Theorem 9.6.1 Properties of Scalar Multiplication

If \( \mathbf{u} \) is any vector in a vector space \( S \) and \( \alpha \) is any scalar, then

\[
\begin{align*}
0\mathbf{u} &= \mathbf{0}, \tag{15a} \\
(-1)\mathbf{u} &= -\mathbf{u}, \tag{15b} \\
\alpha \mathbf{0} &= \mathbf{0}. \tag{15c}
\end{align*}
\]

**Proof:** These results follow from our definition of vector space. To prove (15a), one line of approach is as follows:

\[
\begin{align*}
0\mathbf{u} + \mathbf{u} &= 0\mathbf{u} + \mathbf{1u} \quad \text{by (8)} \\
&= (0 + 1)\mathbf{u} \quad \text{by (6)} \\
&= \mathbf{1u} \\
&= \mathbf{u} \quad \text{by (8)}.
\end{align*}
\]

Then

\[
\begin{align*}
0\mathbf{u} + \mathbf{u} + (-\mathbf{u}) &= \mathbf{u} + (-\mathbf{u}) \\
0\mathbf{u} + 0 &= 0 \quad \text{by (4),}
\end{align*}
\]

The remaining two, (15b) and (15c), are left for the exercises. \( \blacksquare \)

### 9.6.2. Inclusion of inner product and/or norm

Observe that there is no mention of a dot product or a norm either in Definition 9.6.1 or in Examples 1 or 2. Indeed, a vector space \( S \) need not have a dot product (also called an inner product) or a norm defined for it. If it does have an inner product it is called an inner product.

*The second equality holds because \( v(x) + u(x) \) is the ordinary sum of two real numbers; e.g., \( 4 + 3 = 3 + 4 \).*
space; if it has a norm it is called a normed vector space; and if it has both it is called a normed inner product space.

If we do choose to introduce an inner product for \( S \), how is it to be defined? Do you remember the idea of reversing the cart and the horse? That is how we do it. Equations (12a,b,c) in Section 9.5.2 were shown to be properties of the inner product \( \mathbf{u} \cdot \mathbf{v} = u_1v_1 + \cdots + u_nv_n \). We now take those properties and elevate them to axioms, or requirements, that are to be satisfied by any inner product of any vector space.

Similarly, we take the properties (17a,b,c) of the norm, in Section 9.5.3, and elevate them to axioms, or requirements, that are to be satisfied by any norm of any vector space.

Let us tabulate them here:

**REQUIREMENTS OF INNER PRODUCT**

<table>
<thead>
<tr>
<th>Commutative:</th>
<th>( \mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u} ), \hspace{1cm} (16a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonnegative:</td>
<td>( \mathbf{u} \cdot \mathbf{u} &gt; 0 ) for all ( \mathbf{u} \neq \mathbf{0} ), \hspace{1cm} ( \mathbf{u} \cdot \mathbf{u} = 0 ) for ( \mathbf{u} = \mathbf{0} ), \hspace{1cm} (16b)</td>
</tr>
<tr>
<td>Linear:</td>
<td>( (\alpha \mathbf{u} + \beta \mathbf{v}) \cdot \mathbf{w} = \alpha (\mathbf{u} \cdot \mathbf{w}) + \beta (\mathbf{v} \cdot \mathbf{w}) ), \hspace{1cm} (16c)</td>
</tr>
</tbody>
</table>

and

**REQUIREMENTS OF NORM**

| Scaling:              | \( \|\alpha \mathbf{u}\| = |\alpha| \|\mathbf{u}\| \), \hspace{1cm} (17a) |
|-----------------------|----------------------------------------------------------------------------------|
| Nonnegative:          | \( \|\mathbf{u}\| > 0 \) for all \( \mathbf{u} \neq \mathbf{0} \), \hspace{1cm} \( \|\mathbf{u}\| = 0 \) for \( \mathbf{u} = \mathbf{0} \), \hspace{1cm} (17b) |
| Triangle Inequality:  | \( \|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\| \). \hspace{1cm} (17c) |

Let us illustrate.

**EXAMPLE 3.** \( \mathbb{R}^n \)-Space. If we wish to add an inner product to the vector space \( \mathbb{R}^n \), we can use the choice

\[
\mathbf{u} \cdot \mathbf{v} \equiv u_1v_1 + \cdots + u_nv_n = \sum_{j=1}^{n} u_jv_j.
\] \hspace{1cm} (18a)
9.6. Generalized Vector Space

We know that \( (18a) \) satisfies the requirements (16) because the latter were deduced, in Section 9.5.2, as properties that follow from \( (18a) \). A variation of \( (18a) \) that still satisfies \( (16) \) is (Exercise 6)

\[
u \cdot v \equiv u_1v_1 + \cdots + u_nv_n = \sum_{j=1}^{n} w_ju_jv_j, \tag{18b}\]

where the \( w_j \)'s are fixed positive constants known as "weights" because they attach more or less weight to the different components of \( u \) and \( v \). For instance, consider \( \mathbb{R}^2 \) and let \( w_1 = 5 \) and \( w_2 = 3 \). Then if \( u = (2, -4) \) and \( v = (1, 6) \) we have \( u \cdot v = 5(2)(1) + 3(-4)(6) = -62 \).

Note that for \( (18b) \) to be a legitimate inner product we must have \( w_j > 0 \) for each \( j \). For suppose, still in \( \mathbb{R}^2 \), that \( w_1 = 3 \) and \( w_2 = -2 \). Then, for \( u = (1, 5) \), say, we have \( u \cdot u = 3(1)(1) - 2(5)(5) = -47 < 0 \), in violation of \( (16b) \). Or, suppose that \( w_1 = 3 \) and \( w_2 = 0 \). Then, for \( u = (0, 4) \), say, we have \( u \cdot u = 3(0)(0) + 0(4)(4) = 0 \) even though \( u \neq 0 \), again in violation of \( (16b) \).

Now, suppose that we wish to add a norm. If for any vector space \( S \) we already have an inner product, then a legitimate norm can always be obtained from that inner product as \( \| u \| = \sqrt{u \cdot u} \), and that choice is called the natural norm. Thus, the natural norms corresponding to \( (18a) \) and \( (18b) \) are

\[
\| u \| \equiv \sqrt{\sum_{j=1}^{n} w_ju_j^2} \quad \text{and} \quad \| u \| \equiv \sqrt{\sum_{j=1}^{n} w_ju_j^2}, \tag{19a,b}\]

respectively.

However, we do not have to choose the natural norm. For instance, we could use \( (18a) \) as our inner product, and choose

\[
\| u \| \equiv |u_1| + \cdots + |u_n| = \sum_{j=1}^{n} |u_j| \tag{20}\]

as our norm (Exercise 8). The latter is used by Struble in his book on differential equations,* probably because it is algebraically simpler than the Euclidean norm \( (19a) \) or the modified Euclidean norm \( (19b) \). Furthermore, he defines no inner product whatsoever. Struble calls \( (20) \) the taxicab norm since a taxicab driver judges the distance from the corner of 5th Avenue and 34th Street to the corner of 2nd Avenue and 49th Street as 18 blocks, not \( \sqrt{234} \) blocks.

**EXAMPLE 4.** The Function Space of Example 2. How might we choose an inner product for the function space \( S \) defined in Example 2? To motivate our choice, let us imagine approximating any given function (i.e., vector) \( u(x) \) in \( S \) in a piecewise-constant manner as depicted in Fig. 1. That is, divide the \( x \) interval \( (0 \leq x \leq 1) \) into \( n \) equal parts and define the approximating piecewise-constant function, over each subinterval as

---

the value of \( u(x) \) at the left endpoint of that subinterval. If we represent the piecewise-constant function as the \( n \)-tuple \((u_1, \ldots, u_n)\), then we have, in a heuristic sense,
\[
u(x) \approx (u_1, \ldots, u_n).
\]
(21)
Similarly, for any other function \( v(x) \) in \( S \),
\[
v(x) \approx (v_1, \ldots, v_n).
\]
(22)

Figure 1. Staircase approximation of \( u(x) \).

The \( n \)-tuple vectors on the right-hand sides of (21) and (22) are members of \( \mathbb{R}^n \). For that space, let us adopt the inner product
\[
(u_1, \ldots, u_n) \cdot (v_1, \ldots, v_n) = \sum_{j=1}^{n} u_j v_j \Delta x,
\]
that is, (18b) with all of the \( w_j \) weights the same, namely, the subinterval width \( \Delta x \). If we let \( n \to \infty \), the "staircase approximations" approach \( u(x) \) and \( v(x) \), and the sum in (23) tends to the integral \( \int_0^1 u(x) v(x) \, dx \).

This heuristic reasoning suggests the inner product
\[
\langle u(x), v(x) \rangle \equiv \int_0^1 u(x) v(x) \, dx.
\]
(24a)
We can denote it as \( u \cdot v \) and call it the dot product, or we can denote it as \( \langle u(x), v(x) \rangle \) and call it the inner product. For function spaces, the latter notation is somewhat standard, and is our choice in this text.

COMMENT 1. By no means do we claim our staircase idea to be a rigorous derivation of (24a). In fact, it is neither rigorous nor a derivation: it is \textit{heuristic motivation} for the \textit{definition} (24a). We leave it for the exercises to verify that (24a) does satisfy the requirements (16).

COMMENT 2. Just as (18b) is a legitimate generalization of (18a), (if \( w_j > 0 \) for \( 1 \leq j \leq n \)), we expect that
\[
\langle u(x), v(x) \rangle \equiv \int_0^1 u(x) v(x) w(x) \, dx
\]
(24b)
is a legitimate generalization of (24a) [if \( w(x) > 0 \) for \( 0 \leq x \leq 1 \)], proof of which claim is left for the exercises. The inner product (24b) is prominent when we study Fourier series and the Sturm—Liouville theory in Chapter 17.

COMMENT 3. Naturally, if we wish to define a norm as well, we could use a natural norm based on (24a) or (24b), for instance

\[
\|u\| = \sqrt{\langle u(x), w(x) \rangle} = \sqrt{\int_0^1 u^2(x)w(x) \, dx}
\]

(25)
based on (24b).

COMMENT 4. Notice carefully that the concept of the dimension of a vector space has not yet been introduced, although it is in Section 9.10. There, we define dimension and find that \( \mathbb{R}^n \) is \( n \)-dimensional (which claim is probably not a great shock). Since the staircase approximation (21) becomes exact only as \( n \to \infty \), it appears that our function space \( S \) is infinite dimensional!

COMMENT 5. A bit of notation: the set of functions that are defined and continuous on \([0, 1]\) (i.e., \( 0 \leq x \leq 1 \)) is usually denoted as \( C^0[0, 1] \). If not only are the functions continuous but also all derivatives through order \( k \), then the set is denoted as \( C^k[0, 1] \).

Closure. Using \( n \)-space as a ladder, we complete our generalization of vector space by taking the properties of \( \mathbb{R}^n \) (such as \( u + v = v + u \)) and turning them into the axioms, or requirements, to be met by any vector space. Thus, attention shifted from the objects, the vectors, to those requirements. There is no restriction on the nature of the vectors, which can be arrows, \( n \)-tuples, matrices, functions, or oranges. For us, the most important vector spaces are \( \mathbb{R}^n \) and various function spaces: \( \mathbb{R}^n \) is used in the remainder of this chapter and Chapters 10-12, and function spaces are used in Chapter 17 when we study Fourier series and Sturm-Liouville theory.

To illustrate the power of the axiomatic approach, recall the Schwarz inequality \( |u \cdot v| \leq \|u\| \|v\| \), proved in Section 9.5.2 for \( \mathbb{R}^n \). That result holds for any normed inner product space with natural norm \( \|u\| = \sqrt{u \cdot u} \) for it followed from properties of \( \mathbb{R}^n \), which properties are subsequently elevated to axioms for general vector space. Thus, it represents many properties rolled into one. For example, in \( \mathbb{R}^n \), with the dot product (18a) it says

\[
\left| \sum_{j=1}^n u_j v_j \right| \leq \sqrt{\sum_{j=1}^n u_j^2} \sqrt{\sum_{j=1}^n v_j^2},
\]

(26)
in the function space of Examples 2 and 4; with the inner product (24b) and norm (25) it says

\[
\left| \int_0^1 u(x)v(x)w(x) \, dx \right| \leq \sqrt{\int_0^1 u^2(x)w(x) \, dx} \sqrt{\int_0^1 v^2(x)w(x) \, dx},
\]

(27)
and so on.
EXERCISES 9.6

1. Recall that $\mathbb{R}^n$ is the vector space (“real” vector space since all scalars are to be real numbers) in which the vectors are n-tuples $u = (u_1, \ldots, u_n)$, with the definitions

$$u + v = (u_1, \ldots, u_n) + (v_1, \ldots, v_n)$$

$$0 \equiv (0, \ldots, 0),$$

$$-u \equiv (-u_1, \ldots, -u_n),$$

$$\alpha u \equiv (\alpha u_1, \ldots, \alpha u_n).$$

If we make the following modifications, do we still have a vector space? If not, specify all requirements within Definition 9.6.1 that fail to be met.

(a) only vectors of the form $u = (u, u, \ldots, u)$ admitted, where $-\infty < u < \infty$
(b) only vectors of the form $u = (u, 2u, 3u, \ldots, nu)$ admitted, where $-\infty < u < \infty$
(c) only the vector $(0, 0, \ldots, 0)$ admitted (this is an example of a zero vector space, a vector space containing only the zero vector)
(d) $u + v \equiv (u_1 - v_1, \ldots, u_n - v_n)$, in place of (1.1)
(e) $u + v \equiv (0, 0, \ldots, 0)$ for all $u$'s and $v$'s, in place of (1.1)
(f) $\alpha u \equiv (\alpha^2 u_1, \ldots, \alpha^2 u_n)$, in place of (1.4)

2. We noted in Example 1 that the definition (10) of vector addition violates axiom (1). Does it violate any others as well? Explain.

3. Prove (15b), that $-u = (3, 1, -1, 0)$ and $v = (1, 2, 5, -4)$ in $\mathbb{R}^4$, with the inner product (18a)

4. Prove (15c), that $\alpha 0 = 0$.

5. Prove that if $\alpha u = 0$ then $\alpha = 0$ and/or $u = 0$.

6. Show that the inner product (18b) does satisfy the requirement (16).

7. We stated in Example 3 that if for any vector space $S$ we already have an inner product, then a legitimate norm can always be obtained from that inner product as $||u|| = \sqrt{u \cdot u}$, which choice is called the natural norm. Prove that claim.

8. Show that the “taxicab norm” (20) is a legitimate norm — that is, that it satisfies the requirements (17).

9. (a) Does the choice $||u|| = \max_{1 \leq j \leq n} |u_j|$, for $\mathbb{R}^n$, satisfy the requirements (17)? Explain.
(b) How about $||u|| = \min_{1 \leq j \leq n} |u_j|$, for $\mathbb{R}^n$?

10. Let $S$ be the set of real-valued polynomial functions, of degree $n$, defined on $a \leq x \leq b$. If $u = u_0 + a_1 x + \cdots + a_n x^n$ and $v = b_0 + b_1 x + \cdots + b_n x^n$ are any two such functions, and $\alpha$ is any (real) scalar, define the sum $u + v$ and the scalar multiple $\alpha u$ as

$$(u + v)(x) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n,$$

$$(\alpha u)(x) = \alpha a_0 + \alpha a_1 x + \cdots + \alpha a_n x^n,$$

respectively. Further, let $0$ be the function $0 + 0x + \cdots + 0x^n$, and let $-u$ be the function $-a_0 - a_1 x + \cdots - a_n x^n$. Show that $S$ is a vector space.

11. Show that the inner product (24b) does satisfy the requirements (16).

12. (Schwarz inequality) We derive the Schwarz inequality

$$|u \cdot v| \leq ||u|| ||v||$$

for $\mathbb{R}^n$ in Section 9.5.2. The latter holds not only for $\mathbb{R}^n$ but for any normed inner product space with the natural norm $||u|| = \sqrt{u \cdot u}$. In this exercise we simply ask you to verify (12.1) by working out the left- and right-hand sides for these specific cases:

(a) $u = (3, 1, -1, 0)$ and $v = (1, 2, 5, -4)$ in $\mathbb{R}^4$, with the inner product (18a)
(b) $u = (1, 2, 4, -3)$ and $v = (0, 4, 1, 1)$ in $\mathbb{R}^4$, with $u \cdot v = u_1 v_1 + 5u_2 v_2 + 3u_3 v_3 + 2u_4 v_4$
(c) $u = (1, 1, 1, 1)$ and $v = (2, 2, 2, 2)$ in $\mathbb{R}^4$, with $u \cdot v = u_1 v_1 + 2u_2 v_2 + 3u_3 v_3 + 4u_4 v_4 + 5u_5 v_5$
(d) $u = 2 + x$ and $v = 3x^2$ in the function space of Example 4, with the inner product $u \cdot v = \langle u(x), v(x) \rangle = \int_0^1 u(x)v(x) \, dx$
(e) Same as (d), but with $\langle u(x), v(x) \rangle = \int_0^1 u(x)v(x)(2 + 5x) \, dx$

13. (Solution space) (a) Consider a set of $m$ linear homogeneous algebraic equations in the $n$ unknowns $x_1, \ldots, x_n$, and denote each solution of the system as an $n$-tuple vector $x = (x_1, \ldots, x_n)$ in $\mathbb{R}^n$. Show that the set of all such vectors, with the usual definitions $u + v = (u_1 + v_1, \ldots, u_n + v_n)$, $\alpha u = (\alpha u_1, \ldots, \alpha u_n)$, $-u = (-u_1, \ldots, -u_n)$, $0 = (0, \ldots, 0)$, is a vector space. That space is called the solution space of the system.
(b) If the system is nonhomogeneous, is the set of solutions still a vector space? Explain.

14. (Solution space) Show that the solutions of a linear homogeneous differential equation (with the same definitions of \( u + v \), \( au \), \(-u \), and \( 0 \) as in Example 2) constitute a vector space, the so-called solution space of that differential equation.

9.7 Span and Subspace

Here, we begin a sequence of closely related ideas: span, linear dependence, basis, expansion, and dimension. The concepts, definitions, and theorems hold for any vector space, but our illustrative examples are restricted to the \( n \)-space \( \mathbb{R}^n \), this being the case of most interest in Chapters 9–12.

We begin with the idea of the “span” of a set of vectors.

**DEFINITION 9.7.1 Span**

If \( u_1, \ldots, u_k \) are vectors in a vector space \( S \), then the set of all linear combinations of these vectors, that is, all vectors of the form

\[ u = \alpha_1 u_1 + \cdots + \alpha_k u_k, \]

(1)

where \( \alpha_1, \ldots, \alpha_k \) are scalars is called the span of \( u_1, \ldots, u_k \) and is denoted as \( \text{span} \{ u_1, \ldots, u_k \} \).

The set \( \{ u_1, \ldots, u_k \} \) is called the generating set of \( \text{span} \{ u_1, \ldots, u_k \} \).

Let us illustrate with some vector sets in \( \mathbb{R}^2 \) and \( \mathbb{R}^3 \) so we can support the discussion with diagrams.

**EXAMPLE 1.** Determine the span of the single vector

\[ u_1 = (4, 2) \]

(2)

in \( \mathbb{R}^2 \). Then \( \text{span} \{ u_1 \} \) is the set of all vectors that are scalar multiples of \( u_1 \). Hence, \( \text{span} \{ u_1 \} \) is the set of all vectors on the line \( L \) in Fig. 1, such as \( u = 2u_1 = (8, 4) \), \( v = -\frac{1}{2} u_1 = (-2, -1) \), and \( 0 = 0u_1 = (0, 0) \). We say that \( u_1 \) generates the line \( L \).

**EXAMPLE 2.** Determine the span of the two vectors

\[ u_1 = (4, 2), \quad u_2 = (-8, -4). \]

(3)
Span \{\mathbf{u}_1, \mathbf{u}_2\} is, once again, the line \( L \) in Fig. 1 (i.e., the set of all vectors on \( L \)), for both \( \mathbf{u}_1 \) and \( \mathbf{u}_2 \) lie along \( L \), so any linear combination of them, \( \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 \), does too. Similarly, span \{\{(4, 2), (–8, –4), (18, 9), (0, 0)\}\} is the line \( L \).

Observe that the line \( L \), in Examples 1 and 2, is only a subset of the vector space \( \mathbb{R}^2 \). Observe that that subset of \( \mathbb{R}^2 \) is itself a vector space, a so-called “subspace” of \( \mathbb{R}^2 \). For if \( \mathbf{u} \) and \( \mathbf{v} \) are any two vectors on \( L \), then \( \mathbf{u} + \mathbf{v} \) is on \( L \), too, so the set is closed under addition; similarly, if \( \mathbf{u} \) is on \( L \), so is \( \alpha \mathbf{u} \), for any scalar \( \alpha \), so the set is closed under scalar multiplication: \( L \) does contain the zero vector [since we can set all the \( \alpha \)'s in (1) equal to zero]; and for each \( \mathbf{u} \) on \( L \) there is a (unique) vector \( -\mathbf{u} \) on \( L \) such that \( \mathbf{u} + (-\mathbf{u}) = \mathbf{0} \).

**Definition 9.7.2 Subspace**

If a subset \( T \) of a vector space \( S \) is itself a vector space (with the same definitions as \( S \) for vector addition \( \mathbf{u} + \mathbf{v} \), scalar multiplication \( \alpha \mathbf{u} \), zero vector \( \mathbf{0} \), and negative vector \( -\mathbf{u} \)), then \( T \) is a subspace of \( S \).

Usually, a subspace of \( S \) is only a part of \( S \), as the line \( L \) is only a part of \( \mathbb{R}^2 \), but since a subset of a set can be all of that set, a subspace of \( S \) can be all of \( S \). For instance, \( \mathbb{R}^2 \) is a subspace of \( \mathbb{R}^2 \).

**Theorem 9.7.1 Span as Subspace**

If \( \mathbf{u}_1, \ldots, \mathbf{u}_k \) are vectors in a vector space \( S \), then span \{\( \mathbf{u}_1, \ldots, \mathbf{u}_k \)\} is itself a vector space, a subspace of \( S \).

For instance, the line \( L \) in Fig. 1 is a subspace of \( \mathbb{R}^2 \). Proof of Theorem 9.7.1 is left for the exercises.

**Example 3.** Is the span of

\[
\mathbf{u}_1 = (5, 1), \quad \mathbf{u}_2 = (1, 3)
\]

all of \( \mathbb{R}^2 \) or only a part of \( \mathbb{R}^2 \)? To determine the extent of span \{\( \mathbf{u}_1, \mathbf{u}_2 \)\}, let \( \mathbf{v} = (v_1, v_2) \) be any given vector in \( \mathbb{R}^2 \), and try to express

\[
\mathbf{v} = \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2.
\]

That is,

\[
(v_1, v_2) = \alpha_1(5, 1) + \alpha_2(1, 3) = (5\alpha_1, \alpha_1) + (\alpha_2, 3\alpha_2) = (5\alpha_1 + \alpha_2, \alpha_1 + 3\alpha_2).
\]
Equating components, we obtain the linear equations
\[
\begin{align*}
5\alpha_1 + \alpha_2 &= v_1, \\
\alpha_1 + 3\alpha_2 &= v_2
\end{align*}
\]
(7)
in \(\alpha_1, \alpha_2\). Applying Gauss elimination, (7) becomes
\[
\begin{align*}
\alpha_1 + \frac{1}{5}\alpha_2 &= \frac{1}{5}v_1, \\
\alpha_2 &= \frac{5}{14}v_2 - \frac{1}{14}v_1.
\end{align*}
\]
(8)
It is clear from the Gauss-reduced form (8) that the system is consistent (solvable for \(\alpha_1, \alpha_2\)) for every vector \(v\) in \(\mathbb{R}^2\). Hence, we may conclude that span \(\{u_1, u_2\}\) is all of \(\mathbb{R}^2\); we say that \(\{u_1, u_2\}\) spans \(\mathbb{R}^2\). (Here we use “span” as a verb; in Definition 9.7.1 it is introduced as a noun.)

Thus, every \(v\) in \(\mathbb{R}^2\) can be expressed as a linear combination of vector \(u_1\) and \(u_2\). As representative, let \(v = (6, -4)\) so \(v_1 = 6\) and \(v_2 = -4\). Then (8) gives \(\alpha_2 = -\frac{11}{5}\) and \(\alpha_1 = \frac{11}{7}\), so that (5) becomes
\[
v = \frac{11}{7}u_1 - \frac{13}{7}u_2.
\]
(9)
To see this in graphical terms, observe from Fig. 2 that \(v = OA + OB\), where (with the aid of a scale) \(OA \approx 1.6u_1\) and \(OB \approx -1.9u_2\). Thus, \(v \approx 1.6u_1 - 1.9u_2\), in agreement with (9).

COMMENT. Suppose that we add \(u_3 = (2, 2)\) to the set. It should be evident that span \(\{u_1, u_2, u_3\}\) is all of \(\mathbb{R}^2\), again, since \(\{u_1, u_2\}\) spanned \(\mathbb{R}^2\) even “without any help” from \(u_3\). But in case this is not clear, let us go through steps analogous to steps (5) to (8):
\[
v = \alpha_1u_1 + \alpha_2u_2 + \alpha_3u_3
\]
(10)
so \((v_1, v_2) = (5\alpha_1 + \alpha_2 + 2\alpha_3, \alpha_1 + 3\alpha_2 + 2\alpha_3)\). Thus,
\[
\begin{align*}
5\alpha_1 + \alpha_2 + 2\alpha_3 &= v_1, \\
\alpha_1 + 3\alpha_2 + 2\alpha_3 &= v_2
\end{align*}
\]
or
\[
\begin{align*}
\alpha_1 + \frac{1}{5}\alpha_2 + \frac{2}{5}\alpha_3 &= \frac{1}{5}v_1, \\
\alpha_2 + \frac{3}{5}\alpha_3 &= \frac{5}{14}v_2 - \frac{1}{14}v_1.
\end{align*}
\]
(11)
Like (8), (11) is consistent for every \(v\) in \(\mathbb{R}^2\) so \(\{u_1, u_2, u_3\}\) spans \(\mathbb{R}^2\), as claimed. Whereas (8) had a unique solution so that the representation (5) was unique, (11) happens to have an infinity of solutions so that the representation (10) is not unique.

EXAMPLE 4. As a final example, consider the span of
\[
\begin{align*}
u_1 &= (1, 2, 2), \\
u_2 &= (-1, 0, 2)
\end{align*}
\]
(12)
in \(\mathbb{R}^3\). Setting
\[
v = \alpha_1u_1 + \alpha_2u_2,
\]
(13)
we have
\[ \alpha_1 - \alpha_2 = v_1, \]
\[ 2\alpha_1 = v_2, \]
\[ 2\alpha_1 + 2\alpha_2 = v_3, \]
or, after Gauss elimination,
\[ \alpha_1 - \alpha_2 = v_1, \]
\[ \alpha_2 = \frac{1}{2}v_2 - v_1, \]
\[ 0 = v_3 - 2v_2 + 2v_1. \]

Now, \( \text{span} \{ u_1, u_2 \} \) is the set of all possible vectors \( v \) given by (13), i.e., all vectors \( v \) for which the system (14) is consistent, i.e., all vectors \( v = (v_1, v_2, v_3) \) such that
\[ 2v_1 - 2v_2 + v_3 = 0 \]
[so that the last of equations (14) is \( 0 = 0 \) rather than a contradiction].

In geometrical terms, on the other hand, \( \text{span} \{ u_1, u_2 \} \) should be the subset of \( \mathbb{R}^3 \) consisting of the plane that passes through \( u_1 \) and \( u_2 \) (\( u_1 \) and \( u_2 \) are shown in Fig. 3). How does that fact correlate with (15)? As a matter of fact, (15) is the equation of a plane in 3-space, and that plane does pass through the origin, through the tip of \( u_1 \) [i.e., the point \((1, 2, 2)\)], and through the tip of \( u_2 \) [the point \((-1, 0, 2)\)]. Hence, it is the plane through \( u_1 \) and \( u_2 \) so the analytical approach, namely, steps (13) to (15) and our geometrical interpretation are in agreement.

We conclude that \( \text{span} \{ u_1, u_2 \} \) is not all of \( \mathbb{R}^3 \); it is only the subspace of \( \mathbb{R}^3 \) consisting of the plane (i.e., all vectors in the plane) containing the given vectors \( u_1 \) and \( u_2 \).

COMMENT. Since \( \text{span} \{ u_1, u_2 \} \) is a plane, would it be correct to say that \( \text{span} \{ u_1, u_2 \} \) is \( \mathbb{R}^2 \)? No, that would be incorrect; \( \mathbb{R}^2 \) is made up of two-tuples, while the vectors in the above-mentioned plane are three-tuples. Thus, \( \mathbb{R}^2 \) space is not relevant in this problem. All that can be said here is that \( \text{span} \{ u_1, u_2 \} \) is the subspace of \( \mathbb{R}^3 \) consisting of the plane containing the vectors \( u_1 \) and \( u_2 \), that is, the plane defined by (15).

Closure. In leading up to the concept of bases and expansions, the two key ideas are span and linear independence. In this section we introduce the idea of span; in the next section we introduce linear dependence and linear independence. Although the concept of span holds for any vector space, such as \( \mathbb{R}^6 \), we suggest that you focus on the foregoing examples in two- and three-spaces, so that you can use the two- and three-dimensional drawings to promote understanding.
EXERCISES 9.7

1. Show whether the vectors
   (a) \((1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), \ldots, (0, 0, 0, 0, 1)\) span \(\mathbb{R}^n\)
   (b) \((0, 0, 0, 1), (0, 0, 1, 1), (0, 1, 1, 1)\) span \(\mathbb{R}^4\)
   (c) \((1, 2, 0, 4), (2, 3, 1, -1), (1, 0, 0, 1), (0, 0, 0, 0), (1, 1, 2, 3)\)
   (d) \((1, 3, 2, 2), (5, 7, 1, 0), (-1, -2, -4, 3)\) span \(\mathbb{R}^4\)
   (e) \((1, 0, 1), (2, 1, -1), (1, 2, -5)\) span \(\mathbb{R}^3\)
   (f) \((1, 1, 2), (0, 0, 0), (2, 1, 0), (-1, 0, 3)\) span \(\mathbb{R}^3\)
   (g) \((2, 0, 3), (-1, 2, 4), (-5, 2, -2)\) span \(\mathbb{R}^3\)
   (h) \((1, 3, 0), (2, -1, 1), (1, 1, 4)\) span \(\mathbb{R}^3\)
   (i) \((-1, 2, 4), (-5, 2, -2), (2, 0, 3), (1, 2, 3)\) span \(\mathbb{R}^3\)
   (j) \((0, 0, 0), (2, 1, 4), (-1, 3, 5)\) span \(\mathbb{R}^3\)
   (k) \((2, 1, 3), (1, -1, 2)\) span \(\mathbb{R}^3\)
   (l) \((2, 1, -1), (1, 3, 1), (5, 5, -1), (0, 5, 3)\) span \(\mathbb{R}^3\)
   (m) \((-1, 4, 1, 0), (2, 2, 2), (1, 2, 3)\) span \(\mathbb{R}^3\)
   (n) \((-3, 1, 0), (1, 1, 1), (-1, 7, 5)\) span \(\mathbb{R}^3\)
   (o) \((1, 2), (2, 1)\) span \(\mathbb{R}^2\)
   (p) \((1, 2), (2, 1), (4, 5)\) span \(\mathbb{R}^2\)
   (q) \((1, 2), (2, 1), (2, 3), (2, -4)\) span \(\mathbb{R}^2\)

2. (a) Sketch any two vectors that span the space of all vectors in the plane of the paper.
   (b) Sketch any three such vectors.
   (c) Sketch any four such vectors.

3. Are the following vector sets subspaces of \(\mathbb{R}^2\)? (See accompanying figure.) Explain.
   (g) the straight line \(L\) that extends from the origin to infinity
   (b) the wedge-shaped region (including its boundary lines) that extends to infinity in both directions
   (c) the upper half plane \(x_2 \geq 0\)

4. (Solution space) First, review Exercise 13a in Section 9.6.

That solution space is a subspace of \(\mathbb{R}^n\). To illustrate, consider the simple system \(x_1 + 3x_2 = 0\); that is, \(m = 1, n = 2, a_{11} = 1,\) and \(a_{12} = 3\). The solution is \(x_2 = \alpha\) (arbitrary), \(x_1 = -3\alpha, \) or \(x = (x_1, x_2) = \alpha(-3, 1)\) so the solution space is the span of the vector \((-3, 1)\), that is, span \(\{(-3, 1)\}\). In this manner, determine the solution space for each of the following examples.

(a) \(x_1 - x_2 + 4x_3 = 0\) in \(\mathbb{R}^3\)
   (b) \(x_1 + x_2 + x_3 - x_4 = 0\) in \(\mathbb{R}^4\)
   (c) \(x_1 - x_2 + x_3 = 0\)
      \(x_1 + x_2 + x_3 = 0\) in \(\mathbb{R}^3\)
   (d) \(x_1 + 3x_2 - x_3 + x_4 = 0\)
      \(x_1 + 2x_3 + x_4 = 0\) in \(\mathbb{R}^4\)
   (e) \(x_1 - x_2 + x_3 - 2x_4 = 0\)
      \(x_1 - x_2 + x_3 + 2x_5 = 0\) in \(\mathbb{R}^5\)
   (f) \(x_1 + x_2 - x_3 + x_4 = 0\)
      \(x_1 + 2x_2 - x_4 = 0\) in \(\mathbb{R}^4\)
      \(x_1 + 2x_2 - x_3 + x_5 = 0\)
      \(2x_4 + x_5 = 0\) in \(\mathbb{R}^5\)

5. Find any two vectors in \(\mathbb{R}^3\) that span the plane
   (a) \(x_1 - 2x_2 + 4x_3 = 0\)
   (b) \(2x_1 + x_2 - 6x_3 = 0\)
   (c) \(x_1 + 5x_3 = 0\)
   (d) \(x_1 + 4x_2 + x_3 = 0\)
   (e) \(x_2 + 2x_3 = 0\)
   (f) \(3x_1 - x_2 - x_3 = 0\)

6. Show whether the given sets are identical. Explain.
   (a) span \(\{(2, -1, -1), (3, 1, 0)\}\) and span \(\{(2, -1, -1), (5, 5, 2)\}\)
   (b) span \(\{(1, 2, 3), (2, -1, 1)\}\) and span \(\{(1, 2, 3), (3, 1, 5)\}\)
   (c) span \(\{(4, 1, 0), (1, 1, 1)\}\) and span \(\{(1, 1, 1), (2, -1, -2)\}\)
   (d) span \(\{(1, 2, -1), (-3, 0, 0)\}\) and span \(\{(1, 0, 0), (1, 3, 0)\}\)
   (e) span \(\{(1, 0, 1, 2), (-1, -1, 1, 0)\}\) and span \(\{(0, 1, 2, 2), (1, 1, 3, 4)\}\)
   (f) span \(\{(1, 0, 1, 2), (1, 1, 1, 1), (1, 2, 3, 4)\}\) and span \(\{(2, 0, -1, 0), (0, -1, 2, 3), (4, 3, 2, 1)\}\)
   (g) span \(\{(1, 0, 1, 1), (2, 1, 0, 0), (1, 2, 2, 1)\}\) and span \(\{(2, -1, 0, 0), (1, -2, 0, 1), (3, 5, 4, 1)\}\)
   (h) span \(\{(1, 2, 3, 0), (0, 1, 0, 2), (2, 3, 0, 1)\}\) and span \(\{(1, 0, -3, -1), (-1, 1, 3, 0), (1, 2, 1, 1)\}\)

7. Find any two ON (orthonormal) vectors in
   (a) span \(\{(1, 2), (6, -1)\}\)
   (b) span \(\{(1, 2, 4), (2, -1, 3)\}\)
   (c) span \(\{(1, -1, 0), (1, 2, 3)\}\)
(d) \text{span}\{(2,1,0),(0,1,2)\}
(e) \text{span}\{(1,1,0,1),(0,2,-1,1)\}
(f) \text{span}\{(-2,3,1,1),(0,2,-1,1)\}


9.8 Linear Dependence

The definition of the linear dependence or independence of a set of vectors is essentially identical to Definition 3.2.1 for a set of functions, with the word "functions" changed to "vectors:"

**DEFINITION 9.8.1 Linear Dependence and Linear Independence**

A set of vectors \{\mathbf{u}_1, \ldots, \mathbf{u}_k\} is said to be **linearly dependent** if at least one of them can be expressed as a linear combination of the others. If none can be so expressed, then the set is **linearly independent**.

Thus, we urge you to review Section 3.2 in conjunction with your study of this section. As in Chapter 3, we frequently use the abbreviations LD and LI to stand for linearly dependent and linearly independent, respectively.

**EXAMPLE 1.** Let \mathbf{u}_1 = (1,0), \mathbf{u}_2 = (1,1), and \mathbf{u}_3 = (5,4). These are LD since, by inspection, we can express \mathbf{u}_3 as a linear combination of \mathbf{u}_1 and \mathbf{u}_2: \mathbf{u}_3 = \mathbf{u}_1 + 4\mathbf{u}_2. (Alternatively, we could express \mathbf{u}_2 as \frac{1}{4}\mathbf{u}_3 - \frac{1}{3}\mathbf{u}_1, or \mathbf{u}_1 = -4\mathbf{u}_2 + \mathbf{u}_3.)

**EXAMPLE 2.** Let \mathbf{u}_1 = (1,0) and \mathbf{u}_2 = (1,1). These are LI since \mathbf{u}_1 cannot be expressed as a "linear combination of the others," namely, as a scalar multiple of \mathbf{u}_2, nor can \mathbf{u}_2 be expressed as a scalar multiple of \mathbf{u}_1.

**EXAMPLE 3.** Let \mathbf{u}_1 = (2,-1), \mathbf{u}_2 = (0,0), and \mathbf{u}_3 = (0,1). These are LD since we can express \mathbf{u}_2 = 0\mathbf{u}_1 + 0\mathbf{u}_3. (The fact that we cannot express \mathbf{u}_1 as a linear combination of \mathbf{u}_2 and \mathbf{u}_3, nor \mathbf{u}_3 as a linear combination of \mathbf{u}_1 and \mathbf{u}_2 does not alter our conclusion, for recall the words "at least one" in the definition.)

It is implicit in Definition 9.8.1 that \mathbf{u}_1, \ldots, \mathbf{u}_k are all members of the same vector space; in Examples 1 to 3 that space was \mathbb{R}^2. Thus, it would make no sense to ask whether \mathbf{u}_1 = (2,5) and \mathbf{u}_2 = (4,3,0,1) are linearly dependent or not since \mathbf{u}_1 is a member of \mathbb{R}^2 while \mathbf{u}_2 is a member of \mathbb{R}^4.

The preceding examples are simple enough to be worked by inspection. In more complicated cases, the following theorem provides a systematic approach for
determining whether a given vector set is linearly dependent or linearly independent.

**THEOREM 9.8.1 Test for Linear Dependence / Independence**

A finite set of vectors \( \{ u_1, \ldots, u_k \} \) is LD if and only if there exist scalars \( \alpha_j \), not all zero, such that

\[
\alpha_1 u_1 + \cdots + \alpha_k u_k = 0; \quad (1)
\]

if (1) holds only if all the \( \alpha_j \)'s are zero, then the set is LI.

Proof is essentially the same as for Theorem 3.2.1.

**EXAMPLE 4.** Consider the 4-tuples

\[
u_1 = (2, 0, 1, -3), \quad u_2 = (0, 1, 1, 1), \quad u_3 = (2, 2, 3, 0). \quad (2)
\]

To see if these vectors are LI or LD, appeal directly to (1):

\[
\alpha_1 (2, 0, 1, -3) + \alpha_2 (0, 1, 1, 1) + \alpha_3 (2, 2, 3, 0) = (0, 0, 0, 0), \quad (3)
\]

or \( 2\alpha_1 + 2\alpha_3, \alpha_2 + 2\alpha_3, \alpha_1 + \alpha_2 + 3\alpha_3, -3\alpha_1 + \alpha_2 \) = \( 0, 0, 0, 0 \). Thus,

\[
\begin{align*}
2\alpha_1 + 2\alpha_3 &= 0, \\
\alpha_2 + 2\alpha_3 &= 0, \\
\alpha_1 + \alpha_2 + 3\alpha_3 &= 0, \\
-3\alpha_1 + \alpha_2 &= 0.
\end{align*}
\]

Applying Gauss elimination yields

\[
\begin{align*}
2\alpha_1 + 2\alpha_3 &= 0, \\
\alpha_2 + 2\alpha_3 &= 0, \\
\alpha_3 &= 0, \\
0 &= 0.
\end{align*}
\]

This system admits only the trivial solution, \( \alpha_1 = \alpha_2 = \alpha_3 = 0 \) so \( u_1, u_2, u_3 \) are LI.

**EXAMPLE 5.** Consider the 3-tuples

\[
u_1 = (1, 0, 1), \quad u_2 = (1, 1, 1), \quad u_3 = (1, 1, 2), \quad u_4 = (1, 2, 1). \quad (6)
\]

Working from (1), as in Example 4, we have

\[
\begin{align*}
\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 &= 0, \\
\alpha_2 + \alpha_3 + 2\alpha_4 &= 0, \\
\alpha_1 + \alpha_2 + 2\alpha_3 + \alpha_4 &= 0.
\end{align*}
\]

\[
\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 0, \\
\alpha_2 + \alpha_3 + 2\alpha_4 = 0, \\
\alpha_1 + \alpha_2 + 2\alpha_3 + \alpha_4 = 0.
\]

\[
\begin{align*}
\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 &= 0, \\
\alpha_2 + \alpha_3 + 2\alpha_4 &= 0, \\
\alpha_1 + \alpha_2 + 2\alpha_3 + \alpha_4 &= 0.
\end{align*}
\]
or, after Gauss elimination,
\[
\begin{align*}
\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 &= 0, \\
\alpha_2 + \alpha_3 + 2\alpha_4 &= 0, \\
\alpha_3 &= 0.
\end{align*}
\] (8)

This time, there exist nontrivial solutions for the \(\alpha_j\)'s so the vectors \(u_1, u_2, u_3\) are LD. [Specifically, (8) gives \(\alpha_3 = 0, \alpha_4 = 0, \alpha_2 = -2\alpha, \alpha_1 = \alpha\) where \(\alpha\) is arbitrary. With \(\alpha = 1\), say, (1) becomes \(u_1 - 2u_2 + 0u_3 + u_4 = 0\).]

We conclude this section with four modest theorems, the first three being essentially the same as Theorems 3.2.4–3.2.6 for functions.

**THEOREM 9.8.2** Linear Dependence / Independence of Two Vectors
A set of two vectors \(\{u_1, u_2\}\) is LD if and only if one is expressible as a scalar multiple of the other.

**THEOREM 9.8.3** Linear Dependence of Sets Containing the Zero Vector
A set containing the zero vector is LD.

**THEOREM 9.8.4** Equating Coefficients
Let \(\{u_1, \ldots, u_k\}\) be LI. Then, for
\[
a_1 u_1 + \cdots + a_k u_k = b_1 u_1 + \cdots + b_k u_k
\]
to hold, it is necessary and sufficient that \(a_j = b_j\) for each \(j = 1, \ldots, k\). That is, the coefficients of corresponding vectors on the left- and right-hand sides must match.

**THEOREM 9.8.5** Orthogonal Sets
Every finite orthogonal set of (nonzero) vectors is LI.

*Proof of Theorem 9.8.5:* Dot \(u_1\) into both sides of
\[
\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_k u_k = 0.
\] (9)

In other words,
\[
\begin{align*}
0 &= u_1 \cdot (\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_k u_k), \\
0 &= \alpha_1 u_1 \cdot u_1 + \alpha_2 u_1 \cdot u_2 + \cdots + \alpha_k u_1 \cdot u_k, \\
0 &= \alpha_1 \|u_1\|^2 + \cdots + 0 = 0.
\end{align*}
\] (10)
Now \( u_1 \neq 0 \) implies that \( \|u_1\| \neq 0 \) so it follows from (10) that \( \alpha_1 = 0 \). Similarly, dotting \( u_2 \) into (9) gives \( \alpha_2 = 0 \), and so on. Since \( \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0 \), the \( u_j \)'s must be LI, as claimed. \( \blacksquare \)

**EXAMPLE 6.** The set \{\\((2, 1), (1, 5)\\)\} in \( \mathbb{R}^2 \) is LI because neither vector can be expressed as a scalar multiple of the other. \( \blacksquare \)

**EXAMPLE 7.** Let 
\[
\begin{align*}
u_1 &= (4, -1, 1, 2), \\
u_2 &= (3, 0, 2, 5), \\
u_3 &= (0, 0, 0, 0)
\end{align*}
\] in \( \mathbb{R}^4 \). The set is LD, according to Theorem 9.8.3 because it contains the zero vector \( u_3 = 0 \). That is, \( u_3 \) can be expressed as a linear combination of \( u_1 \) and \( u_2 \): 
\[
u_3 = -1/u_1 + 0/u_2.
\]
If the preceding sentence is not clear, rewrite the equation as 
\[
u_3 = 0/u_1 + 0/u_2 - 1,
\] and observe that the \( \alpha_j \) coefficients (0, 0, and -1) are not all zero. \( \blacksquare \)

**Closure.** The foregoing discussion of the linear dependence / independence of vectors is essentially the same as the discussion of the linear dependence / independence of functions in Section 3.2, except that the Wronskian determinant test did not carry over.

**EXERCISES 9.8**

1. (a) Can a set be neither LD nor LI? Explain. 
   (b) Can a set be both LD and LI? Explain. 

2. Show that the following sets are LD by expressing one of
   the vectors as a linear combination of the others. 
   (a) \{\\((1, 1), (1, 2), (3, 4)\\)\} 
   (b) \{\\((1, 4), (2, 8), (3, -1)\\)\} 
   (c) \{\\((1, -1), (4, 2), (-3, 3)\\)\} 
   (d) \{\\((1, 2, 3), (3, 2, 1), (5, 5, 5)\\)\} 
   (e) \{\\((1, 0, 0), (0, 1, 0), (3, 3, 0), (2, -7, 9)\\)\}

3. Determine whether the following set is LI or LD. If it is LD, then give a linear relation among the vectors. 
   (a) \{(1, 3), (2, 0), (-1, 3), (7, 3)\\} 
   (b) \{(1, 3), (2, 0), (1, 2), (-1, 5)\\} 
   (c) \{(2, 3, 0), (1, -2, 3)\\} 
   (d) \{(2, 3, 0), (1, -2, 4), (1, 1, 0), (1, 1, 1)\\} 
   (e) \{(0, 0, 2), (0, 0, 3), (2, -1, 5), (1, 2, 4), (7, 9, 1), (2, 0, -4)\\} 
   (f) \{(2, 3, 0, 0), (1, -5, 0, 2), (3, 1, 2, 2)\\} 
   (g) \{(1, 3, 2, 0), (1, 1, -2, -2), (0, 2, 0, 3), (4, 7, 1, 2)\\} 
   (h) \{(2, 0, 1, -1, 0), (1, 2, 0, 3, 1), (4, -4, 3, -9, -2)\\} 
   (i) \{(1, 3, 0), (0, 1, -1), (0, 0, 0)\\} 
   (j) \{(1, 1, 0, 0), (1, -1, 0, 0), (0, 0, -2, 2), (0, 0, 1, 1)\\} 
   (k) \{(1, -3, 0, 2, 1), (-2, 6, 0, -4, -2)\\} 
   (l) \{(5, 4, 1, 1), (0, 0, 0, 0), (1, 9, -7, 2)\\} 
   (m) \{(1, 2, 3, 4), (2, 3, 4, 5)\\} 
   (n) \{(2, 1, -1), (1, 4, 2), (3, -2, -4)\\} 
   (o) \{(7, 1, 0), (-1, 1, 4), (2, 3, 5)\\} 
   (p) \{(1, 2, -1), (1, 0, 1), (3, -2, 5)\\} 
   (q) \{(3, 1, 0, 0), (1, -2, 4, 1), (2, 1, 6, 5)\\} 
   (r) \{(2, 4, 0, 1), (1, 0, 1, 2), (1, -3, 1, 2), (1, 1, -1, 1)\\}

4. Show, by graphical means, that the vector sets shown below, and lying in the plane of the paper, are LD. (The emphasis here is on the method and ideas, not on graphical precision.)
5. If \( u_1 \) and \( u_2 \) are LI, \( u_1 \) and \( u_3 \) are LI and \( u_2 \) and \( u_3 \) are LI, does it follow that \( \{u_1, u_2, u_3\} \) is LI? Prove or disprove.

6. Prove or disprove:
   (a) \( v \) is in span \( \{u_1, \ldots, u_k\} \) if \( \{v, u_1, \ldots, u_k\} \) is LD.
   (b) \( v \) is not in span \( \{u_1, \ldots, u_k\} \) if \( \{v, u_1, \ldots, u_k\} \) is LI.
   (c) \( v \) is not in span \( \{u_1, \ldots, u_k\} \) if and only if \( \{v, u_1, \ldots, u_k\} \) is LI.

7. (a) Prove Theorem 9.8.2.
    (b) Prove Theorem 9.8.3.
    (c) Prove Theorem 9.8.4.

9.9 Bases, Expansions, Dimension

9.9.1. Bases and expansions. In the calculus we learn that a given function \( f(x) \) can be "expanded" as a linear combination of powers of \( x \) (namely \( 1, x, x^2, \ldots \)),

\[
f(x) = a_0 + a_1 x + a_2 x^2 + \cdots. \tag{1}
\]

We call \( a_0, a_1, a_2, \ldots \) the "expansion coefficients," and these can be computed from \( f(x) \) as \( a_j = f^{(j)}(0)/j! \). Such representation of a given function is important, and examples such as \( e^x = 1 + x + \frac{1}{2!} x^2 + \frac{1}{3!} x^3 + \cdots \) and \( \sin x = x - \frac{1}{3!} x^3 + \frac{1}{5!} x^5 - \cdots \) are familiar to us.

Likewise useful, in Chapters 9–12, are the expansion of a given vector \( u \) in terms of a set of "base vectors" \( e_1, \ldots, e_k \):

\[
u = \alpha_1 e_1 + \cdots + \alpha_k e_k. \tag{2}
\]

How do we come up with such sets of base vectors and, once we know the \( e_j \)'s and the given \( u \), how do we compute the expansion coefficients \( \alpha_j \)? The story is simpler than for the power series of functions because whereas (1) is an infinite series and one needs to deal with the sophisticated issue of convergence, our vector expansions in Chapters 9–12 entail only a finite number of terms.

Beginning simply, consider the vector space \( \mathbb{R}^2 \), the set of all vectors in the plane of the paper. In particular, consider the vectors \( e_1 \) and \( e_2 \) shown in Fig. 1a. It should be evident (Theorem 9.8.2) that \( e_1 \) and \( e_2 \) are LI and that they span the space so that any given vector, such as \( u \) in Fig. 1b and \( v \) in Fig. 1c, can be expressed as a linear combination of them.
For the vector \( \mathbf{u} \), for example, \( \mathbf{u} = \mathbf{OA} + \mathbf{OB} \); with the aid of a scale, \( \mathbf{OA} = 1.6\mathbf{e}_1 \) and \( \mathbf{OB} = 2\mathbf{e}_2 \), so that

\[
\mathbf{u} = 1.6\mathbf{e}_1 + 2\mathbf{e}_2.
\]

Similarly (Fig. 1c),

\[
\mathbf{v} = 2\mathbf{e}_1 - 2.5\mathbf{e}_2,
\]

and so on, for any given vector in the plane. Of course, the zero vector is simply \( \mathbf{0} = 0\mathbf{e}_1 + 0\mathbf{e}_2 \).

The formulas (3) and (4) are examples of the expansion of a given vector \( \mathbf{u} \) in (3), \( \mathbf{v} \) in (4) in terms of a set of base vectors [the set \( \{\mathbf{e}_1, \mathbf{e}_2\} \)].

**Definition 9.9.1 Basis**

A finite set of vectors \( \{\mathbf{e}_1, \ldots, \mathbf{e}_k\} \) in a vector space \( S \) is a basis for \( S \) if each vector \( \mathbf{u} \) in \( S \) can be expressed (i.e., "expanded") uniquely in the form

\[
\mathbf{u} = \alpha_1\mathbf{e}_1 + \cdots + \alpha_k\mathbf{e}_k = \sum_{j=1}^{k} \alpha_j\mathbf{e}_j.
\]

By the expansion (5) being unique, we mean that the \( \alpha_j \) expansion coefficients are uniquely determined.

**Theorem 9.9.1 Test for Basis**

A finite set \( \{\mathbf{e}_1, \ldots, \mathbf{e}_k\} \) in a vector space \( S \) is a basis for \( S \) if and only if it spans \( S \) and is LI.

**Proof:** First, it follows from the definition of the verb span that every vector \( \mathbf{u} \) in \( S \) can be expanded as in (5) if and only if the set \( \{\mathbf{e}_1, \ldots, \mathbf{e}_k\} \) spans \( S \). Turning to the question of the uniqueness of the expansion, suppose that both expansions

\[
\mathbf{u} = \alpha_1\mathbf{e}_1 + \cdots + \alpha_k\mathbf{e}_k, \quad (6)
\]

\[
\mathbf{u} = \beta_1\mathbf{e}_1 + \cdots + \beta_k\mathbf{e}_k \quad (7)
\]

hold for any given vector \( \mathbf{u} \) in \( S \). Subtracting (7) from (6) gives

\[
(\alpha_1 - \beta_1)\mathbf{e}_1 + \cdots + (\alpha_k - \beta_k)\mathbf{e}_k = \mathbf{0}. \quad (8)
\]

Now, each of the coefficients \( (\alpha_1 - \beta_1), \ldots, (\alpha_k - \beta_k) \) in (8) must be zero, in which case \( \alpha_1 = \beta_1, \ldots, \alpha_k = \beta_k \) and expansions (6) and (7) are identical if and only if
the set \( \{e_1, \ldots, e_k\} \) is LI. Hence, the expansion (5) is unique if and only if the set is LI, and this completes the proof. \( \blacksquare \)

The key idea revealed in the foregoing proof is that a basis needs to contain enough vectors but not too many: enough so that the set spans the space and can therefore be used to expand any given vector in the space, but not too many, in order that such expansions will be unique.

**EXAMPLE 1.** Consider the vectors

\[
e_1 = (-2, 1), \quad e_2 = (2, 4).
\] (9)

As may be verified, the set (9) is LI and spans \( \mathbb{R}^2 \) and is therefore a basis for \( \mathbb{R}^2 \).

Using that set to expand the vector \( u = (6, 2) \), say, we express

\[
u = \alpha_1 e_1 + \alpha_2 e_2,
\] (10)

or \((6, 2) = (-2\alpha_1, \alpha_1) + (2\alpha_2, 4\alpha_2)\). Hence,

\[
-2\alpha_1 + 2\alpha_2 = 6,
\]

\[
\alpha_1 + 4\alpha_2 = 2.
\] (11)

Solving (11), \( \alpha_1 = -2 \) and \( \alpha_2 = 1 \) so the expansion (10) is

\[
u = -2e_1 + e_2,
\] (12)

as displayed in Fig. 2a.

It is to be emphasized that the basis (9) shown in Fig. 2a is by no means the only basis for \( \mathbb{R}^2 \); there are slews of them. For example, it is readily verified that another is

\[
e'_1 = (4, -1), \quad e'_2 = (-1, 5),
\] (13)

and in this case the expansion of \( u = (6, 2) \) is found to be

\[
u = \frac{32}{19} e'_1 + \frac{14}{19} e'_2,
\] (14)

as depicted in Fig. 2b.

**COMMENT.** The difference between the expansions (12) and (13) is not at odds with the notion of uniqueness since the two expansions are with respect to different bases. In other words, (12) is the unique expansion of \( u \) in terms of the \( e_1, e_2 \) basis, and (14) is the unique expansion of \( u \) in terms of the \( e'_1, e'_2 \) basis. \( \blacksquare \)

**9.9.2. Dimension.** If we always worked in 2-space or 3-space, the concept of dimension would hardly need elaboration; for example, 3-space is three-dimensional, a plane within it is two-dimensional, and a line within it is one-dimensional. However, having generalized our vector concept beyond 3-space, we need to clarify the idea of dimension.

**DEFINITION 9.9.2 Dimension**

If the greatest number of LI vectors that can be found in a vector space \( S \) is \( k \),
where \( 1 \leq k < \infty \), then \( S \) is \textit{k-dimensional}, and we write
\[
\dim S = k.
\]
If \( S \) is the zero vector space (i.e., if it contains only the zero vector), we define \( \dim S = 0 \). If an arbitrarily large number of LI vectors can be found in \( S \), we say that \( S \) is \textit{infinite-dimensional}.*

To determine the dimension of a given vector space, it may be more convenient to use the following theorem than to work directly from Definition 9.9.2.

\begin{center}
\underline{THEOREM 9.9.2 Test for Dimension}
\end{center}

If a vector space \( S \) admits a basis consisting of \( k \) vectors, then \( S \) is \textit{k-dimensional}.

\textit{Proof}: Let \( \{e_1, \ldots, e_k\} \) be a basis for \( S \). Because these vectors form a basis, they must be LI. Hence, we have \textit{at least} \( k \) LI vectors in \( S \), and it remains to show that \textit{no more than} \( k \) LI vectors can be found in \( S \). Suppose that vectors \( e'_1, \ldots, e'_{k+1} \) in \( S \) are LI. Each of these can be expanded in terms of the given base vectors, as
\[
\begin{align*}
e'_1 &= a_{11}e_1 + \cdots + a_{1k}e_k, \\
e'_{k+1} &= a_{k+1,1}e_1 + \cdots + a_{k+1,k}e_k,
\end{align*}
\]
(15)
say. Putting these expressions into the equation
\[
\alpha_1 e'_1 + \alpha_2 e'_2 + \cdots + \alpha_{k+1} e'_{k+1} = 0
\]
(16)
and grouping terms gives
\[
(\alpha_1 a_{11} + \cdots + \alpha_{k+1} a_{k+1,1}) e_1 + \cdots + (\alpha_1 a_{1k} + \cdots + \alpha_{k+1} a_{k+1,k}) e_k = 0.
\]
But the set \( \{e_1, \ldots, e_k\} \) is LI since it is a basis, so each coefficient in the preceding equation must be zero:
\[
\begin{align*}
a_{11}\alpha_1 + \cdots + a_{k+1,1}\alpha_{k+1} &= 0, \\
\vdots
\end{align*}
\]
(17)
These are \( k \) linear homogeneous equations in the \( k + 1 \) unknowns \( \alpha_1 \) through \( \alpha_{k+1} \), and such a system necessarily admits nontrivial solutions (Theorem 8.3.4). Thus, the \( \alpha \)'s in (17) are not all necessarily zero so the vectors \( e'_1, \ldots, e'_{k+1} \) could

*Infinite-dimensional function spaces will be studied in Chapter 17.
Chapter 9. Vector Space

not have been LI after all. Hence, it is not possible to find more than $k$ LI vectors in $S$, and this completes the proof. ■

The spaces of chief concern in Chapters 9–12 are the $n$-tuple spaces $\mathbb{R}^n$ and subspaces thereof. For $\mathbb{R}^n$ we can say the following.

**THEOREM 9.9.3 Dimension of $\mathbb{R}^n$**
The dimension of $\mathbb{R}^n$ is $n$: $\dim \mathbb{R}^n = n$.

**Proof:** The vectors

$$
e_1 = (1, 0, \ldots, 0),
\quad e_2 = (0, 1, 0, \ldots, 0),
\quad \vdots
\quad e_n = (0, \ldots, 0, 1)
$$

constitute a basis for $\mathbb{R}^n$ because any vector $u = (u_1, \ldots, u_n)$ in $\mathbb{R}^n$ can be expanded uniquely as $u = u_1e_1 + \cdots + u_ne_n$. Since this basis contains $n$ vectors, it follows from Theorem 9.9.2 that $\mathbb{R}^n$ is $n$-dimensional. ■

Indeed, we might well have questioned the reasonableness of our definition of dimension if $\mathbb{R}^n$ had turned out to be other than $n$-dimensional! The ON basis (18) is called the **standard basis** for $\mathbb{R}^n$ (and is the $n$-space generalization of the "i,j,k" ON basis that might be known to you from other courses).

Finally, what about the dimension of a **subspace**, for example, the subspace of $\mathbb{R}^3$ that is spanned by two given vectors?

**THEOREM 9.9.4 Dimension of Span $\{u_1, \ldots, u_k\}$**
The dimension of span $\{u_1, \ldots, u_k\}$, where the $u_i$'s are not all zero, denoted as $\dim \text{span} \{u_1, \ldots, u_k\}$, is equal to the greatest number of LI vectors within the generating set $\{u_1, \ldots, u_k\}$.

**Proof:** Denote the generating set $\{u_1, \ldots, u_k\}$ as $U$. Let the greatest number of LI vectors in $U$ be $N$, where $1 \leq N \leq k$. It may be assumed, without loss of generality, that the members of $U$ have been numbered so that $u_1, \ldots, u_N$ are LI. Then each of the remaining members of $U$, namely $u_{N+1}, \ldots, u_k$, can be expressed as a linear combination of $u_1, \ldots, u_N$. Surely, then, each vector in span $U$ can similarly be expressed as a linear combination of $u_1, \ldots, u_N$. Now $\{u_1, \ldots, u_N\}$ is LI and spans span $U$. According to Theorem 9.9.3, then, the dimension of span $U$ is $N$; that is, it is the same as the greatest number of LI vectors in $U$, as was to be proved. ■
EXAMPLE 2. Let

\[ u_1 = (3, -1, 2, 1), \quad u_2 = (1, 1, 0, -1), \quad u_3 = (4, 0, 2, 0). \]

These vectors are, of course, members of \( \mathbb{R}^4 \). But since \( u_1, u_2, u_3 \) are only three vectors, \( \dim \{ \text{span} \{ u_1, u_2, u_3 \} \} \) is at most three. In fact, it is not three since we see that \( u_3 = u_1 + u_2 \). But \( u_1 \) and \( u_2 \), say, are LI since neither is a scalar multiple of the other. Thus, there are only two LI vectors within the generating set so \( \dim \{ u_1, u_2 \} = 2 \).

In Example 2 we determined that the greatest number of LI vectors in the generating set was 2 by inspection. What if we wish to determine \( \dim \{ u_1, \ldots, u_k \} \) where the \( u_j \)'s are members of \( \mathbb{R}^8 \), and \( k = 6 \), say? For such a large problem we cannot expect “inspection” to work. Yet, what are we to do, test the \( u_j \)'s for linear independence one at a time, two at a time, three at a time, and so on, until we determine the greatest number of LI vectors in \( \{ u_1, \ldots, u_k \} \)? That would be quite tedious. No, we will see later, in Chapter 10, that the best way to determine the greatest number of LI vectors in a given set is to determine the “rank” of a certain matrix, and that can be done by the extremely efficient method of elementary row operations. Meanwhile, in the present section, we “get by” by keeping the examples and exercises simple enough so that we can rely on inspection.

Let us return, now, to our discussion of bases and expansions.

9.9.3. Orthogonal bases. If, as in Example 1, there are many bases for a given space, then how do we decide which one to select? We will find that in most applications the most convenient basis to use is dictated by the context, so let us not worry about that now. This point is addressed in Chapter 11 as well as in the chapters on PDEs.

However, we do wish to show here, that orthogonal bases are to be preferred whenever possible. For observe from Example 1 that to expand \( u \) (that is, to compute the \( \alpha_j \) expansion coefficients) we needed to solve the system (11) of two equations in two unknowns. Similarly, if we seek to expand a given vector in \( \mathbb{R}^8 \), then there will be eight base vectors (because \( \mathbb{R}^8 \) is eight-dimensional) and eight \( \alpha_j \) expansion coefficients, and these will be found by solving a system [analogous to (11)] of eight equations in the eight unknown \( \alpha_j \)'s. Thus, the expansion process can be quite laborious.

On the other hand, suppose that \( \{ e_1, \ldots, e_k \} \) is an orthogonal basis for \( S \); that is, it is not only a basis but also happens to be an orthogonal set:

\[ e_i \cdot e_j = 0 \quad \text{if} \quad i \neq j. \]  \hspace{1cm} (19)

Suppose that we wish to expand a given vector \( u \) in \( S \) in terms of that basis; that is, we wish to determine the coefficients \( \alpha_1, \ldots, \alpha_k \) in the expansion

\[ u = \alpha_1 e_1 + \alpha_2 e_2 + \cdots + \alpha_k e_k. \] \hspace{1cm} (20)

To accomplish this, dot (20) with \( e_1, e_2, \ldots, e_k \), in turn. Doing so, and using
Chapter 9. Vector Space

(19), we obtain the linear system

\[
\begin{align*}
\mathbf{u} \cdot \mathbf{e}_1 &= (\mathbf{e}_1 \cdot \mathbf{e}_1) \alpha_1 + 0 \alpha_2 + \cdots + 0 \alpha_k, \\
\mathbf{u} \cdot \mathbf{e}_2 &= 0 \alpha_1 + (\mathbf{e}_2 \cdot \mathbf{e}_2) \alpha_2 + 0 \alpha_3 + \cdots + 0 \alpha_k, \\
\vdots \\
\mathbf{u} \cdot \mathbf{e}_k &= 0 \alpha_1 + \cdots + 0 \alpha_{k-1} + (\mathbf{e}_k \cdot \mathbf{e}_k) \alpha_k,
\end{align*}
\]

where all of the quantities \(\mathbf{u} \cdot \mathbf{e}_1, \ldots, \mathbf{u} \cdot \mathbf{e}_k, \mathbf{e}_1 \cdot \mathbf{e}_1, \ldots, \mathbf{e}_k \cdot \mathbf{e}_k\) are computable since \(\mathbf{u}, \mathbf{e}_1, \ldots, \mathbf{e}_k\) are known. The crucial point is that even though (21) is still \(k\) equations in the \(k\) unknown \(\alpha_j\)'s, the system is *uncoupled* (i.e., the only unknown in the first equation is \(\alpha_1\), the only one in the second is \(\alpha_2\), and so on) and readily gives

\[
\alpha_1 = \frac{\mathbf{u} \cdot \mathbf{e}_1}{\mathbf{e}_1 \cdot \mathbf{e}_1}, \quad \alpha_2 = \frac{\mathbf{u} \cdot \mathbf{e}_2}{\mathbf{e}_2 \cdot \mathbf{e}_2}, \quad \ldots, \quad \alpha_k = \frac{\mathbf{u} \cdot \mathbf{e}_k}{\mathbf{e}_k \cdot \mathbf{e}_k},
\]

provided, of course, that none of the denominators vanish. But these quantities cannot vanish because \(\mathbf{e}_j \cdot \mathbf{e}_j = \|\mathbf{e}_j\|^2\), which is zero if and only if \(\mathbf{e}_j = \mathbf{0}\), and this cannot be because if any \(\mathbf{e}_j\) were \(\mathbf{0}\), then the set \(\{\mathbf{e}_1, \ldots, \mathbf{e}_k\}\) would be LD (Theorem 9.8.3), and hence not a basis.

Thus, if the \(\{\mathbf{e}_1, \ldots, \mathbf{e}_k\}\) basis is orthogonal, the expansion of any given \(\mathbf{u}\) is simply

\[
\mathbf{u} = \left(\frac{\mathbf{u} \cdot \mathbf{e}_1}{\mathbf{e}_1 \cdot \mathbf{e}_1}\right) \mathbf{e}_1 + \cdots + \left(\frac{\mathbf{u} \cdot \mathbf{e}_k}{\mathbf{e}_k \cdot \mathbf{e}_k}\right) \mathbf{e}_k = \sum_{j=1}^{k} \left(\frac{\mathbf{u} \cdot \mathbf{e}_j}{\mathbf{e}_j \cdot \mathbf{e}_j}\right) \mathbf{e}_j,
\]

If, besides being orthogonal, the \(\mathbf{e}_j\)'s are normalized (\(\|\mathbf{e}_j\| = 1\)) so that they constitute an ON (orthonormal) basis, then (23) simplifies slightly to

\[
\mathbf{u} = (\mathbf{u} \cdot \hat{\mathbf{e}}_1) \hat{\mathbf{e}}_1 + \cdots + (\mathbf{u} \cdot \hat{\mathbf{e}}_k) \hat{\mathbf{e}}_k = \sum_{j=1}^{k} (\mathbf{u} \cdot \hat{\mathbf{e}}_j) \hat{\mathbf{e}}_j,
\]

where we recall that carets denote unit vectors.

**EXAMPLE 3.** Expand \(\mathbf{u} = (4, 3, -3, 6)\) in terms of the orthogonal base vectors \(\mathbf{e}_1 = (1, 0, 2, 0), \mathbf{e}_2 = (0, 1, 0, 0), \mathbf{e}_3 = (-2, 0, 1, 5), \mathbf{e}_4 = (-2, 0, 1, -1)\) of \(\mathbb{R}^4\). This basis is orthogonal but not ON so we use (23) rather than (24). Computing \(\mathbf{u} \cdot \mathbf{e}_1 = -2, \mathbf{e}_1 \cdot \mathbf{e}_1 = 5\), and so on, (23) gives

\[
\mathbf{u} = -\frac{2}{5} \mathbf{e}_1 + 3 \mathbf{e}_2 + \frac{19}{30} \mathbf{e}_3 - \frac{17}{6} \mathbf{e}_4.
\]
Alternatively, we could have inferred, from \( \mathbf{u} = \alpha_1 \mathbf{e}_1 + \cdots + \alpha_4 \mathbf{e}_4 \), the four equations

\[
\begin{align*}
\alpha_1 - 2\alpha_3 - 2\alpha_4 &= 4, \\
\alpha_2 &= 3, \\
2\alpha_1 + \alpha_3 + \alpha_4 &= -3, \\
5\alpha_3 - \alpha_4 &= 6
\end{align*}
\]

on the four unknown \( \alpha_j \)'s, and solved these by Gauss elimination, but it is much easier to "cash in" on the orthogonality of the basis and to use (23). If we choose to work with an ON basis, we can scale the \( \mathbf{e}_j \)'s as \( \hat{\mathbf{e}}_1 = \frac{1}{\sqrt{5}}(1, 0, 2, 0) \), \( \hat{\mathbf{e}}_2 = (0, 1, 0, 0) \), \( \hat{\mathbf{e}}_3 = \frac{1}{\sqrt{30}}(-2, 0, 1, 5) \), \( \hat{\mathbf{e}}_4 = \frac{1}{\sqrt{6}}(-2, 0, 1, -1) \). Then (24) gives

\[
\mathbf{u} = -\frac{2}{\sqrt{5}} \hat{\mathbf{e}}_1 + 3\hat{\mathbf{e}}_2 + \frac{19}{\sqrt{30}} \hat{\mathbf{e}}_3 - \frac{17}{\sqrt{6}} \hat{\mathbf{e}}_4, \tag{27}
\]

which result is equivalent to (25). \( \blacksquare \)

Given a nonorthogonal basis there are three possibilities. First, one can use it and face up to the tedious expansion process. Second, one can "trade the nonorthogonal basis in" for an orthogonal basis using the Gram-Schmidt orthogonalization procedure, which procedure is introduced briefly in the exercises and discussed in detail in the next section. Third, one can retain the nonorthogonal basis but streamline the expansion process by computing and utilizing a set of dual, or reciprocal, vectors corresponding to the given basis, as described in the exercises.

**Closure.** This section is about the expansion of vectors, in a given vector space \( \mathcal{S} \), in terms of a set of base vectors. A set of vectors \( \{\mathbf{e}_1, \ldots, \mathbf{e}_k\} \) in \( \mathcal{S} \) is a basis for \( \mathcal{S} \) if each vector \( \mathbf{u} \) in \( \mathcal{S} \) can be expanded as a unique linear combination of the \( \mathbf{e}_j \)'s. We showed (Theorem 9.9.1) that \( \{\mathbf{e}_1, \ldots, \mathbf{e}_k\} \) is indeed a basis for \( \mathcal{S} \) if and only if it spans \( \mathcal{S} \) (so each \( \mathbf{u} \) can be expanded) and is LI (so the expansion is unique). The number of vectors in any basis for \( \mathcal{S} \) is called the dimension of \( \mathcal{S} \). For instance, \( \mathbb{R}^n \) admits the standard basis (18), comprised of \( n \) vectors, so \( \mathbb{R}^n \) is \( n \)-dimensional. And the greatest number of LI vectors in a set \( \{\mathbf{u}_1, \ldots, \mathbf{u}_k\} \) is the dimension of their span.

We found that the expansion process (i.e., the determination of the expansion coefficients) can be quite laborious if there are many base vectors, but is extremely simple if the basis is orthogonal, or ON, in which case the expansions are given by (23) or (24), respectively. You should remember those two formulas and be able to derive them as well.
EXERCISES 9.9

1. Show whether the following is a basis.
   (a) \((1, 0), (1, 1), (1, 2)\) for \(\mathbb{R}^2\)
   (b) \((3, 2), (-1, -5)\) for \(\mathbb{R}^2\)
   (c) \((1, 1)\) for \(\mathbb{R}^2\)
   (d) \((2, 0, 1), (5, -1, 2), (1, -1, 0)\) for \(\mathbb{R}^3\)
   (e) \((5, -1, 2), (2, 0, 1), (1, -1, 1)\) for \(\mathbb{R}^3\)
   (f) \((2, 1, 0, 6), (7, -1, -2, 3), (4, 3, 2, 1)\) for \(\mathbb{R}^4\)
   (g) \((4, 3, -2, 1), (5, 0, 0, 0), (2, 1, -3, 0), (1, 2, 4, 5)\) for \(\mathbb{R}^4\)
   (h) \((4, 2, 0, 0), (1, 2, 3, 0), (5, -2, 3, 1), (0, -6, 0, 1)\) for \(\mathbb{R}^4\)
   (i) \((1, 0, 0, 0), (1, 1, 0, 0), (1, 1, 1, 0), (1, 1, 1, 1)\) for \(\mathbb{R}^4\)
   (j) \((3, 0, 0, 1), (2, 0, 0, 1), (1, 3, 5, 6), (4, -2, 1, 3)\) for \(\mathbb{R}^4\)
   (k) \((1, 3, -1, 2), (1, 2, 4, 3), (2, 5, 3, 5), (3, 7, 7, 8)\) for \(\mathbb{R}^4\)
   (l) \((2, 3, 5, 0), (1, -1, 2, 3), (4, 1, 2, 3), (5, 4, 1, 0), (1, 2, 4, 6)\) for \(\mathbb{R}^4\)
   (m) \((1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 0)\) for \(\mathbb{R}^4\)
   (n) \((1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), (0, 0, 0, 0)\) for \(\mathbb{R}^4\)
   (o) \((1, 1, 2), (4, -2, -1)\) for span \(\{(2, -4, -5), (1, -1, -1)\}\)
   (p) \((1, 1, 2), (4, -2, -1)\) for span \(\{(3, -5, -6), (1, 2, 1)\}\)
   (q) \((1, 1, 1), (1, -1, 2)\) for span \(\{(2, 4, 1), (1, 7, -2)\}\)
   (r) \((1, 2, 3), (1, 0, 4)\) for span \(\{(3, 2, 0), (1, 1, -1)\}\)

2. Expand each vector \(u\) in terms of the orthogonal basis \(\{e_1, e_2, e_3\}\) of \(\mathbb{R}^3\), where \(e_1 = (2, 1, 3), e_2 = (1, -2, 0), e_3 = (3, 3, -5)\).
   (a) \(u = (9, -2, 4)\)
   (b) \(u = (1, 0, 0)\)
   (c) \(u = (0, 1, 5)\)
   (d) \(u = (3, -1, 1)\)
   (e) \(u = (0, 5, 0)\)
   (f) \(u = (1, 2, 3)\)

3. (a)–(f) Expand each of the \(u\) vectors in Exercise 2 in terms of the ON basis \(\{e_1, e_2, e_3\}\) of \(\mathbb{R}^3\), where \(e_1, e_2, e_3\) are normalized versions of \(e_1, e_2, e_3\) given in Exercise 2.

4. Expand each vector \(u\) in terms of the orthogonal basis \(\{e_1, \ldots, e_4\}\) of \(\mathbb{R}^4\), where \(e_1 = (2, 0, -1, -5), e_2 = (2, 0, -1, 1), e_3 = (0, 1, 0, 0), e_4 = (1, 0, 2, 0)\).
   (a) \(u = (1, 0, 0, 0)\)
   (b) \(u = (0, 6, 0, 0)\)
   (c) \(u = (2, 5, 1, -3)\)
   (d) \(u = (4, 3, -2, 0)\)
   (e) \(u = (1, 2, 0, 5)\)
   (f) \(u = (2, -7, 4, 1)\)
   (g) \(u = (0, 0, 0, 9)\)
   (h) \(u = (2, 3, -2, 1)\)
   (i) \(u = (0, 0, 5, 0)\)
   (j) \(u = (1, 1, 1, 1)\)

5. Verify that the \(\{e_1, \ldots, e_4\}\) vectors given in Example 3 are a basis for \(\mathbb{R}^4\). Also, solve (26) by Gauss elimination and verify that the \(a_\ell\)'s thus obtained agree with those given in (25).

6. If \(\{e_1, \ldots, e_k\}\) is an orthogonal set in a vector space \(S\), is it a basis?

(a) for \(S\)?
(b) for span \(\{e_1, \ldots, e_k\}\)?

7. (Zero vector space) Show that a zero vector space (i.e., a vector space consisting of the zero vector alone) has no basis.

8. Let \(u_1 = (1, 0, 0), u_2 = (0, 1, 0), u_3 = (0, 0, 1), u_4 = (1, 1, 0), u_5 = (0, 1, 1), u_6 = (1, 1, 1),\) and \(u_7 = 2\).
   Evaluate each of the following.
   (a) \(\dim \text{span} \{u_1\}\)
   (b) \(\dim \text{span} \{u_1, u_2\}\)
   (c) \(\dim \text{span} \{u_1, u_2, u_3\}\)
   (d) \(\dim \text{span} \{u_1, u_2, u_3, u_4\}\)
   (e) \(\dim \text{span} \{u_1, u_2, u_4\}\)
   (f) \(\dim \text{span} \{u_1, u_2, u_5\}\)
   (g) \(\dim \text{span} \{u_5, u_6, u_7\}\)
   (h) \(\dim \text{span} \{u_1, u_3, u_6, u_7\}\)

9. Let \(u_1 = (1, 0, 0, 0), u_2 = (1, 1, 0, 0), u_3 = (1, 1, 1, 0), u_4 = (1, 1, 1, 1), u_5 = (0, 0, 0, 1), u_6 = (3, 3, 3, 3)\). Evaluate each of the following.
   (a) \(\dim \text{span} \{u_1, u_3, u_5\}\)
   (b) \(\dim \text{span} \{u_1, u_4, u_6\}\)
   (c) \(\dim \text{span} \{u_2, u_4, u_6\}\)
   (d) \(\dim \text{span} \{u_5\}\)
   (e) \(\dim \text{span} \{u_3, u_4\}\)
   (f) \(\dim \text{span} \{u_3, u_4, u_5, u_6\}\)
   (g) \(\dim \text{span} \{u_3, u_6\}\)
   (h) \(\dim \text{span} \{u_2, u_3, u_4, u_5, u_6\}\)

10. (a)–(f) Determine the dimension of the solution space in Exercise 4 of Section 9.7.

11. (Gram–Schmidt orthogonalization process) Given \(k\) LI vectors \(v_1, \ldots, v_k\), it is possible to obtain from them \(k\) ON vectors, say \(e_1, \ldots, e_k\), in span \(\{v_1, \ldots, v_k\}\) by the Gram–Schmidt process, after Jürgen P. Gram (1850–1916) and Erhardt Schmidt (1876–1959), by taking \(e_1\) equal to \(v_1\), taking \(e_2\) equal to a suitable linear combination of \(v_1, v_2\), taking \(e_3\) equal to a suitable linear combination of \(v_1, v_2, v_3\), and so on, and then normalizing the results. The resulting ON set is as
follows:
\[
\hat{e}_1 = \frac{v_1}{\|v_1\|}, \\
\hat{e}_2 = \frac{v_2 - (v_2 \cdot \hat{e}_1)\hat{e}_1}{\|v_2 - (v_2 \cdot \hat{e}_1)\hat{e}_1\|}, \\
\vdots \\
\hat{e}_j = \frac{v_j - \sum_{i=1}^{j-1} (v_j \cdot \hat{e}_i)\hat{e}_i}{\|v_j - \sum_{i=1}^{j-1} (v_j \cdot \hat{e}_i)\hat{e}_i\|} \quad \text{through \(j = k\)}.
\]

We now state the problem: Verify that each \(\hat{e}_j\) defined by (11.1) is a linear combination of \(v_1, \ldots, v_j\), and that the \(\hat{e}_j\)'s are ON. [In verifying that \(\|\hat{e}_j\| = 1\), be sure to show that each denominator in (11.1) is nonzero.]

12. In each case use the Gram-Schmidt formula (11.1) in Exercise 11 to obtain an ON set from the given LI set.

(a) \((4, 0), (2, 1)\)
(b) \((1, -2), (3, 4)\)
(c) \((1, 0, 0), (1, 1, 0), (1, 1, 1)\)
(d) \((1, 1, 0), (2, -1, 1), (1, 0, 3)\)
(e) \((1, 1, 1), (2, 0, -1)\)
(f) \((1, 1, 1), (1, 0, 1), (1, 1, 0)\)
(g) \((1, 2, 1), (1, -1, 2), (-1, 3, 1)\)
(h) \((2, 0, 1), (1, 1, 1), (-2, 0, 3)\)
(i) \((2, 1, 1, 0), (1, 5, -1, 2)\)
(j) \((6, -1, 1, 2, 1), (2, 3, -1, 1, 4)\)

13. (The dual or reciprocal vectors) For definiteness, consider our vector space \(S\) to be \(\mathbb{R}^n\).

(a) If \(\{\hat{e}_1, \ldots, \hat{e}_n\}\) is an ON basis for \(\mathbb{R}^n\), and \(u\) is in \(\mathbb{R}^n\), then by dotting \(\hat{e}_i\) into both sides of the equation \(u = \sum_{j=1}^{n} \alpha_j e_j\), we find that \(\alpha_i = u \cdot \hat{e}_i\), so that the expansion of \(u\) in terms of the given basis is
\[
u = \sum_{j=1}^{n} (u \cdot \hat{e}_j)\hat{e}_j.
\]

If, instead, we have a basis \(\{e_1, \ldots, e_n\}\) which is not ON, then, as noted in the text, the expansion process is not so simple. However, suppose that we can find a set \(\{e_1^*, \ldots, e_n^*\}\) such that
\[
e_i \cdot e_j^* = \begin{cases} 1, & i = j \\
0, & i \neq j.
\end{cases}
\]

Then show that dotting \(e_i^*\) into \(u = \sum_{j=1}^{n} \alpha_j e_j\) gives \(\alpha_i = u \cdot e_i^*\) so that
\[
u = \sum_{j=1}^{n} (u \cdot e_j^*)e_j.
\]

The set \(\{e_1^*, \ldots, e_n^*\}\) is called the dual, or reciprocal, set corresponding to the original set \(\{e_1, \ldots, e_n\}\). (We will see in the last exercise in Section 10.6 that the dual set exists, is unique, and is itself a basis for \(\mathbb{R}^n\), the so-called dual or reciprocal basis.)

(b) Given the basis \(e_1 = (1, 0), e_2 = (1, 1)\) for \(\mathbb{R}^2\), use equation (13.2) to determine the dual vectors \(e_1^*, e_2^*\). Then use equation (13.3) to expand \(u = (3, 1)\). Sketch \(e_1, e_2, e_1^*, e_2^*\) to scale, and verify the expansion graphically, that is, by means of the parallelogram rule of vector addition.

(c) Repeat part (b), for \(e_1 = (2, 1), e_2 = (0, 2), u = (-3, 2)\).

(d) Repeat part (b), for \(e_1 = (1, -1), e_2 = (2, 1), u = (0, 4)\).

(e) Given the basis \(e_1 = (1, 0, 0), e_2 = (1, 1, 0), e_3 = (1, 1, 1)\) for \(\mathbb{R}^3\), use equation (13.2) to determine the dual vectors \(e_1^*, e_2^*, e_3^*\). Then use equation (13.3) to expand each of the vectors \(u = (4, -1, 5), v = (0, 0, 2), w = (5, -2, 3)\). Be sure to see that the dual vectors get computed once and for all, for a given basis \(\{e_1, \ldots, e_n\}\); once we have got them, expansions of the form (13.3) are simple.

(f) Repeat part (e) for \(e_1 = (2, 0, 1), e_2 = (1, 1, 0), e_3 = (1, -1, 3),\) and \(u = (6, 1, 0), v = (1, 2, 4), w = (0, 3, 0)\).

(g) Show that if the \(\{e_1, \ldots, e_n\}\) basis does happen to be ON, then the dual vectors coalesce with the \(e_j\)'s, i.e., \(e_j^* = e_j\) for \(j = 1, 2, \ldots, n\).
9.10 Best Approximation

Let $\mathcal{S}$ be a normed inner product vector space (i.e., a vector space with both a norm and an inner, or dot, product defined), and let the norm be the “natural norm” $\|u\| = \sqrt{u \cdot u}$. We know that if $\{e_1, \ldots, e_N\}$ is a basis for $\mathcal{S}$, then any vector $u$ in $\mathcal{S}$ can be (uniquely) expanded in the form $u = \sum_{j=1}^{N} c_j e_j$. If the basis is orthogonal, then the expansion process is easy, with the $c_j$'s computed, from the given vector $u$ and the base vectors $e_j$, as $c_j = (u \cdot e_j)/(e_j \cdot e_j)$. And if the basis is not only orthogonal but ON, then $u = \sum_{j=1}^{N} c_j \hat{e}_j$, where $c_j = u \cdot \hat{e}_j$.

However, what if we do not have a “full deck?” That is, what if $\{\hat{e}_1, \ldots, \hat{e}_N\}$ is ON, but falls short of being a basis for $\mathcal{S}$ (i.e., $N < \dim \mathcal{S}$)? If $u$ happens to fall within span $\{\hat{e}_1, \ldots, \hat{e}_N\}$, which subspace of $\mathcal{S}$ we denote as $\mathcal{T}$, then it can still be expanded in terms of $\hat{e}_1, \ldots, \hat{e}_N$, but if it is not in $\mathcal{T}$, then it cannot be so expanded.

In the latter case the question arises, what is the best approximation of $u$ in terms of $\hat{e}_1, \ldots, \hat{e}_N$? In this section we answer that question in general, and illustrate the results for the case where $\mathcal{S}$ is $\mathbb{R}^n$. Later in this book, when we study Fourier series and partial differential equations, our interest will be in function spaces instead.

9.10.1. Best approximation and orthogonal projection. The best approximation problem, which we address is this: given a vector $u$ in $\mathcal{S}$, and an ON set $\{\hat{e}_1, \ldots, \hat{e}_N\}$ in $\mathcal{S}$, what is the best approximation

$$u \approx c_1 \hat{e}_1 + \cdots + c_N \hat{e}_N = \sum_{j=1}^{N} c_j \hat{e}_j? \tag{1}$$

That is, how do we compute the $c_j$ coefficients so as to render the error vector $E = u - \sum_{j=1}^{N} c_j \hat{e}_j$ as small as possible? In other words, how do we choose the $c_j$'s so as to minimize the norm of the error vector $\|E\|$? If $\|E\|$ is a minimum, then so is $\|E\|^2$, so let us minimize $\|E\|^2$ (to avoid square roots), where

$$\|E\|^2 = E \cdot E = \left( u - \sum_{j=1}^{N} c_j \hat{e}_j \right) \cdot \left( u - \sum_{j=1}^{N} c_j \hat{e}_j \right)$$

$$= u \cdot u - 2 \sum_{j=1}^{N} c_j (u \cdot \hat{e}_j) + \sum_{j=1}^{N} c_j^2, \tag{2}$$

and where the step

$$\left( \sum_{j=1}^{N} c_j \hat{e}_j \right) \cdot \left( \sum_{j=1}^{N} c_j \hat{e}_j \right) = (c_1 \hat{e}_1 + \cdots + c_N \hat{e}_N) \cdot (c_1 \hat{e}_1 + \cdots + c_N \hat{e}_N)$$

$$= c_1^2 + \cdots + c_N^2 = \sum_{j=1}^{N} c_j^2 \tag{3}$$
follows from the orthonormality of the \( \hat{e}_j \)'s.

Defining \( \mathbf{u} \cdot \hat{e}_j \equiv \alpha_j \) and noting that \( \mathbf{u} \cdot \mathbf{u} = \| \mathbf{u} \|^2 \), we may express (2) as

\[
\| \mathbf{E} \|^2 = \sum_{j=1}^{N} c_j^2 - 2 \sum_{j=1}^{N} \alpha_j c_j + \| \mathbf{u} \|^2 ,
\]

or, completing the square, as

\[
\| \mathbf{E} \|^2 = \sum_{j=1}^{N} (c_j - \alpha_j)^2 + \| \mathbf{u} \|^2 - \sum_{j=1}^{N} \alpha_j^2 .
\]  

Observe that \( \mathbf{u} \) and the ON set \( \{ \hat{e}_1, \ldots, \hat{e}_N \} \) are given so that \( \| \mathbf{u} \| \) and the \( \alpha_j \)'s in (4) are fixed computable quantities: \( \| \mathbf{u} \| = \sqrt{\mathbf{u} \cdot \mathbf{u}} \) and \( \alpha_j = \mathbf{u} \cdot \hat{e}_j \) for \( j = 1, 2, \ldots, N \). Thus, in seeking to minimize the right-hand side of (4), the only control we exercise is in our choice of the \( \alpha_j \)'s. The right-hand side of (4) is greater than or equal to zero,* and so is the \( \sum_{j=1}^{N} (c_j - \alpha_j)^2 \) term containing the \( \alpha_j \)'s. Thus, the best that we can do is to set \( c_j = \alpha_j \) (\( j = 1, 2, \ldots, N \)). With that choice, our best approximation (1) becomes

\[
\mathbf{u} \approx \sum_{j=1}^{N} (\mathbf{u} \cdot \hat{e}_j) \hat{e}_j.
\]  

Let us summarize these results.

**THEOREM 9.10.1 Best Approximation**

Let \( \mathbf{u} \) be any vector in a normed inner product vector space \( S \) with natural norm \( (\| \mathbf{u} \| = \sqrt{\mathbf{u} \cdot \mathbf{u}}) \), and let \( \{ \hat{e}_1, \ldots, \hat{e}_N \} \) be an ON set in \( S \). Then the best approximation (1) is obtained when the \( \alpha_j \)'s are given by \( \alpha_j = \mathbf{u} \cdot \hat{e}_j \), as indicated in (5).

**EXAMPLE 1.** Let \( S = \mathbb{R}^2 \), \( N = 1 \), \( \hat{e}_1 = \frac{1}{13}(12, 5) \), and \( \mathbf{u} = (1, 1) \), as shown in Fig. 1. Find the best approximation \( \mathbf{u} \approx c_1 \hat{e}_1 \), that is, the best approximation of \( \mathbf{u} \) in span \( \{ \hat{e}_1 \} \) (which is the line \( L \)). Theorem 9.10.1 gives \( c_1 = \mathbf{u} \cdot \hat{e}_1 = 17/13 \), and hence the best approximation

\[
\mathbf{u} \approx \frac{17}{13} \hat{e}_1 .
\]  

which is the vector \( OA \) in Fig. 1.

COMMENT. Observe from the figure that the best approximation \( OA \) is the **orthogonal projection** of \( \mathbf{u} \) onto span \( \{ \hat{e}_1 \} \), which orthogonality is verified by the calculation

\[
\| \mathbf{E} \|^2 = \| \mathbf{u} \|^2 - \sum_{j=1}^{N} \alpha_j^2 .
\]  

*This fact may not be obvious due to the minus sign in front of the last summation. But remember that the right-hand side of (4) is equal to \( \| \mathbf{E} \|^2 \), and surely \( \| \mathbf{E} \|^2 \geq 0 \).
Chapter 9. Vector Space

Figure 2. Best approximation of \( \mathbf{u} \) in \( \text{span} \{ \mathbf{e}_1, \mathbf{e}_2 \} \).

**Figure 2.** Best approximation of \( \mathbf{u} \) in \( \text{span} \{ \mathbf{e}_1, \mathbf{e}_2 \} \).

\[
\mathbf{A} \mathbf{B} \cdot \mathbf{e}_1 = (\mathbf{u} - \mathbf{OA}) \cdot \mathbf{e}_1 = (\mathbf{u} - \frac{17}{15} \mathbf{e}_1) \cdot \mathbf{e}_1 = \frac{17}{15} \frac{17}{15} = 0.
\]
That result makes perfect sense since if \( c_1 \mathbf{e}_1 \) is to be the best approximation to \( \mathbf{u} \), then the distance from the tip of \( \mathbf{u} \) to the tip of \( c_1 \mathbf{e}_1 \) (which is some point on \( L \)) should be as small as possible. That shortest distance is the perpendicular distance from the tip of \( \mathbf{u} \) to the line \( L \).

**Example 2.** Let \( S \) be \( \mathbb{R}^3 \), let \( N = 2 \) with \( \mathbf{e}_1 = (1, 0, 0) \) and \( \mathbf{e}_2 = (0, 1, 0) \), and let \( \mathbf{u} = (a, b, c) \), as shown in Fig. 2. Computing the coefficients in (5) as \( \mathbf{u} \cdot \mathbf{e}_1 = a \) and \( \mathbf{u} \cdot \mathbf{e}_2 = b \), (5) becomes

\[
\mathbf{u} \approx a \mathbf{e}_1 + b \mathbf{e}_2.
\]

The latter is an equality if \( c = 0 \). That is, (7) is an equality if \( \mathbf{u} \) happens to lie in \( \text{span} \{ \mathbf{e}_1, \mathbf{e}_2 \} \), but if \( c \neq 0 \) then the best approximation \( a \mathbf{e}_1 + b \mathbf{e}_2 \) to \( \mathbf{u} \) is the orthogonal projection of \( \mathbf{u} \) onto \( \text{span} \{ \mathbf{e}_1, \mathbf{e}_2 \} \).

In Examples 1 and 2, \( S \) was \( \mathbb{R}^2 \) and \( \mathbb{R}^3 \), respectively, so we were able to draw useful pictures. In each case we discovered that the best approximation of \( \mathbf{u} \) on the subspace \( T \) of \( S \) spanned by \( \mathbf{e}_1, \ldots, \mathbf{e}_N \) was the orthogonal projection of \( \mathbf{u} \) onto \( T \). Is that result true in all cases? That is, is the error vector \( \mathbf{E} \) necessarily orthogonal to \( T \)? Since the error vector is

\[
\mathbf{E} = \mathbf{u} - \sum_{j=1}^{N} (\mathbf{u} \cdot \mathbf{e}_j) \mathbf{e}_j,
\]

we have

\[
\mathbf{E} \cdot \mathbf{e}_k = \left[ \mathbf{u} - \sum_{j=1}^{N} (\mathbf{u} \cdot \mathbf{e}_j) \mathbf{e}_j \right] \cdot \mathbf{e}_k
= \mathbf{u} \cdot \mathbf{e}_k - (\mathbf{u} \cdot \mathbf{e}_k)(1) = 0
\]

for each \( k = 1, 2, \ldots, N \), where the second equality follows from the fact that \( \mathbf{e}_j \cdot \mathbf{e}_k = 0 \) if \( j \neq k \) and 1 if \( j = k \).

Since \( \mathbf{E} \) is orthogonal to every one of the \( \mathbf{e}_j \)'s, it is therefore orthogonal to every vector in \( T \). In that sense we say that the right-hand side of (5) is the orthogonal projection of \( \mathbf{u} \) onto \( T \), and denote it as \( \text{proj}_T \mathbf{u} \):

\[
\text{proj}_T \mathbf{u} = \sum_{j=1}^{N} (\mathbf{u} \cdot \mathbf{e}_j) \mathbf{e}_j.
\]

The idea that the best approximation of \( \mathbf{u} \) in \( T \) is the orthogonal projection of \( \mathbf{u} \) onto \( T \) lends a welcome geometrical interpretation to the problem of best approximation. In fact, let us rephrase Theorem 9.10.1 in terms of orthogonal projection.

**Theorem 9.10.1'** Best Approximation by Orthogonal Projection

Let \( \mathbf{u} \) be any vector in a normed inner product vector space \( S \) with natural norm
9.10. Best Approximation

\( \|u\| = \sqrt{u \cdot u} \), and let \( \{\hat{e}_1, \ldots, \hat{e}_N\} \) be an ON set in \( S \). Denote the subspace span \( \{\hat{e}_1, \ldots, \hat{e}_N\} \) of \( S \) as \( T \). Then the best approximation of \( u \) in \( T \) (i.e., of the form \( c_1\hat{e}_1 + \cdots + c_N\hat{e}_N \)) is given by the orthogonal projection of \( u \) onto \( T \), namely, by \( \text{proj}_T u \).

9.10.2. Kronecker delta. When working with ON sets it is convenient to use the Kronecker delta symbol \( \delta_{jk} \), defined as

\[
\delta_{jk} = \begin{cases} 
1, & j = k \\
0, & j \neq k
\end{cases}
\]

and named after Leopold Kronecker (1823–1891), who contributed to algebra and the theory of equations. The subscripted \( j \) and \( k \) are usually positive integers. Clearly, \( \delta_{jk} \) is symmetric in its indices \( j \) and \( k \):

\[
\delta_{jk} = \delta_{kj}.
\]

To illustrate the use of the Kronecker delta, suppose that \( \{\hat{e}_1, \ldots, \hat{e}_N\} \) is an ON basis for some space \( S \), and that we wish to expand a given \( u \) in \( S \) as

\[
u = \sum_{j=1}^{N} c_j \hat{e}_j.
\]

To determine the \( c_j \)'s, dot \( \hat{e}_k \) into both sides, where \( k \) is any integer such that \( 1 \leq k \leq N \), and use the fact that \( \hat{e}_j \cdot \hat{e}_k = \delta_{jk} \) (because the \( \hat{e}_j \)'s are ON):

\[
u \cdot \hat{e}_k = \left( \sum_{j=1}^{N} c_j \hat{e}_j \right) \cdot \hat{e}_k = \sum_{j=1}^{N} c_j (\hat{e}_j \cdot \hat{e}_k) = \sum_{j=1}^{N} c_j \delta_{jk} = c_k.
\]

Thus, \( c_k = u \cdot \hat{e}_k \) for each \( k = 1, 2, \ldots, N \) so (13) becomes

\[
u = \sum_{j=1}^{N} (u \cdot \hat{e}_j) \hat{e}_j.
\]

Closure. Principal interest, in this brief section, is in the best approximation of a given vector \( u \) in a normed inner product vector space \( S \) in terms of an ON set \( \{\hat{e}_1, \ldots, \hat{e}_N\} \) which falls short of being a basis for \( S \) inasmuch as \( N < \dim S \). Of course, if \( N = \dim S \) so the set is a basis, then we have the equality (15), but if \( N < \dim S \), then the best approximation of \( u \) is given by (5), best in the vector sense; that is, the norm of the error vector [i.e., the norm of the difference between
Chapter 9. Vector Space

EXERCISES 9.10

1. We concluded from (4) that the best choice for the \( c_j \)'s is

\[
c_j = \alpha_j = u \cdot \hat{e}_j.
\]

Show that this same result is obtained from (2) by setting \( \frac{\partial \|E\|^2}{\partial c_j} = 0 \), and verify that the extremum thus obtained is a minimum.

2. Let \( S \) be \( \mathbb{R}^5 \), and let \( N = 3 \) with \( \hat{e}_1 = \frac{1}{\sqrt{3}}(1, 0, 2, 0, 0) \),

\[
\hat{e}_2 = \frac{1}{\sqrt{3}}(2, 0, -1, 0, 1), \quad \hat{e}_3 = (0, 0, 0, 1, 0).
\]

Find the best approximation to the given \( u \) vector within span \( \{\hat{e}_1, \hat{e}_2, \hat{e}_3\} \), and the norm of the error vector.

(a) \( (3, -2, 0, 0, 5) \) \hspace{1cm} (b) \( (0, 0, 0, 2, 1) \) \hspace{1cm} (c) \( (3, 0, 1, 4, 1) \)

(d) \( (1, 1, 0, 1, 1) \) \hspace{1cm} (e) \( (0, 2, 0, 0, 0) \) \hspace{1cm} (f) \( (1, 0, -3, 3, 1) \)

(g) \( (0, 7, 0, 3, 0) \) \hspace{1cm} (h) \( (1, 2, 3, 4, 5) \) \hspace{1cm} (i) \( (5, 4, 3, 2, 1) \)

3. Let \( S \) be \( \mathbb{R}^4 \), and let

\[
\hat{e}_1 = \frac{1}{\sqrt{3}}(1, 1, 0, -1), \quad \hat{e}_2 = \frac{1}{\sqrt{3}}(1, -1, -1, 0),
\]

\[
\hat{e}_3 = \frac{1}{\sqrt{3}}(1, 0, 1, 1), \quad \hat{e}_4 = \frac{1}{\sqrt{3}}(0, 1, -1, 1).
\]

Find the best approximation to \( u = (4, -2, 1, 6) \) within span \( \{\hat{e}_1\} \), span \( \{\hat{e}_1, \hat{e}_2\} \), span \( \{\hat{e}_1, \hat{e}_2, \hat{e}_3\} \), and span \( \{\hat{e}_1, \hat{e}_2, \hat{e}_3, \hat{e}_4\} \), and in each case compute the norm of the error vector, \( \|E\| \).

4. Same as Exercise 3, but for the given \( u \) vector.

(a) \( (4, 1, 0, -1) \) \hspace{1cm} (b) \( (3, -1, 1, 2) \) \hspace{1cm} (c) \( (0, 0, 2, 5) \)

(d) \( (1, 2, 4, 4) \) \hspace{1cm} (e) \( (0, 5, 3, -1) \) \hspace{1cm} (f) \( (2, 0, -1, -1) \)

5. (Bessel Inequality) Beginning with (4), derive the Bessel inequality

\[
\sum_{j=1}^{N} (u \cdot \hat{e}_j)^2 \leq \|u\|^2.
\] (5.1)

Notice that if \( u \) happens to be in span \( \{\hat{e}_1, \ldots, \hat{e}_k\} \), or if \( \dim S = N \), then (5.1) becomes an equality. In two and three dimensions that equality is actually the Pythagorean theorem, and in more than three dimensions it amounts to an abstract extension of that theorem.

6. (A Different Inner Product) In Examples 1 and 2 we use the "usual" inner product for \( \mathbb{R}^n \), \( u \cdot v = u_1v_1 + \cdots + u_nv_n \), but that is not the only acceptable one. In Exercise 3 of Section 9.6 we see that another acceptable inner product is

\[
u \cdot v = w_1u_1v_1 + \cdots + w_nu_nv_n,
\]

(6.1)

where the \( w_j \)'s are fixed positive constants, or "weights."

(a) Rework Example 1 using the modified inner product \( u \cdot v = 2u_1v_1 + u_2v_2 \) and its corresponding natural norm \( \|u\| = \sqrt{2u_1^2 + u_2^2} \). Show that the resulting best approximation, \( \frac{29}{13} \) \( \sqrt{12} \), which is not the same as the best approximation \( \frac{11}{16} \) \( \sqrt{12} \) given by (6). HINT: You will need to rescale \( e_1 \) so as to be a unit vector according to the new norm.

(b) Whereas the error vector \( \mathbf{AB} \) (Fig. 1) is orthogonal and perpendicular to span \( \{e_1\} \), show that in this exercise the error vector is indeed orthogonal to span \( \{\hat{e}_1\} \), as promised in the text, but not perpendicular to it. To explain this "paradox," show that for the modified inner product the orthogonality of two nonzero vectors does not imply their perpendicularity.

7. Verify the last step in (14), that \( \sum_{j=1}^{N} c_j \delta_{jk} = c_k \).

8. Verify the following, where \( (i, j, k, l) \) run from 1 to \( N \).

(a) \( \sum_{j=1}^{N} \delta_{ij} = 1 \)

(b) \( \sum_{j} \delta_{ij} \delta_{jk} = \delta_{ik} \)

(c) \( \sum_{j} \sum_{k} \delta_{ij} \delta_{jk} \delta_{kl} = \delta_{il} \)
Chapter 9 Review

We begin with the two- and three-dimensional "arrow vector" concept that is probably already familiar to you from an introductory course in physics, where the vectors denoted forces, velocities, and so on. For such vectors, vector addition $\mathbf{u} + \mathbf{v}$, scalar multiplication $(\alpha \mathbf{u})$, a zero vector $(\mathbf{0})$, a negative inverse $[-\mathbf{u} = (-1)\mathbf{u}]$, a norm $(\|\mathbf{u}\|)$, a dot product

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta,$$

and the angle $\theta = \cos^{-1}\left(\frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|}\right)$ between $\mathbf{u}$ and $\mathbf{v}$ are all defined.

From there, we generalize to abstract $n$-space, where $\mathbf{u} = (u_1, \ldots, u_n)$, by defining vector addition, and so on, in such a way that they agree with the corresponding arrow vector definitions when $n = 2$ and $n = 3$. For instance,

$$\mathbf{u} \cdot \mathbf{v} = \sum_{j=1}^{n} u_j v_j,$$

$$\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}} = \sqrt{\sum_{j=1}^{n} u_j^2},$$

and

$$\theta = \cos^{-1}\left(\frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|}\right).$$

From these definitions, we derived various properties such as

$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u},$$

(commutative)

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

(associative)

and so on, along with the following properties of the dot product and norm.

**Dot Product**

Commutative: $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$,

Nonnegative: $\mathbf{u} \cdot \mathbf{u} > 0$ for all $\mathbf{u} \neq \mathbf{0}$

$$= 0 \quad \text{for} \quad \mathbf{u} = \mathbf{0},$$

Linear: $(\alpha \mathbf{u} + \beta \mathbf{v}) \cdot \mathbf{w} = \alpha(\mathbf{u} \cdot \mathbf{w}) + \beta(\mathbf{v} \cdot \mathbf{w})$,

**Norm**

Scaling: $\|\alpha \mathbf{u}\| = |\alpha| \|\mathbf{u}\|$, 

Nonnegative: $\|\mathbf{u}\| > 0$ for all $\mathbf{u} \neq \mathbf{0}$

$$= 0 \quad \text{for} \quad \mathbf{u} = \mathbf{0},$$

Triangular Inequality: $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$. 

To complete the extension to generalized vector space, we reverse the cart and the horse by elevating these various properties to the level of axioms, or requirements. That is, we let the fundamental objects, the vectors, be whatever we choose them to be, and then define addition and scalar multiplication operations, a zero vector, a negative inverse, a dot or "inner" product (if we wish), and a norm (if we wish), so that those axioms are satisfied. Our chief interest, in introducing generalized vector space, is in function spaces, but we will not work with function spaces until Chapter 17, when we study Fourier series and the Sturm–Liouville theory.

Next, we introduce the concept of span and linear dependence, primarily so that we can develop the idea of the expansion of a given vector in a vector space $S$ in terms of a set of base vectors for $S$. We define a set of vectors $\{e_1, \ldots, e_k\}$ to be a basis for $S$ if each vector $u$ in $S$ can be expressed ("expanded") uniquely in the form $u = \alpha_1 e_1 + \cdots + \alpha_k e_k$, and prove that a set $\{e_1, \ldots, e_k\}$ is a basis for $S$ if and only if it spans $S$ and is LI (linearly independent). In particular, orthogonal bases are especially convenient because of the ease with which one can compute the expansion coefficients $\alpha_j$. The result is

$$u = \left(\frac{u \cdot e_1}{e_1 \cdot e_1}\right) e_1 + \cdots + \left(\frac{u \cdot e_k}{e_k \cdot e_k}\right) e_k$$

(9)

if the basis is orthogonal, and

$$u = (u \cdot \hat{e}_1) \hat{e}_1 + \cdots + (u \cdot \hat{e}_k) \hat{e}_k$$

(10)

if it is ON (orthonormal); (9) and (10) should be understood and remembered.

Finally, we study the question of the best approximation of a given vector $u$ in a vector space $S$ in terms of an ON set $\{\hat{e}_1, \ldots, \hat{e}_N\}$ which falls short of being a basis for $S$. We show that the best approximation (i.e., the one that minimizes the norm of the error vector) is

$$u \approx \sum_{j=1}^{N} (u \cdot \hat{e}_j) \hat{e}_j$$

(11)

which, in geometrical language, is the orthogonal projection of $u$ onto the span of $\hat{e}_1, \ldots, \hat{e}_N$. 
Chapter 10

Matrices and Linear Equations

10.1 Introduction

We have already met matrices in Section 8.3.3, but they were introduced there only as a notational convenience for the implementation of Gauss elimination and Gauss–Jordan reduction. In the present chapter we focus on matrix theory itself, which will enable us to obtain additional important results regarding the solution of systems of linear algebraic equations.

One way to view matrix theory is to think in terms of a parallel with function theory. In our mathematical training, we first study numbers — the points on a real number axis. Then we study functions, which are mappings, or transformations, from one real axis to another. For instance, \( f(x) = x^2 \) maps the point \( x = 3 \), say, on an \( x \) axis to the point \( f = 9 \) on an \( f \) axis. Just as functions act upon numbers, we shall see that matrices act upon vectors and are mappings from one vector space to another. Having studied vectors, in Chapter 9, we can now turn our attention to matrices.

Historically, matrix theory did not become a part of undergraduate engineering science curricula until around 1960, when digital computers became widely available in academia.

10.2 Matrices and Matrix Algebra

A matrix is a rectangular array of quantities that are called the elements of the matrix. Normally, the elements will be real numbers, although they may occasionally be other objects such as differential operators or even matrices. Some of these cases will be met as we go along; for the present, however, let us consider the elements to be real numbers. The complex case is studied in Chapter 12.
Chapter 10. Matrices and Linear Equations

Specifically, any matrix \( A \) may be expressed as

\[
A = \begin{bmatrix}
  a_{11} & a_{12} & \cdots & a_{1n} \\
  a_{21} & a_{22} & \cdots & a_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix},
\]

(1)

where the brackets (or, in some texts, parentheses) are used to emphasize that the entire array is to be regarded as a single entity. A horizontal line of elements is called a row, and a vertical line is called a column. Counting rows from the top and columns from the left, then

\[
a_{21} \ a_{22} \ \cdots \ a_{2n} \ \text{and} \ a_{13} \\
a_{23} \\
\vdots \\
a_{m3}
\]

say, are the second row and third column, respectively. Thus, we call the first subscript on \( a_{ij} \) the row index, and the second subscript the column index.

We usually use boldface capital letters to denote matrices and lightface lowercase letters to denote their elements. The matrix \( A \) in (1) is seen to have \( m \) rows and \( n \) columns and is therefore said to be \( m \times n \) (read "\( m \) by \( n \")); we shall refer to this as the form of \( A \). In some applications \( m \) and/or \( n \) may be infinite, but here we shall consider only matrices of finite size: \( 1 \leq m < \infty, 1 \leq n < \infty \). Furthermore, \( m \) and \( n \) may, but need not, be equal. If \( A \) is small we may wish to dispense with the subscript notation for the elements. For example, if \( m = n = 2 \), we may prefer

\[
A = \begin{bmatrix}
  a & b \\
  c & d
\end{bmatrix}
\]

(2)

but if \( A \) is large this becomes inconvenient. The double-subscript notation employed in (1) is especially convenient for digital computer calculations.

In view of the subscript notation in (1), one also writes

\[
A = \{a_{ij}\}
\]

(3)

for short, where \( a_{ij} \) is called the \( ij \) element and \( i = 1, \ldots, m \) and \( j = 1, \ldots, n \). Some authors write \( a_{i,j} \) in place of \( a_{ij} \) to avoid ambiguity — for example, to prevent us from reading \( a_{21} \) as a-sub-twenty-one, but we will omit the commas, except when such ambiguity is not easily resolved from the context.
EXAMPLE 1. The matrices

\[
A = \begin{bmatrix} 3 & -1 \\ 0 & 2 \\ 7 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 8 & -6 & 0 \\ 1 & 3 & 27 \\ 2 & 9 & 4 \end{bmatrix}, \quad \text{and} \quad C = [8, -7, 0, 4, 3]
\]

are 3 × 2, 3 × 3, and 1 × 5, respectively. If we denote \(B = \{b_{ij}\}\), then \(b_{11} = 8, b_{12} = -6, b_{22} = 9\), and so on. \(C\) happens to be a single row and it seems best to separate the elements by commas, but the commas are not essential.

Two matrices are said to be equal if they are of the same form and if their corresponding elements are equal. For instance, none of the matrices above are equal, but if \(D = [8, -7, 0, 4, 3]\) then, \(D = C\).

One may be tempted to identify \(C\), above, as a 5-tuple vector rather than as a 1 × 5 matrix. That would be a bit premature since vectors are not merely objects; they have rules for vector addition and scalar multiplication defined, whereas our matrices are, thus far, just mathematical "objects." In fact, our next step is to define some arithmetic operations for matrices so that they may be manipulated in useful ways. For vectors we defined two arithmetic operations, vector addition and scalar multiplication; for matrices we define three: matrix addition, scalar multiplication, and the multiplication of matrices.

Matrix addition. If \(A = \{a_{ij}\}\) and \(B = \{b_{ij}\}\) are any two matrices of the same form, say \(m \times n\), then their sum \(A + B\) is defined as

\[
A + B \equiv \{a_{ij} + b_{ij}\}
\]

and is itself an \(m \times n\) matrix. If \(A\) and \(B\) are of the same form, they are said to be conformable for addition; if they are not of the same form, then \(A + B\) is not defined.

EXAMPLE 2. If

\[
A = \begin{bmatrix} 2 & 0 & -6 \\ 1 & 3 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 2 & 0 \\ 15 & 6 & 3 \end{bmatrix}, \quad \text{and} \quad C = \begin{bmatrix} 4 & 2 \\ -1 & 0 \end{bmatrix}
\]

then,

\[
A + B = \begin{bmatrix} 1 & 2 & -6 \\ 16 & 9 & 7 \end{bmatrix}
\]

but \(A + C\) and \(B + C\) are not defined since \(A\) and \(B\) are 2 × 3 while \(C\) is 2 × 2.

Scalar multiplication. If \(A = \{a_{ij}\}\) is any \(m \times n\) matrix and \(c\) is any scalar, their product is defined as

\[
cA \equiv \{ca_{ij}\}
\]
and is itself an \( m \times n \) matrix; we do not distinguish between \( cA \) and \( Ac \). Furthermore, we denote
\[
-A \equiv (-1)A.
\]
In place of \( A + (-B) \), we simply write \( A - B \), and call it the difference of \( A \) and \( B \), or \( A \) minus \( B \).

**EXAMPLE 3.** If \( A \) and \( C \) are the matrices in Example 1, then
\[
3A = \begin{bmatrix} 9 & -3 \\ 0 & 6 \\ 21 & 15 \end{bmatrix} \quad \text{and} \quad -C = [-8, 7, 0, -4, -3].
\]

We shall list the important properties of matrix addition and scalar multiplication in a moment, but first let us define the so-called zero matrix \( 0 \) to be any \( m \times n \) matrix all the elements of which are zero. For example,
\[
\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = 0 \quad \text{and} \quad \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0,
\]
the first being \( 2 \times 3 \), the second being \( 3 \times 1 \).

**THEOREM 10.2.1** Properties of Matrix Addition and Scalar Multiplication
If \( A, B, \) and \( C \) are \( m \times n \) matrices, \( 0 \) is an \( m \times n \) zero matrix, and \( \alpha, \beta \) are any scalars, then
\[
\begin{align*}
A + B &= B + A, \\
(A + B) + C &= A + (B + C), \\
A + 0 &= A, \\
A + (-A) &= 0, \\
\alpha(\beta A) &= (\alpha\beta)A, \\
(\alpha + \beta)A &= \alpha A + \beta A, \\
\alpha(A + B) &= \alpha A + \beta B, \\
1A &= A, \\
0A &= 0, \\
\alpha0 &= 0.
\end{align*}
\]

The proof follows from the foregoing definitions and is left for the exercises.
Observe that there are no surprises in (9); the usual rules of arithmetic are seen to apply. For the special case where \( A \) consists of a single row (or column), we see that the definitions of addition and scalar multiplication above are identical to those introduced in Section 9.4 for \( n \)-tuple vectors. Thus, we may properly refer to the matrices
\[
A = \begin{bmatrix} a_{11}, \ldots, a_{1n} \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} a_{11} \\ \vdots \\ a_{n1} \end{bmatrix}
\]
as \( n \)-dimensional row and column vectors, respectively.

**Matrix multiplication.** Judging from the rather natural way in which matrix addition and scalar multiplication are defined, by (4) and (7), one might well expect the multiplication of two matrices \( A = \{a_{ij}\} \) and \( B = \{b_{ij}\} \) to be defined only if \( A \) and \( B \) are of the same form, with the definition \( AB = \{a_{ij}b_{ij}\} \). In fact, this is not the case. Instead, the standard definition of matrix multiplication is the one suggested by Cayley.* Called the **Cayley product**, it is as follows: if \( A = \{a_{ij}\} \) is any \( m \times n \) matrix and \( B = \{b_{ij}\} \) is any \( n \times p \) matrix (so that the number of columns of \( A \) is equal to the number of rows of \( B \)), then the product \( AB \) is defined as
\[
AB = \left\{ \sum_{k=1}^{n} a_{ik}b_{kj} \right\}; \quad (1 \leq i \leq m, \ 1 \leq j \leq p)
\]
that is, if we denote \( AB = C = \{c_{ij}\} \), then
\[
c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}.
\]
(12)

If the number of columns of \( A \) is equal to the number of rows of \( B \), then \( A \) and \( B \) are said to be **conformable for multiplication**; if not, the product \( AB \) is **not defined**.

**NOTE:** The relative forms of \( A, B, \) and their product \( C \) are important and, as stated above, are as follows:
\[
A \quad \text{times} \quad B = C.
\]
\[
m \times n \quad n \times p \quad m \times p
\]

**EXAMPLE 4.** Suppose that
\[
A = \begin{bmatrix} 2 & 0 & -5 \\ 1 & 3 & 2 \\ 4 & 1 & -1 \\ 0 & 2 & 7 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 5 & 1 \\ -2 & 3 \\ 1 & 0 \end{bmatrix}.
\]

*Arthur Cayley (1821–1895) produced around 200 papers in a 15-year period during which he was engaged in the practice of law. In 1863, he accepted a professorship of mathematics at Cambridge.
Then \( A \) is \( 4 \times 3 \) and \( B \) is \( 3 \times 2 \). Since the number of columns of \( A \) (namely, 3) is the same as the number of rows of \( B \), the product \( AB \) is defined and, according to (13), will be \( 4 \times 2 \). According to the definition (12),

\[
AB = \begin{bmatrix} 2 & 0 & -5 \\ 1 & 3 & 2 \\ 4 & 1 & -1 \\ 0 & 2 & 7 \end{bmatrix} \begin{bmatrix} 5 & 1 \\ -2 & 3 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 5 & 2 \\ 1 & 10 \\ 17 & 7 \\ 3 & 6 \end{bmatrix} = C. \tag{14}
\]

To compute \( c_{32} \), for example, (12) gives

\[
c_{32} = \sum_{k=1}^{3} a_{3k} b_{k2} = (4)(1) + (1)(3) + (-1)(0) = 7,
\]

as indicated by the arrows in (14). One more:

\[
c_{11} = \sum_{k=1}^{3} a_{1k} b_{k1} = (2)(5) + (0)(-2) + (-5)(1) = 5.
\]

We move across the rows of the first matrix and down the columns of the second.

**COMMENT.** Observe that \( c_{32} \) is the dot product of the third row of \( A \), considered as a 3-tuple vector, with the second column of \( B \). More generally, if \( AB = C = \{c_{ij}\} \), then \( c_{ij} \) is the dot product of the \( i \)th row of \( A \) with the \( j \)th column of \( B \). Thus, the number of elements in the rows of \( A \) (namely, the number of columns in \( A \)) must equal the number of elements in the columns of \( B \) (namely, the number of rows of \( B \)).

**EXAMPLE 5.** Two more examples:

\[
\begin{bmatrix} 4 & 0 & -1 \\ 5 & 2 & 3 \\ 1 & 0 & 6 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix} = \begin{bmatrix} (5)(3) + (0)(2) + (-1)(5) \\ (2)(3) + (3)(2) + (4)(5) \\ (1)(3) + (0)(2) + (6)(5) \\ (0)(3) + (0)(2) + (1)(5) \end{bmatrix} = \begin{bmatrix} 10 \\ 32 \\ 33 \\ 5 \end{bmatrix},
\]

and

\[
\begin{bmatrix} -3 & 1 \\ 10 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} (-3)(1) + (1)(2) & (-3)(0) + (1)(4) \\ (10)(1) + (2)(2) & (10)(0) + (2)(4) \end{bmatrix} = \begin{bmatrix} -1 & 4 \\ 14 & 8 \end{bmatrix}.
\]
It is extremely important to see that matrix multiplication is not, in general, commutative; that is,

\[ AB \neq BA, \]

except in exceptional cases. For suppose that \( A = m \times n \) and \( B = n \times p \) (i.e., \( A \) is \( m \times n \) and \( B \) is \( n \times p \)) so that \( AB \) is at least defined. However, \( BA = (n \times p)(m \times n) \) is not even defined, let alone equal to \( AB \), unless \( p = m \). Assuming that that is the case,

\[ BA = (n \times m)(m \times n) = n \times n, \]

whereas

\[ AB = (m \times n)(n \times m) = m \times m. \]

Comparing (16a) with (16b), we see that we must also have \( m = n \) if \( AB \) and \( BA \) are to be of the same form and hence possibly equal. Thus, a necessary condition for \( AB \) to equal \( BA \) (i.e., for \( A \) and \( B \) to commute under multiplication) is that \( A \) and \( B \) both be \( n \times n \).

**EXAMPLE 6.** If

\[
A = \begin{bmatrix} 2 & 3 \\ 3 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} -2 & 3 \\ 3 & -4 \end{bmatrix}, \quad C = \begin{bmatrix} 2 & 3 \\ 1 & 0 \end{bmatrix},
\]

we find that \( A \) and \( B \) commute \((AB = BA)\), but \( A \) and \( C \) do not \((AC \neq CA)\), nor do \( B \) and \( C \).

Thus, the condition that \( A \) and \( B \) must both be \( n \times n \) for \( A \) and \( B \) to commute, is necessary but not sufficient. In view of the importance of which factor is first and which is second in a matrix product \( AB \), we sometimes say that \( B \) is pre-multiplied by \( A \), and \( A \) is post-multiplied by \( B \) so as to leave no doubt as to which factor is first and which is second.

The lack of commutativity, in general, is a major setback so we must wonder why Cayley's definition has been adopted rather than the simpler one that comes to mind, \( AB \equiv \{a_{ij}b_{ij}\} \), which would surely yield commutativity since \( BA = \{b_{ij}a_{ij}\} = \{a_{ij}b_{ij}\} = AB \) (the second equality following from the commutativity of the multiplication of ordinary numbers). A sufficiently compelling reason to use Cayley's definition involves the application of matrix notation to systems of linear algebraic equations for it turns out that, with Cayley's definition of multiplication, the system of \( m \) linear algebraic equations

\[
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = c_1,
\]
\[
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = c_2,
\]
\[
\vdots
\]
\[
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = c_m
\]

in the \( n \) unknowns \( x_1, \ldots, x_n \) is equivalent to the single compact matrix equation

\[ Ax = c. \]
where

\[ A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \text{and} \quad c = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}. \quad (19) \]

A is called the coefficient matrix, and \( x \) is the unknown; that is, its components are the unknowns \( x_1, \ldots, x_n \). To verify the claimed equivalence, work out the product \( Ax \), and set the result equal to \( c \). That step gives

\[ \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ a_{21}x_1 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix}. \quad (20) \]

These two \( m \times 1 \) matrices (or \( m \)-dimensional column vectors) will be equal if and only if each of their corresponding \( m \) elements (or components) are equal. Thus, (18) is equivalent to the \( m \) scalar equations (17), so (17) and (18) are equivalent, as claimed. This important result is of course a consequence of (11) and provides strong support for adopting that definition of matrix multiplication.

Any \( n \times n \) matrix \( A = \{a_{ij}\} \) is said to be square, and of order \( n \), and the elements \( a_{11}, a_{22}, \ldots, a_{nn} \) are said to lie on the main diagonal of \( A \) — that is, the diagonal from the upper left corner to the lower right corner. Notice that to be able to multiply any matrix \( A \) with itself, \( A \) needs to be square. For suppose that \( A \) is \( m \times n \); then we have

\[ A \cdot A \quad m \times n \quad m \times n \]

and we need \( n \) (the number of columns in the first matrix) to equal \( m \) (the number of rows in the second) for the multiplication to be defined. If \( A \) is square and \( p \) is any positive integer, we define

\[ A \cdot A \cdots A \equiv A^p. \quad (p \text{ factors}) \quad (21) \]

The familiar laws of exponents,

\[ A^p A^q = A^{p+q}, \quad (A^p)^q = A^{pq} \quad (22) \]

follow for any positive integers \( p \) and \( q \).

*Thus, we distinguish between form and order: the form of an \( m \times n \) matrix is \( m \times n \), the order of an \( n \times n \) matrix is \( n \).
If, in particular, the only nonzero elements of a square matrix lie on the main diagonal, \( \mathbf{A} \) is said to be a diagonal matrix. For example,

\[
\mathbf{A} = \begin{bmatrix} 3 & 0 \\ 0 & -5 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 7 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 0 \end{bmatrix}
\]

are diagonal, and any diagonal matrix of order \( n \) can be denoted as

\[
\mathbf{D} = \begin{bmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & d_{nn} \end{bmatrix}, \tag{23}
\]

where not all of the \( d_{ij} \)'s are zero,* and all of the off-diagonal elements are zero; that is, \( d_{ij} = 0 \) if \( i \neq j \). It is left for the exercises to show that

\[
\mathbf{D}^p = \begin{bmatrix} d_{11}^p & 0 & \cdots & 0 \\ 0 & d_{22}^p & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & d_{nn}^p \end{bmatrix}, \tag{24}
\]

for any positive integer \( p \).

If, furthermore, \( d_{11} = d_{22} = \cdots = d_{nn} = 1 \), then \( \mathbf{D} \) is called the identity matrix \( \mathbf{I} \). Thus,

\[
\mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 \end{bmatrix} = \{ \delta_{ij} \}, \tag{25}
\]

where \( \delta_{ij} \) is the Kronecker delta symbol defined in Section 9.10.2, namely,

\[
\delta_{ij} = \begin{cases} 1 \text{ if } i = j, \\ 0 \text{ if } i \neq j. \end{cases} \tag{26}
\]

It is sometimes convenient to include a subscript \( n \) to indicate the order of \( \mathbf{I} \). For example,

\[
\mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.
\]

*If all of the diagonal elements were zero as well, then all the elements would be zero, and it would be more reasonable to describe \( \mathbf{D} \) as a zero matrix, \( \mathbf{0} \).
The key property of the identity matrix is that if \( A \) is any square matrix of the same order as \( I \), then

\[
IA = AI = A,
\]

proof of which is left for the exercises. In other words, \( I \) is the matrix analog of the number 1 in scalar arithmetic! From the first equality in (27), we see that one case in which commutativity does hold is when one of the matrices is the identity matrix \( I \).

Finally, it is convenient to extend our definition of \( A^p \) [recall (21)] to the case where \( p = 0 \). If \( A \) is any \( n \times n \) matrix, we define

\[
A^0 = I,
\]

where \( I \) is an \( n \times n \) identity matrix.

Perhaps we should take a moment to mention that whereas the identity matrix \( I \) is necessarily square, the zero matrix \( 0 \) is simply \( m \times n \), not necessarily square. It is readily verified that

\[
0A = 0 \quad \text{and} \quad A0 = 0
\]

for any matrix \( A \). In the first of these equations, suppose that the 0 on the left is \( m \times n \) and that \( A \) is \( n \times p \). Then the 0 on the right is \( m \times p \); that is, it is not necessarily of the same form as the one on the left.

In view of the general failure of commutativity, as stated in (15), we may well wonder if any other familiar arithmetic rules fail to hold for the multiplication of matrices. The answer is yes; the following rules for real numbers \( (a, b, c) \) do not carry over to matrices:

1. \( ab = ba \) (commutativity).
2. If \( ab = ac \) and \( a \neq 0 \), then \( b = c \) (cancellation rule).
3. If \( ab = 0 \), then \( a = 0 \) and/or \( b = 0 \).
4. If \( a^2 = 1 \), then \( a = +1 \) or \(-1 \).

To add emphasis, we state these difficulties as a theorem.

**THEOREM 10.2.2** "Exceptional" Properties of Matrix Multiplication

(i) \( AB \neq BA \) in general.
(ii) Even if \( A \neq 0 \), \( AB = AC \) does not imply that \( B = C \).
(iii) \( AB = 0 \) does not imply that \( A = 0 \) and/or \( B = 0 \).
(iv) \( A^2 = I \) does not imply that \( A = +I \) or \(-I \).

The first of these has already been discussed, and the others are discussed in the exercises. Theorem 10.2.2 notwithstanding, several important properties do carry over from the multiplication of real numbers to the multiplication of matrices:
THEOREM 10.2.3 "Ordinary" Properties of Matrix Multiplication
If \( \alpha, \beta \) are scalars, and the matrices \( A, B, C \) are suitably conformable, then

\[
\begin{align*}
(\alpha A)B &= A(\alpha B) = \alpha(AB), \quad \text{(associativity)} \quad (30a) \\
A(BC) &= (AB)C, \quad \text{(associativity)} \quad (30b) \\
(A + B)C &= AC + BC, \quad \text{(distributivity)} \quad (30c) \\
C(A + B) &= CA + CB, \quad \text{(distributivity)} \quad (30d) \\
A(\alpha B + \beta C) &= \alpha AB + \beta AC. \quad \text{(linearity)} \quad (30e)
\end{align*}
\]

Proof is left for the exercises.

Partitioning. Let us close this section with a discussion of the partitioning of matrices. The idea is that any matrix \( A \) (which is larger than \( 1 \times 1 \)) may be partitioned into a number of smaller matrices called blocks by vertical lines that extend from bottom to top, and horizontal lines that extend from left to right.

EXAMPLE 7. Partitioning.

\[
A = \begin{bmatrix} 2 & 0 & -3 \\ 5 & 2 & 7 \\ 1 & 3 & 0 \\ 0 & 4 & 6 \end{bmatrix} = \begin{bmatrix} 2 & 0 & -3 \\ 5 & 2 & 7 \\ 1 & 3 & 0 \\ 0 & 4 & 6 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix}, \quad (31)
\]

where the blocks are

\[
A_{11} = \begin{bmatrix} 2 & 0 \\ 5 & 2 \end{bmatrix}, \quad A_{12} = \begin{bmatrix} -3 \\ 7 \end{bmatrix}, \quad A_{21} = [1, 3], \quad A_{22} = [0],
\]

and so on. Clearly, the partition is not unique. In the present example we could also have set

\[
A = \begin{bmatrix} 2 & 0 & -3 \\ 5 & 2 & 7 \\ 1 & 3 & 0 \\ 0 & 4 & 6 \end{bmatrix} = \begin{bmatrix} 2 & 0 & -3 \\ 5 & 2 & 7 \\ 1 & 3 & 0 \\ 0 & 4 & 6 \end{bmatrix} = [A_{11}, A_{12}, A_{13}], \quad (32)
\]

say.  

While the matrices used here as illustrations are kept small for convenience, those encountered in modern applications may be quite large, for example \( 600 \times 800 \). Even with modern computers such large matrices create special computational problems, and it is often advantageous to work instead with a number of smaller matrices through the use of partitioning. Such advantages might well prove illusory, however, were it not for the fact that the usual matrix arithmetic can be carried out with partitioned matrices.
Specifically, if $A$ and $B$ are partitioned as

$$
A = \begin{bmatrix}
A_{11} & A_{12} & \cdots & A_{1n} \\
\vdots & \vdots & & \vdots \\
A_{m1} & A_{m2} & \cdots & A_{mn}
\end{bmatrix}
$$

and

$$
B = \begin{bmatrix}
B_{11} & B_{12} & \cdots & B_{1q} \\
\vdots & \vdots & & \vdots \\
B_{p1} & B_{p2} & \cdots & B_{pq}
\end{bmatrix},
$$

then

$$
\alpha A = \begin{bmatrix}
\alpha A_{11} & \alpha A_{12} & \cdots & \alpha A_{1n} \\
\vdots & \vdots & & \vdots \\
\alpha A_{m1} & \alpha A_{m2} & \cdots & \alpha A_{mn}
\end{bmatrix};
$$

if $m = p$ and $n = q$ and each $A_{ij}$ block is of the same form as the corresponding $B_{ij}$ block, then

$$
A + B = \begin{bmatrix}
A_{11} + B_{11} & A_{12} + B_{12} & \cdots & A_{1n} + B_{1n} \\
\vdots & \vdots & & \vdots \\
A_{m1} + B_{m1} & A_{m2} + B_{m2} & \cdots & A_{mn} + B_{mn}
\end{bmatrix};
$$

and if $n = p$ and we denote $AB = C$, then

$$
C_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj},
$$

provided that the number of columns in each $A_{ik}$ is the same as the number of rows in the corresponding $B_{kj}$, so that the products in (36) are defined.

Verification of these three claims, (34) to (36), is left for the exercises.

**EXAMPLE 8.** If

$$
A = \begin{bmatrix}
2 & 4 & 1 \\
-1 & 3 & 0 \\
5 & 4 & 6
\end{bmatrix} = \begin{bmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{bmatrix},
$$

and

$$
B = \begin{bmatrix}
0 & 1 & 3 \\
2 & -4 & 8 \\
5 & 2 & 6
\end{bmatrix} = \begin{bmatrix}
B_{11} & B_{12} \\
B_{21} & B_{22}
\end{bmatrix},
$$

then

$$
AB = \begin{bmatrix}
A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\
A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22}
\end{bmatrix}.
$$

Working out the the elements,

$$
A_{11}B_{11} + A_{12}B_{21} = \begin{bmatrix}
2 & 4 \\
-1 & 3
\end{bmatrix} \begin{bmatrix}
0 & 1 \\
2 & -4
\end{bmatrix} + \begin{bmatrix}
1 & 0
\end{bmatrix} [5, 8]
$$
10.2. Matrices and Matrix Algebra

\[
\begin{bmatrix}
8 & -14 \\
6 & -13
\end{bmatrix}
+ 
\begin{bmatrix}
5 & 8 \\
0 & 0
\end{bmatrix}
= 
\begin{bmatrix}
13 & -6 \\
6 & -13
\end{bmatrix},
\]

\[
A_{11}B_{12} + A_{12}B_{22} = 
\begin{bmatrix}
2 & 4 \\
-1 & 3
\end{bmatrix}
+ 
\begin{bmatrix}
3 & 1 \\
-1 & 0
\end{bmatrix}
= 
\begin{bmatrix}
5 & 4 \\
-6 & -6
\end{bmatrix},
\]

\[
A_{21}B_{11} + A_{22}B_{21} = 
\begin{bmatrix}
0 & 1 \\
2 & -4
\end{bmatrix}
+ 
[6][5, 8]
= 
[8, -11] + [30, 48] = [38, 37],
\]

and

\[
A_{21}B_{12} + A_{22}B_{22} = [5, 4] 
\begin{bmatrix}
3 \\
-1
\end{bmatrix}
+ [6][2] = [23],
\]

so

\[
AB = 
\begin{bmatrix}
13 & -6 & 4 \\
6 & -13 & -6 \\
38 & 37 & 23
\end{bmatrix}
= 
\begin{bmatrix}
13 & -6 & 4 \\
6 & -13 & -6 \\
38 & 37 & 23
\end{bmatrix},
\]

which is the same result as obtained by the multiplication of the unpartitioned matrices \(A\) and \(B\).

COMMENT 1. By no means do we claim that partitioning made the preceding calculation easier; our aim was simply to illustrate the idea of partitioning.

COMMENT 2. The partition

\[
B = 
\begin{bmatrix}
0 & 1 & 3 \\
2 & -4 & -1 \\
5 & 8 & 2
\end{bmatrix}
= 
\begin{bmatrix}
B_{11} & B_{12} \\
B_{21} & B_{22}
\end{bmatrix},
\]

in place of (38), would also be conformable with (37) for the multiplication \(AB\) and would lead to the same result, (39). On the other hand, neither of the partitions

\[
B = 
\begin{bmatrix}
0 & 1 & 3 \\
2 & -4 & -1 \\
5 & 8 & 2
\end{bmatrix}
\quad \text{or} \quad
B = 
\begin{bmatrix}
0 & 1 & 3 \\
2 & -4 & -1 \\
5 & 8 & 2
\end{bmatrix}
\]

would be conformable with (37) for the product \(AB\). For example, the term \(A_{11}B_{11}\) would not be defined then since \(A_{11}\) has two columns whereas the \(B_{11}\)'s in (41) have only one row.

Let us close by giving two results, for reference, that will be used later on. Both use partitioning to work out the product of two matrices, \(A\) and \(B\). First, if we partition \(B\) into columns \(c_1, \ldots, c_n\), then

\[
AB = A \begin{bmatrix} c_1, \ldots, c_n \end{bmatrix} = \begin{bmatrix} Ac_1, \ldots, Ac_n \end{bmatrix}.
\]
Chapter 10. Matrices and Linear Equations

Second, if we also partition $A$ into rows $r_1, \ldots, r_m$, then

$$AB = \begin{bmatrix} r_1 \\ \vdots \\ r_m \end{bmatrix} \begin{bmatrix} c_1 & \cdots & c_n \end{bmatrix} = \begin{bmatrix} r_1 \cdot c_1 & \cdots & r_1 \cdot c_n \\ \vdots & \vdots \\ r_m \cdot c_1 & \cdots & r_m \cdot c_n \end{bmatrix},$$

(43)

That is, the $i, j$ element of $AB$ is $r_i$ dotted with $c_j$.

Closure. We define matrices and three arithmetic operations or matrices: addition, multiplication by a scalar, and multiplication. Subtraction is accounted for by addition and multiplication by a scalar: $A - B = A + (-1)B$. But no division operation is defined for matrices.

It is emphasized that multiplication is not commutative (i.e., $AB \neq BA$ in general). This "failure" and several others are listed in Theorem 10.2.2. It is suggested that these shortcomings of the Cayley definition of matrix multiplication are more than offset by the fact that it permits us to express a system of $m$ linear algebraic equations in the $n$ unknowns $x_1, \ldots, x_n$ in compact matrix form as $Ax = c$.

Computer software. As mentioned in Section 8.3, the Maple system contains many linear algebra commands within the linalg package, among which evalm is especially useful. For instance, to evaluate $(AB)^2P^3 - 5Q$, where

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -1 \\ 0 & 2 \end{bmatrix}, \quad P = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad Q = \begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix},$$

first enter

\begin{verbatim}
with(linalg):
A := array([[1, 2], [3, 0]]):
\end{verbatim}

and return. Then enter

\begin{verbatim}
A := array([[1, 2], [3, 0]]):
\end{verbatim}

and return. (If you wish the A matrix to be printed, type a semicolon in place of the final colon.) Similarly, for

$$B := \text{array}([[1, -1], [0, 2]]):$$

(45)

$$P := \text{array}([[1, 1], [1, 1]]):$$

(46)

$$Q := \text{array}([[1, -1], [2, 3]]):$$

(47)

Then enter

\begin{verbatim}
evalm((A &* B) &* P &* 3 - 5 * Q):
\end{verbatim}

and return. The printed output is the result

$$\begin{bmatrix} 11 & 21 \\ 38 & 33 \end{bmatrix}$$
Note that matrix multiplication of A and B is denoted as $A \times B$ (not $A \ast B$). Exponentiation to a positive integer power is denoted with $^n$, and multiplication of a matrix by a scalar is denoted by $\ast$. If we want $((AB)^2P^3 - 5Q)^4$, we can use a quotation mark to carry the matrix \[
abla \begin{bmatrix}
1 & 11 \\
21 & 21 \\
38 & 33 \\
\end{bmatrix}
\] forward. Thus, enter
\[\text{evalm}("4")\]
and return. The result is
\[
\begin{bmatrix}
2389489 & 2592744 \\
4691632 & 5105697 \\
\end{bmatrix}
\]

Alternative to the array format indicated above, we can use a matrix format. For instance,
\[
A := \text{matrix}(2, 2, [1, 2, 3, 0])
\]
is equivalent to the array format shown above, where the "2, 2" denotes that A is a $2 \times 2$ matrix.

**EXERCISES 10.2**

1. Given the matrices
\[
A = \begin{bmatrix}
0 & 3 \\
2 & -5 \\
1 & 10
\end{bmatrix}, \quad B = \begin{bmatrix}
5 & -1 \\
0 & 2
\end{bmatrix},
\]
work out whichever of the products $AB$, $BA$, $Ax$, $xA$, $xB$, $yB$, $A^2$, $B^2$, $x^2$, $xy$, and $yx$ are defined.

2. Let $A$ be $6 \times 4$, $B$ be $4 \times 4$, $C$ be $4 \times 3$, $D$ and $E$ be $3 \times 1$.
Determine which of the following are defined, and for those that are, give the form of the resulting matrix.
(a) $A^{10}$
(b) $B^{10}$
(c) $ABC$
(d) $ABCD$
(e) $ACBD$
(f) $CD + E$
(g) $C(2D - E)$
(h) $AB + AC$
(i) $BC + CB$
(j) $3BA - 5CD$

3. Evaluate the products
\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & 2 & 0 \\
0 & 0 & 3
\end{bmatrix}\begin{bmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9
\end{bmatrix}
\text{and}
\]

4. Suppose that $A$ is $n \times n$, $x$ is $n \times 1$, and $c$ is a scalar. Can we re-express $Ax = cx$ as $(A - c)x = 0$? Explain.

5. If $A$ and $B$ are square matrices of the same order, are the following correct? Explain.
(a) $(A + B)^2 = A^2 + 2AB + B^2$
(b) $(A + B)(A - B) = A^2 - B^2$
(c) $(AB)^2 = A^2B^2$
(d) $(AB)^3 = A^3B^3$

6. (a) If $p$ is a positive integer, does $A$ need to be square for $A^p$ to be defined? Explain.
(b) Let $A$ be $m \times n$ and $B$ be $p \times q$. What restrictions, if any, need must be satisfied by $m, n, p, q$ if $(AB)^2$ is to exist (i.e., be defined)?

7. Expand each of the following [e.g., the "expanded" version of $(A + B)C$ would be $AC + BC$], assuming that all of the matrices are suitably conformable. Justify each step by citing the relevant equation number in Theorem 10.2.1 or 10.2.3.
(a) $(2A + B)(A + 2B)$
(b) $(A + B)(C + D + E)$
(c) $(A + B)^2$
(d) $(A - 3I)(2A + I)$
8. Given \( A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, C = \begin{bmatrix} 0 & 3 \\ 0 & 0 \end{bmatrix}, \) and \( D = \begin{bmatrix} 0 & 5 & 7 \\ 0 & 0 & 8 \\ 0 & 0 & 0 \end{bmatrix}, \) evaluate each of the following.

(a) \( A^{100} \)
(b) \( B^{100} \)
(c) \( C^{100} \)
(d) \( D^{100} \)
(e) \( (ABC)^3 \)
(f) \( (CBA)^3 \)
(g) \( B^4C^4 \) and \( (BC)^4 \)
(h) \( C^3B^3 \) and \( (CB)^3 \)

9. Any diagonal matrix whose diagonal elements are all equal is called a **scalar matrix**. If

\[
S = \begin{bmatrix} k & 0 & \cdots & 0 \\ 0 & k \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & k \end{bmatrix}
\]

is \( n \times n \), and \( A \) is any \( n \times n \) matrix, show that

\[
AS = SA = kA.
\]

What can be said if, instead, \( A \) is \( m \times n \) \((m \neq n)\)?

10. If for any given vector \( x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \), the product \( Ax \) is the column vector given below, find \( A \).

(a) \( \begin{bmatrix} x_1 - 3x_4 \\ x_2 + x_4 - x_1 \\ x_3 + x_2 \\ x_3 + x_4 \end{bmatrix} \)
(b) \( \begin{bmatrix} x_2 + x_4 - x_1 \\ x_1 + 5x_3 \end{bmatrix} \)
(c) \( \begin{bmatrix} 2x_1 - x_3 - x_4 \\ x_1 + x_2 \\ x_3 + x_4 \end{bmatrix} \)
(d) \( \begin{bmatrix} x_2 + x_4 \\ x_1 - x_3 - x_4 \\ -2x_1 + x_3 - x_4 \end{bmatrix} \)
(e) \( \begin{bmatrix} x_4 \\ x_3 \\ x_2 \\ x_1 \end{bmatrix} \)
(f) \( \begin{bmatrix} x_1 + 3x_4 \\ x_2 - 2x_4 \\ x_3 + x_1 \\ x_3 - 2x_1 \end{bmatrix} \)

11. Make up a specific pair of matrices, \( A \) and \( B \), both nonzero, such that \( AB = 0 \), where

(a) \( A \) is \( 2 \times 2 \) and \( B \) is \( 2 \times 2 \)
(b) \( A \) is \( 5 \times 2 \) and \( B \) is \( 2 \times 2 \)
(c) \( A \) is \( 1 \times 2 \) and \( B \) is \( 2 \times 4 \)
(d) \( A \) is \( 4 \times 3 \) and \( B \) is \( 3 \times 2 \)

12. Given the partitioned matrices \( A \) and \( B \), below, carry out the products \( A^2 \) and \( AB \) for those cases in which the partitioning is suitable, i.e., conformable. If the partitioning is not suitable, explain why it is not.

\[
(a) A = \begin{bmatrix} 2 & 0 & -1 \\ 1 & -1 & 0 \\ 5 & 2 & 4 \end{bmatrix}, B = \begin{bmatrix} 3 & 6 & 2 \\ 0 & 1 & 0 \\ 3 & -4 & 7 \end{bmatrix}
\]
\[
(b) A = \begin{bmatrix} 2 & 0 & -1 \\ 1 & -1 & 0 \\ 5 & 2 & 4 \end{bmatrix}, B = \begin{bmatrix} 3 & 6 & 2 \\ 0 & 1 & 0 \\ 3 & -4 & 7 \end{bmatrix}
\]
\[
(c) A = \begin{bmatrix} 1 & -1 & 0 \\ 5 & 2 & 4 \end{bmatrix}, B = \begin{bmatrix} 0 & 1 & 0 \\ 3 & -4 & 7 \end{bmatrix}
\]

13. Show that \( c_1x_1 + c_2x_2 + \cdots + c_nx_n \) can be expressed in the form

\[
\begin{bmatrix} x_1, x_2, \ldots, x_n \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}
\]

14. (a) If two unpartitioned matrices are not conformable for addition, can they be rendered conformable by suitable partitioning? Explain.
(b) Same as part (a), but for multiplication.

15. If there is some positive integer \( p \) such that \( A^p = 0 \), then \( A \) is said to be **nilpotent** (mnemonic: “potentially nil”). A square matrix \( A = \{a_{ij}\} \) such that \( a_{ij} = 0 \) for all \( i > j \) is said to be an upper triangular matrix; if \( a_{ij} = 0 \) for all \( i < j \), then \( A \) is a lower triangular matrix. A matrix is said to be triangular if it is either upper triangular or lower triangular.

(a) Every upper triangular matrix with null main diagonal (so that \( a_{ij} = 0 \) for all \( i \geq j \)) is nilpotent. Verify this result for second-, third-, and fourth-order matrices.
(b) In fact, show that every upper triangular matrix (of finite order) with null main diagonal is nilpotent. HINT: Use partitioning and induction.
(c) Is every lower diagonal matrix with null main diagonal nilpotent? Explain.
(d) If \( A^p = 0 \), show that \((I + A + A^2 + \cdots + A^{p-1})(I - A) = (I - A)(I + A + A^2 + \cdots + A^{p-1}) = I.\)

16. If \( A^2 = I \), then \( A \) is called **involutory**.

(a) Show, using Theorems 10.2.1 to 10.2.3 and (27), that \( A \) is involutory if and only if

\[
(I - A)(I + A) = 0.
\]

(b) Give an example of an involutory matrix other than \( I \) and \(-I\). Thus, observe that \( A^2 = I \) does not imply that \( A = \pm I \).
(c) Determine the most general \( 2 \times 2 \) matrix that is involutory.

17. (a) Prove (9a) to (9c).
(b) Prove (9d) to (9f).
(c) Prove (9g) to (9i).

18. In Theorem 10.2.2, prove
(a) (i)  (b) (ii)  (c) (iii)  (d) (iv)

19. (a) Verify (24).
(b) Verify (27).
(c) Verify (30a) and (30b).
(d) Verify (30c) and (30d).
(e) Show that (30e) follows from (30a)–(30d).

20. Show that the most general matrix that commutes with
\[
\begin{bmatrix}
1 & 2 \\
3 & 4
\end{bmatrix}
\]
is
\[
\begin{bmatrix}
\alpha & \beta \\
2\beta/3 & \alpha
\end{bmatrix},
\]
where \(\alpha\) and \(\beta\) are arbitrary.

21. Given \(A\), find the most general matrix \(B\) such that \(AB = 0\).
(a) \(A = \begin{bmatrix} 1 & 2 \\
3 & 4 \end{bmatrix}\)
(b) \(A = \begin{bmatrix} 2 & 3 \\
0 & 0 \end{bmatrix}\)
(c) \(A = \begin{bmatrix} 0 & 0 \\
5 & 0 \end{bmatrix}\)
(d) \(A = \begin{bmatrix} 2 & 3 \\
4 & 6 \end{bmatrix}\)

22. Explore and discuss the advantages and disadvantages of defining \(c_{ij} = a_{ij}b_{ij}\) in place of the Cayley product (13).

23. (Transition probability matrix) Selling their valuables, professors \(A\) and \(B\) raise \$2 apiece, and proceed to match coins at \$1 per match. There arise five possible states:

\[
\begin{array}{cccccc}
S_1 & S_2 & S_3 & S_4 & S_5 \\
04 & 13 & 22 & 31 & 40
\end{array}
\]

In state \(S_2\), for example, \(A\) has \$1 and \(B\) has \$3. If either player is bankrupt (\(A\) is bankrupt in state \(S_1\), \(B\) in state \(S_5\)), the game is over. Let \(p_{ij}^{(n)}\) be the \(n\)-step transition probability, i.e., the probability of changing from state \(S_i\) to state \(S_j\) in \(n\) matches. For \(n = 1\) we see that

\[
p^{(1)} = \{p_{ij}^{(1)}\} = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 \\
1/2 & 1/2 & 0 & 0 & 0 \\
0 & 1/3 & 1/3 & 1/3 & 0 \\
0 & 0 & 1/2 & 1/2 & 1/2 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix}.
\]

For example, beginning in \(S_2\), say, one match will necessarily move us to \(S_1\) or \(S_3\), with 50% probability in each case; thus \(p_{21}^{(1)} = p_{23}^{(1)} = 1/2\). Precisely \$1 will change hands so that \(p_{22}^{(1)} = p_{24}^{(1)} = p_{26}^{(1)} = 0\). Further, once in \(S_1\) (or \(S_5\)) we remain there (according to the rules), so \(p_{11}^{(1)} = 1\, p_{12}^{(1)} = p_{13}^{(1)} = p_{14}^{(1)} = p_{15}^{(1)} = 0\). Show, by any convincing arguments or discussion, that

\[
p_{ij}^{(2)} = \sum_{k=1}^{5} p_{ik}^{(1)} p_{kj}^{(1)} \quad \text{and} \quad p_{ij}^{(3)} = \sum_{k=1}^{5} p_{ik}^{(2)} p_{kj}^{(1)} \quad \text{etc.}
\]
or, in matrix notation, \(P^{(2)} = [P^{(1)}]_2, P^{(3)} = [P^{(1)}]_3, \quad \text{etc.}\)
Use this result to determine \(P^{(2)}\) and \(P^{(3)}\). What is the probability that \(A\) is bankrupt after (at most) three matches if \(A\) starts with \$2? \$3? \$1? NOTE: \(P^{(1)}\) is an example of a Markov matrix. We meet Markov matrices again in Chapter 11.

24. Let

\[
A = \begin{bmatrix}
2 & -1 \\
3 & 0 \\
1 & 4
\end{bmatrix}, \quad B = \begin{bmatrix}
5 & 3 & 25 \\
2 & 0.1 & -6
\end{bmatrix},
\]

\[
C = \begin{bmatrix}
9 & 1 & -1 \\
2 & 0 & 7 \\
0 & 4 & 6
\end{bmatrix}, \quad F = \begin{bmatrix}
6 & 5 & 1 \\
4 & 1 & -6
\end{bmatrix}.
\]

NOTE: The letters \(D, E, I, O, S,\) and \(W\) are "protected," in Maple, for other purposes. The problem: use computer software to evaluate

(a) \(6AB - 9C\)
(b) \((AB)^3 + 5C^2\)
(c) \(6AFB - 2C^3\)
(d) \(4BAF\)
(e) \((BCAF)^4\)
(f) \(2CA + 37.3A\)
(g) \((CAB)^2\)
(h) \(0.73BA + 1.6F^6\)

10.3 The Transpose Matrix

We continue the development of Section 10.2 by introducing the "transpose" of a matrix. Given any \(m \times n\) matrix \(A = \{a_{ij}\}\), we define the transpose of \(A\), denoted
Chapter 10. Matrices and Linear Equations

482

as \( A^T \) and read as "A-transpose," as

\[
A^T = \{ a_{ji} \} = \begin{bmatrix}
  a_{11} & a_{21} & \cdots & a_{m1} \\
  a_{12} & a_{22} & \cdots & a_{m2} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{1n} & a_{2n} & \cdots & a_{mn}
\end{bmatrix},
\]

(1)

that is, the \( n \times m \) matrix is obtained by interchanging the rows and columns of \( A \): the first row of \( A \) becomes the first column of \( A^T \), the second row of \( A \) becomes the second column of \( A^T \), and so on. Or, the first column of \( A \) becomes the first row of \( A^T \), and so on. That is, if we denote the \( i, j \) elements of \( A \) and \( A^T \) as \( a_{ij} \) and \( a^T_{ij} \), respectively, then

\[
a^T_{ij} = a_{ji}.
\]

(2)

Be clear that \( A^T \) is not \( A \) to the \( T \)th power; it is the transpose of \( A \).

**EXAMPLE 1.** If

\[
A = \begin{bmatrix}
  2 & 0 & 1 \\
-1 & 3 & 5 \\
 4 & 6 & 7
\end{bmatrix}, \quad B = \begin{bmatrix}
  2 \\
  6 \\
  7
\end{bmatrix}, \quad \text{and} \quad C = [1, -8, 9],
\]

then

\[
A^T = \begin{bmatrix}
  2 & -1 & 4 \\
  0 & 3 & 6 \\
  1 & 5 & 7
\end{bmatrix}, \quad B^T = [2, 6, 7], \quad \text{and} \quad C^T = \begin{bmatrix}
  1 \\
  -8 \\
  9
\end{bmatrix}.
\]

**THEOREM 10.3.1** Properties of the Transpose

\[
(A^T)^T = A, \quad (A + B)^T = A^T + B^T, \quad (\alpha A)^T = \alpha A^T, \quad (AB)^T = B^T A^T,
\]

(3a) (3b) (3c) (3d)

where it is assumed in (3b) that \( A \) and \( B \) are conformable for addition, and in (3d) that they are conformable for multiplication.

**Proof:** Proof of (3a)–(3c) is left for the exercises. To prove (3d), let \( AB \equiv C = \{ c_{ij} \} \). By the definition of matrix multiplication,

\[
c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj},
\]

(4)
Thus,
\[ c_{ij}^T = c_{ji} = \sum_{k=1}^{n} a_{jk} b_{ki} = \sum_{k=1}^{n} b_{ki} a_{jk} = \sum_{k=1}^{n} b_{kj}^T a_{kj}. \] (5)

Having returned, at the end of (5), to the pattern \( (\cdot)_{ij} = \sum (\cdot)_{ik} (\cdot)_{kj} \) as in (4), we can conclude from (5) that
\[ C^T = B^T A^T \quad \text{or} \quad (AB)^T = B^T A^T, \]
as was to be proved. Understand that the third equality in (5) is not equivalent to the matrix statement \( AB = BA \) which, we recall from Section 10.2, is generally untrue. It is simply the scalar statement that \( a_{jk} b_{ki} = b_{ki} a_{jk} \), which is true because the multiplication of scalars is commutative [e.g., (2)(3) = (3)(2) = 6].

The striking feature of (3d) is the reversal in the order: \( (AB)^T \) on the left, \( B^T A^T \) on the right. Notice how (3d) checks “dimensionally”:
\[
\begin{align*}
[(m \times n)(n \times p)]^T &= (n \times p)^T (m \times n)^T \\
(m \times p)^T &= (p \times n)(n \times m) \\
(p \times m) &= (p \times m).
\end{align*}
\] (6)

Naturally, (6) does not prove (3d), but it provides a useful check, just as we check the physical units (such as force, mass, length, and time) of an equation to be sure that they are consistent.

**EXAMPLE 2.** If \( A = \begin{bmatrix} 4 & 2 & -5 \\ 0 & 1 & 3 \end{bmatrix} \) and \( B = \begin{bmatrix} 6 \\ 0 \\ -1 \end{bmatrix} \), say, then
\[
AB = \begin{bmatrix} 4 & 2 & -5 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 29 \\ -3 \end{bmatrix}, \quad \text{so} \quad (AB)^T = [29, -3]
\]

and
\[
B^T A^T = [6, 0, -1] \begin{bmatrix} 4 & 0 & -5 \\ 2 & 1 & 3 \end{bmatrix} = [29, -3],
\]
in agreement with (3d).

Furthermore, it follows from (3d) that
\[ (ABC)^T = C^T B^T A^T, \quad (ABCD)^T = D^T C^T B^T A^T, \] (7)
and so on (Exercise 2).
Chapter 10. Matrices and Linear Equations

Using lowercase boldface letters for matrices that happen to be column vectors from now on, let

\[ x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}. \]

Then the standard dot product

\[ x \cdot y = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = \sum_{j=1}^{n} x_j y_j \]

can be expressed compactly, in matrix language, as

\[ x \cdot y = x^T y \] (8)

or, equivalently, as \( y^T x \), although not as \( xy^T \) or \( yx^T \), which expressions represent \( n \times n \) matrices!

**EXAMPLE 3.** If \( x = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \) and \( y = \begin{bmatrix} 5 \\ 2 \end{bmatrix} \), say, then

\[ x \cdot y = x^T y = [3, 1] \begin{bmatrix} 5 \\ 2 \end{bmatrix} = 15 + 2 = 17, \]

or

\[ x \cdot y = y^T x = [5, 2] \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 15 + 2 = 17, \]

whereas

\[ xy^T = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \begin{bmatrix} 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 15 & 6 \\ 5 & 2 \end{bmatrix} \neq x \cdot y, \]

and

\[ yx^T = \begin{bmatrix} 5 \\ 2 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 15 & 5 \\ 6 & 2 \end{bmatrix} \neq x \cdot y. \]

Finally, two more definitions: if

\[ A^T = A \] (9)

we say that \( A \) is symmetric, and if

\[ A^T = -A. \] (10)

*There is a small difficulty here: \( x \cdot y \) is to be scalar, whereas \( x^T y \) is a \( 1 \times 1 \) matrix. Thus, what we really intend, by \( x^T y \) in (8), is not the \( 1 \times 1 \) matrix, but rather the scalar element inside it. That could be noted by writing \( x \cdot y = (x^T y)_{11} \) instead, but it will be simpler to leave (8) intact, with the understanding that the right-hand side is a scalar.*
we say that it is **skew-symmetric** (or antisymmetric). For either of these properties to apply \(A\) must be square, since otherwise \(A^T\) and \(A\) would be of different form. And for \(A\) to be skew-symmetric all of its diagonal elements must be zero, since (10) implies that \(a_{ji} = -a_{ij}\), or if we set \(i = j\), \(a_{ii} = -a_{ii}\); thus \(2a_{ii} = 0\), so \(a_{ii} = 0\) for each \(i\).

It would be reasonable to imagine that the likelihood of encountering purely symmetric or skew-symmetric matrices in applications would be slim. On the contrary, we shall see that symmetric matrices arise frequently, and that their symmetry is often a consequence of fundamental physical principles, rather than chance.

**Closure.** The key points in this section are the defining of the transpose of any matrix, and the results \((AB)^T = B^T A^T\) and \(x \cdot y = x^T y\) (or \(y^T x\)). Note that the transpose notation is sometimes used to save space. For instance, it takes less vertical space on the page to write \(x^T = [7, 2]\) than \([7, 2]\).

**Computer software.** The relevant *Maple* function, to take the transpose of a matrix \(A\), is the command `transpose`, within the `linalg` package. For instance, to obtain the transpose of \(A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}\) enter

\[
\text{with(linalg):}
\]

to access the `transpose(A)` command. Enter

\[
A := \text{array}([[1, 2, 3], [4, 5, 6]]);
\]

and return, then enter

\[
\text{transpose}(A);
\]

and return. The output is

\[
\begin{bmatrix}
1 & 4 \\
2 & 5 \\
3 & 6
\end{bmatrix}
\]

**EXERCISES 10.3**

1. (a) If \(x = [3, -3]^T\) and \(y = [1, 2]^T\), work out \(x^T y\) and \(xy^T\).
   (b) If \(x = [4, -1, 0]^T\) and \(y = [1, 2, 3]^T\), work out \(x^T y\) and \(xy^T\).
   (c) If \(x = [0, 4, -2, 1]^T\) and \(y = [3, 0, 1, -2]^T\), work out \(x^T y\) and \(xy^T\).

2. Show that (7) follows from (3d).

3. Recall that in general \(AB \neq BA\), and that a necessary (but not sufficient) condition for equality to hold is that both \(A\) and \(B\) be square and of the same order. Perhaps a sufficient condition is that \(A\) and \(B\) both be of the same order and symmetric. Prove or disprove this hypothesis.

4. Verify \((ABC)^T = C^T B^T A^T\) directly, for
   \[
   (a) \ A = \begin{bmatrix} 5 & -2 \\ 0 & 1 \\ 1 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 7 \end{bmatrix}, \quad C = [3, 1, 2, 9]\]
(b) $A = \begin{bmatrix} 1 & 2 \\ 5 & 4 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$, $C = \begin{bmatrix} 4 & -1 & 5 \\ 0 & 3 & 0 \end{bmatrix}$

(c) $A = [5, 3, 0]$, $B = \begin{bmatrix} -2 & 1 \\ 6 & 4 \\ 5 & 3 \end{bmatrix}$, $C = \begin{bmatrix} 4 \\ 8 \end{bmatrix}$

(d) $A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 3 & 4 \\ 1 & 2 \\ 0 & 1 \end{bmatrix}$, $C = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$

5. Prove the properties (3a), (3b), and (3c).

6. Even if a (square) matrix is neither symmetric nor skew-symmetric it can be decomposed as the sum of a symmetric matrix and a skew-symmetric matrix. Specifically, writing

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T)$$

show that $A_1$ is symmetric and $A_2$ is skew-symmetric. NOTE: Equation (6.1) is but one such decomposition. For instance, we saw in Chapter 9 that a vector $v$ in 3-space can be decomposed as the sum of two vectors, one lying in a given plane (the "projection of $v$" onto that plane) and the other perpendicular to that plane; when we study vector fields we will see that any vector field can be decomposed as the sum of two fields, one irrational and the other solenoidal; when we study Fourier series we will see that any function $f(x)$ can be decomposed as the sum of two functions, one even and the other odd; and so on. Thus, such decompositions are not uncommon in mathematics.

7. Decompose the given matrix as the sum of two matrices, one symmetric and one skew symmetric, as explained in Exercise 6.

(a) \begin{bmatrix} 3 & 2 \\ 1 & -5 \end{bmatrix}  \quad (b) \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}  \quad (c) \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \end{bmatrix}  \quad (d) \begin{bmatrix} 8 & -2 \\ 4 & 0 \end{bmatrix}  \quad (e) \begin{bmatrix} 9 & 8 & 7 \\ 6 & 5 & 4 \\ 3 & 2 & 1 \end{bmatrix}  \quad (f) \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \\ 2 \\ -1 \\ 0 \\ 3 \\ 0 \end{bmatrix}

8. (Quadratic forms) The quadratic function $ax^2 + by^2 + cxy$ is said to be a quadratic form in $x$ and $y$, $ax^2 + by^2 + cz^2 + dxy + exz + fyz$ is a quadratic form in $x, y, z$, and so on. Every quadratic form in the $n$ variables $x_1, \ldots, x_n$ can be expressed, in matrix notation, as $x^T Ax$, where $A$ is a symmetric $n \times n$ matrix. For each given quadratic form, determine the $A$ matrix. NOTE: Quadratic forms will be important to us in Chapter 11.

(a) $6x_1^2 + x_2^2 - 8x_1x_2$ in $x_1, x_2$
(b) $x_1^2 - 3x_1^2 + 6x_1x_2$ in $x_1, x_2$
(c) $4x_1^2 + x_2^2 - x_3^2 + 8x_1x_2 + 3x_1x_3 - 2x_2x_3$ in $x_1, x_2, x_3$
(d) $x_1^2 - 4x_2^2 + 2x_1x_3 - 10x_2x_3$ in $x_1, x_2, x_3$
(e) $3x_2^2 + x_3^2 - 6x_1x_3$ in $x_1, x_2, x_3$
(f) $x_1^2 + x_2^2 + x_3^2 + x_1x_3 + x_1x_4 + x_3x_4$ in $x_1, x_2, x_3, x_4$

9. Show that if $A$ is an $m \times n$ matrix, then $AA^T$ is symmetric. HINT: There is often an inclination to work out a problem like this using "brute force," i.e., by actually writing out the $A$ and $A^T$ matrices, multiplying them, and examining the resulting matrix to see if it is symmetric. Whenever possible, we advise against such an approach. In this problem, for example, we wish to show that $C^T = C$, where $C$ is short for $AA^T$; i.e., we wish to show that $(AA^T)^T = AA^T$, and this can be done (in one or two short lines) using the properties stated in Theorem 10.3.1.

10. Prove that the product $AB$ need not be symmetric, even if $A$ and $B$ are both symmetric and of the same order.

11. Let

$$A = \begin{bmatrix} 4 & 1 & 2 \\ 0 & 5 & 7 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & -4 & 2 \\ 8 & 1 & 4 \end{bmatrix}$$

Use computer software, such as Maple, to evaluate

(a) $AB^T$  \quad (b) $BA^T$
(c) $(AB^T)^5$  \quad (d) $(BA^T)^T$
(e) $(B^T A)^8$  \quad (f) $2A^T - 7.3B^T$

### 10.4 Determinants

In this section we introduce a scalar quantity associated with every square matrix, the so-called "determinant" of the matrix. We denote the determinant of an $n \times n$
matrix $A = \{a_{ij}\}$ as

$$\det A = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}, \tag{1}$$

that is, with straight line braces instead of square brackets. Determinants are prominent in Chapter 3 in connection with the linear dependence or independence of sets of functions, especially with sets of solutions of linear homogeneous differential equations. More generally, they play a key role in the theory of systems of linear algebraic equations, as discussed in Section 10.5.

The determinant of an $n \times n$ matrix $A = \{a_{ij}\}$ is defined by the cofactor expansion

$$\det A \equiv \sum_{k=1}^{n} a_{jk} A_{jk}, \tag{2}$$

where the summation is carried out on $j$ for any fixed value of $k \ (1 \leq k \leq n)$ or on $k$ for any fixed value of $j \ (1 \leq j \leq n)$. $A_{jk}$ is called the cofactor of the $a_{jk}$ element and is defined as

$$A_{jk} \equiv (-1)^{j+k} M_{jk}, \tag{3}$$

where $M_{jk}$ is called the minor of $a_{jk}$, namely, the determinant of the $(n-1) \times (n-1)$ matrix that survives when the row and column containing $a_{jk}$ (the $j$th row and the $k$th column) are struck out.

For example, if

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

then

$$M_{11} = \begin{vmatrix} 3 & 2 \\ 5 & 6 \end{vmatrix}, \quad M_{12} = \begin{vmatrix} 0 & 2 \\ 1 & 6 \end{vmatrix}, \quad \text{and} \quad M_{23} = \begin{vmatrix} 4 & 7 \\ 1 & 5 \end{vmatrix}. \quad \text{Thus, if} \ A \ \text{is} \ n \times n, \ \text{then the right-hand side of (2) is a linear combination of} \ n \ \text{determinants, each of which is} \ (n-1) \times (n-1). \ \text{Each of these, in turn, may be expressed as a linear combination of} \ (n-2) \times (n-2) \ \text{determinants, and so on, until we have a (perhaps large) number of} \ 1 \times 1 \ \text{determinants. Thus, the definition (2) is logically incomplete until we define a} \ 1 \times 1 \ \text{determinant, which we do as follows:}$$

$$\det \begin{bmatrix} a_{11} \end{bmatrix} = \begin{vmatrix} a_{11} \end{vmatrix} = a_{11}. \tag{4}$$

That is, the determinant of a $1 \times 1$ matrix $\begin{bmatrix} a_{11} \end{bmatrix}$ is simply $a_{11}$ itself. CAUTION: In the present context, the braces around $a_{11}$, in the middle member of (4), denote determinant, not absolute value. For instance, $\det[-6] = |-6| = -6$. 

10.4. Determinants 487
Using (2) and (4), let us work out the determinant of any 2 × 2 matrix. Recalling that we can sum on \( j \) with \( k \) fixed, or vice versa, let us sum on \( k \) with \( j = 1 \), say. Then

\[
\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \sum_{k=1}^{2} a_{1k} A_{1k} = a_{11} A_{11} + a_{12} A_{12} = a_{11}(+1)M_{11} + a_{12}(-1)M_{12} = a_{11} A_{11} + a_{12} A_{12} = a_{11}(-1)^{1+1}M_{11} + a_{12}(-1)^{1+2}M_{12} = a_{11}(+1)|a_{22}| + a_{12}(-1)|a_{21}| = a_{11}a_{22} - a_{12}a_{21},
\]

which result is probably familiar to you from earlier studies. Observe that the \((-1)^{j+k}\) in (3) is simply +1 if \( j + k \) is even, and -1 if \( j + k \) is odd.

**EXAMPLE 1.** Let

\[
A = \begin{bmatrix} 0 & 2 & -1 \\ 4 & 3 & 5 \\ 2 & 0 & -4 \end{bmatrix}.
\]

Using (2) with \( j = 1 \), say,

\[
\det A = \sum_{k=1}^{3} a_{1k} A_{1k} = a_{11} A_{11} + a_{12} A_{12} + a_{13} A_{13} = a_{11}(+1)M_{11} + a_{12}(-1)M_{12} + a_{13}(-1)M_{13} = a_{11} M_{11} - a_{12} M_{12} + a_{13} M_{13} = (0) \begin{vmatrix} 3 & 5 \\ 0 & -4 \end{vmatrix} - (2) \begin{vmatrix} 4 & 5 \\ 2 & -4 \end{vmatrix} + (-1) \begin{vmatrix} 4 & 3 \\ 2 & 0 \end{vmatrix} = 0 - (2)(-16 - 10) + (-1)(0 - 6) = 58.
\]

This particular choice \((j = 1, \text{ summing on } k)\) is said to be cofactor expansion about the first row since (7) is the sum of the first-row elements \( a_{11}, a_{12}, a_{13} \), multiplied by their cofactors.

According to (2), we can expand about any row or column. Let us illustrate that the same answer is indeed obtained if we expand about other rows or columns.

*Expansion about second row* [i.e., set \( j = 2 \) in (2), and sum on \( k \)]:

\[
\det A = \sum_{k=1}^{3} a_{2k} A_{2k} = a_{21} A_{21} + a_{22} A_{22} + a_{23} A_{23} = a_{21}(-1)M_{21} + a_{22}(+1)M_{22} + a_{23}(-1)M_{23} = -(4) \begin{vmatrix} 2 & -1 \\ 0 & -4 \end{vmatrix} + (3) \begin{vmatrix} 0 & -1 \\ 2 & -4 \end{vmatrix} - (5) \begin{vmatrix} 0 & 2 \\ 2 & 0 \end{vmatrix} = -(4)(-8) + (3)(2) - (5)(-4) = 58,
\]

again.
Expansion about third column \((k = 3, \text{ sum on } j)\):

\[
\det A = \sum_{j=1}^{3} a_{j3} A_{j3} = a_{13} A_{13} + a_{23} A_{23} + a_{33} A_{33}
\]

\[
= a_{13}(+1)M_{13} + a_{23}(-1)M_{23} + a_{33}(+1)M_{33}
\]

\[
= (-1) \begin{vmatrix} 4 & 3 \\ 2 & 0 \end{vmatrix} - (5) \begin{vmatrix} 0 & 2 \\ 2 & 0 \end{vmatrix} + (-4) \begin{vmatrix} 0 & 2 \\ 4 & 3 \end{vmatrix}
\]

\[
= (-1)(-6) - (5)(-4) + (-4)(-8) = 58,
\]

once more. 

Since we may expand about any row or column, it is convenient in hand calculations to choose the row or column with the most zeros in it since those terms in the expansion then drop out.

Notice carefully that for large \(n\) the cofactor expansion process is exceedingly laborious. Even if \(n = 10\), say, which is still quite modest, (2) gives a linear combination of ten \(9 \times 9\) determinants. In turn, each of these ten is evaluated as a linear combination of nine \(8 \times 8\) determinants, and, so on! Let us see just how serious this predicament is. For estimating purposes, let us count each multiplication, addition, and subtraction as one "calculation." It can be shown (Exercise 18a) that the number of calculations \(N(n)\) required in the evaluation (by cofactor expansion) of an \(n \times n\) determinant is

\[
N(n) \sim e n!
\]

as \(n \to \infty\), where \(e \approx 2.718\) is the base of the natural logarithm, and \(n!\) is \(n\) factorial. If each calculation takes approximately one microsecond, then some time estimates are as follows. (Before reading on, we urge you to guess how long such a computer would take to evaluate a \(25 \times 25\) determinant.)

<table>
<thead>
<tr>
<th>(n)</th>
<th>Computing Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.0003 sec</td>
</tr>
<tr>
<td>10</td>
<td>10 sec</td>
</tr>
<tr>
<td>15</td>
<td>(4 \times 10^6) sec \approx 40 days</td>
</tr>
<tr>
<td>20</td>
<td>(7 \times 10^{12}) sec \approx 210,000 years</td>
</tr>
<tr>
<td>25</td>
<td>(4 \times 10^{19}) sec \approx 10^{12} years</td>
</tr>
</tbody>
</table>

It is interesting that faster computers offer no hope. For instance, even a computer that is a million times as fast would still take around \(10^9\) years to evaluate a \(25 \times 25\) determinant. And scientific calculations can easily involve determinants that are \(250 \times 250\).

It is tempting to conclude that "determinants are worthless," but let us see if we can come up with a more efficient algorithm than the cofactor expansion. A logical starting point is to first determine the various properties of determinants so that we can use them to design a better algorithm.
First, we need to introduce the idea of a "triangular" matrix. A square matrix \( A = \{a_{ij}\} \) is upper triangular if \( a_{ij} = 0 \) for all \( j < i \) and lower triangular if \( a_{ij} = 0 \) for all \( j > i \). That is,

\[
\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
0 & a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & a_{nn}
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
a_{11} & 0 & \cdots & 0 \\
a_{21} & a_{22} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} & \cdots & \cdots & a_{nn}
\end{bmatrix}
\] (9)

are upper triangular and lower triangular, respectively. If a matrix is upper triangular or lower triangular it is said to be triangular.

Here are the properties that we will need.

**PROPERTIES OF DETERMINANTS**

**D1.** If any row (or column) of \( A \) is modified by adding \( \alpha \) times the corresponding elements of another row (or column) to it, yielding a new matrix \( B \), then \( \det B = \det A \).

*Symbolically:* \( r_j \rightarrow r_j + \alpha r_k \)

**D2.** If any two rows (or columns) of \( A \) are interchanged, yielding a new matrix \( B \), then \( \det B = -\det A \).

*Symbolically:* \( r_j \leftrightarrow r_k \)

**D3.** If \( A \) is triangular, then \( \det A \) is simply the product of the diagonal elements,

\[ \det A = a_{11}a_{22}\cdots a_{nn}. \]

Of these, D3 is easily proved. For consider the general upper triangular matrix in (9). Doing a cofactor expansion about the first column gives \( a_{11} \) times an \( (n-1) \times (n-1) \) minor determinant, which is again upper triangular. Expanding the latter about its first column gives \( a_{22} \) times an \( (n-2) \times (n-2) \) minor determinant, which is again upper triangular. Repeating the process leads to \( \det A = a_{11}a_{22}\cdots a_{nn} \). Similarly for the general lower triangular matrix in (9), except that in that case we expand about the first row, repeatedly, rather than the first column.

Let us illustrate the use of the properties D1–D3 instead of the cofactor expansion.

**EXAMPLE 2.** Consider the \( A \) matrix of Example 1 again.

\[
det A = \begin{vmatrix} 0 & 2 & -1 \\ 4 & 3 & 5 \\ 2 & 0 & -4 \end{vmatrix} = -\begin{vmatrix} 2 & 0 & -4 \\ 4 & 3 & 5 \\ 0 & 2 & -1 \end{vmatrix} = -\begin{vmatrix} 2 & 0 & -4 \\ 0 & 3 & 13 \\ 0 & 2 & -1 \end{vmatrix} = -(2)(3) \left( \frac{-29}{3} \right) = 58,
\]

as obtained in Example 1. In the second equality we interchanged the first and third rows \( (r_1 \leftrightarrow r_2) \), thereby changing the sign of the determinant (D2) so we compensated by
putting the minus sign out in front. In the third equality we modified the second row by adding \(-2\) times the first row to it \(r_2 \rightarrow r_2 - 2r_1\), which step left the determinant unchanged \((D1)\). In the fourth equality we modified the third row by adding \(\frac{3}{2}\) times the second row to it \(r_3 \rightarrow r_3 - \frac{3}{2}r_2\), which step left the determinant unchanged \((D1)\). Since those steps produced a triangular matrix, we could then use D3.

The point, then, is to use some combination of D1 and D2 steps to reduce the determinant to triangular form, in which case it is evaluated easily by D3. Of course the method is quite similar to Gauss elimination, described in Section 8.3. For instance, compare D1 and D2 with the first and third elementary equation operations listed in Section 8.3. For reference purposes, we will call the method illustrated in Example 2 the method of triangularization.

It is hard to tell, from the \(3 \times 3\) calculation in Example 2, whether the method is more efficient than the cofactor expansion. However, in Exercise 18b it is shown that using triangularization the number of calculations \(N(n)\) is

\[
N(n) \sim \frac{2n^3}{3}
\]

as \(n \rightarrow \infty\). Again assuming one microsecond per calculation, (10) gives a computing time of around 0.005 second for \(n = 20\) and 0.01 second for \(n = 25\), compared with 210,000 years and \(10^{12}\) years, respectively! (Comparing (8) and (10), we can see how much faster \(n!\) grows than \(n^3\).]

The upshot is that except for small hand calculations we should avoid the cofactor expansion, and should use triangularization instead.

Although properties D1–D3 suffice for the efficient calculation of determinants, other properties are sometimes useful as well, and are listed below.

**ADDITIONAL PROPERTIES OF DETERMINANTS**

D4. If all the elements of any row or column are zero, then \(\det \mathbf{A} = 0\).

D5. If any two rows or columns are proportional to each other, then \(\det \mathbf{A} = 0\).

D6. If any row (column) is a linear combination of other rows (columns), then \(\det \mathbf{A} = 0\).

D7. If all the elements of any row or column are scaled by \(\alpha\), yielding a new matrix \(\mathbf{B}\), then \(\det \mathbf{B} = \alpha \det \mathbf{A}\).

D8. \(\det(\alpha \mathbf{A}) = \alpha^n \det \mathbf{A}\).

D9. If any one row (or column) \(\mathbf{a}\) of \(\mathbf{A}\) is separated as \(\mathbf{a} = \mathbf{b} + \mathbf{c}\), then

\[
\det \mathbf{A}_a = \det \mathbf{A}_b + \det \mathbf{A}_c
\]

where \(\mathbf{A}_a\) denotes the \(\mathbf{A}\) matrix with \(\mathbf{a}\) intact, \(\mathbf{A}_b\) denotes the \(\mathbf{A}\) matrix with \(\mathbf{b}\) in place of \(\mathbf{a}\), and similarly for \(\mathbf{A}_c\). For example,

\[
\begin{vmatrix}
6 + 2 & -3 + 1 & 5 + 4 \\
3 & 0 & 2 \\
1 & -6 & 7
\end{vmatrix}
= \begin{vmatrix}
6 & -3 & 5 \\
3 & 0 & 2 \\
1 & -6 & 7
\end{vmatrix}
+ \begin{vmatrix}
2 & 1 & 4 \\
3 & 0 & 2 \\
1 & -6 & 7
\end{vmatrix}
\]
D10. The determinant of $A$ and its transpose are equal,
\[ \det(A^T) = \det A. \]

D11. In general,
\[ \det(A + B) \neq \det A + \det B. \]

D12. The determinant of a product equals the product of their determinants,
\[ \det(AB) = (\det A)(\det B). \quad (11) \]

These properties are not independent of each other. For example, D5 follows from D1 and D4, and D4 follows from D6. Keep in mind that $\det()$ is not linear. That is, if $\alpha, \beta$ are scalars and $A, B$ are $n \times n$ matrices, then
\[ \det(\alpha A + \beta B) \neq \alpha \det A + \beta \det B, \quad (12) \]
in general. For instance, if $\beta = 0$ is $\det(\alpha A) = \alpha \det A$? No, according to D7 it is $\alpha^0 \det A$. Or, with $\alpha = \beta = 1$, is $\det(A + B) = \det A + \det B$? Not in general, according to D11. This result may come as a surprise since we are studying “linear algebra.” Also surprising is the truth of D12, if we contrast the complexity of the matrix multiplication $AB$ on the left with the simplicity of the outcome, expressed on the right-hand side. This result was proved by Cauchy in 1815.*

Closure. Every $n \times n$ matrix $A$ has a value associated with it called its determinant and denoted as $\det A$. Though $\det A$ is defined, traditionally, by the cofactor expansion (2), we find that that formula is useless, computationally, unless $n$ is quite small. Thus, we study various properties of the determinant and put forward a computational algorithm called triangularization, based upon properties D1–D3, that is incredibly efficient compared to the cofactor expansion.

Computer software. Using Maple, one can evaluate determinants using the $\text{det}(A)$ command. For instance, to evaluate the determinant of the matrix $A$ given by (6), enter
\[
\text{with(linalg):}
\]
to access the $\text{det}(A)$ command. Then enter
\[
\text{det}([[0, 2, -1], [4, 3, 5], [2, 0, -4]]);
\]
and return. The output is 58. Alternatively, the sequence
\[
\text{with(linalg):}
\]
\[
A := \text{array}([[0, 2, -1], [4, 3, 5], [2, 0, -4]]);
\]
\[
\text{det}(A);
\]

*Augustin-Louis Cauchy (1789–1857) is among the great mathematicians. Unlike his contemporary, Gauss who published little of his work, Cauchy published more than 700 papers. Among the subjects on which he worked were determinants, ordinary and partial differential equations, complex variable theory, and the wave theory of light.
gives the same result.

EXERCISES 10.4

1. In (5) we evaluated the determinant of a general 2 × 2 matrix using a cofactor expansion about the first row. Evaluate it again, using a cofactor expansion about the second row instead, then about the first column, and then about the second column, showing that the answer is the same in each case.

2. Evaluate each, using a cofactor expansion about the first and last rows, and also about the last column.

\[
\begin{align*}
(a) & \quad \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix} & (b) & \begin{bmatrix} 2 & -3 & 0 \\ 1 & 4 & 2 \\ -6 & 1 & 5 \end{bmatrix} \\
(c) & \quad \begin{bmatrix} -4 & 1 & 0 \\ 3 & 2 & 0 \\ 1 & 5 & 7 \end{bmatrix} & (d) & \begin{bmatrix} 3 & 3 & 12 \\ 0 & 6 & -1 \\ 4 & 0 & 0 \end{bmatrix} \\
(e) & \quad \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix} & (f) & \begin{bmatrix} -5 & 2 & 1 \\ 4 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \\
(g) & \quad \begin{bmatrix} 2 & 0 & 1 \\ 0 & 3 & 1 \\ 1 & 2 & 3 \end{bmatrix} & (h) & \begin{bmatrix} 0 & 1 & 2 \\ 3 & -1 & 4 \\ 0 & 2 & 1 \end{bmatrix} \\
(i) & \quad \begin{bmatrix} a & 0 & 0 \\ 0 & b & c \\ 0 & d & e \end{bmatrix} & (j) & \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \\n(0) & \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} & (1) & \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\
\end{align*}
\]

3. (a)–(j) Same as Exercise 2, but expanding about the second row, and about the first column.

4. (a)–(h) Same as Exercise 2, but using the method of triangularization.

5. (a)–(h) Same as Exercise 2, but using computer software.

6. Evaluate, by any means other than computer software, showing your steps or logic. You may use any of the properties D1–D12.

\[
\begin{align*}
(a) & \quad \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \\ 0 & 1 & -3 & 5 \end{bmatrix} & (b) & \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 0 \end{bmatrix} \\
\end{align*}
\]

(c) \begin{bmatrix} 0 & 0 & a \\ 0 & b & c \\ d & e & f \end{bmatrix} \quad \text{and} \quad (d) \begin{bmatrix} a & b & c \\ d & e & 0 \\ f & 0 & 0 \end{bmatrix}

7. A mnemonic device often put forward for evaluating 2 × 2 and 3 × 3 determinants is as shown below.

\[
\text{In other words, the determinants are the sums of the indicated products, with each product carrying the indicated sign. For example, the 2 × 2 case this device gives}
\]

\[
\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = + (a_{11} a_{22}) - (a_{21} a_{12}),
\]

which does agree with (5). We now state the problem: write out the mnemonic result for the 3 × 3 case, and verify (by cofactor expansion) that it is correct. CAUTION: This device does not hold, in general, for n × n determinants if n ≥ 4.

8. Let an n × n matrix \( A = \{a_{ij}\} \) be diagonal. Show that

\[
\det A = a_{11} a_{22} \cdots a_{nn}. \quad (8.1)
\]

9. (a) Suppose that an n × n matrix \( A \) can be partitioned into the block-diagonal form

\[
A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & A_m \end{bmatrix},
\]

where \( A_1, \ldots, A_m \) are all square, although not necessarily all of the same order. Show that

\[
\det A = (\det A_1)(\det A_2) \cdots (\det A_m). \quad (9.1)
\]

This result may be regarded as a generalization of (8.1), above, wherein \( A_1, \ldots, A_m \) were all 1 × 1's.

(b) Does (9.1) still hold if the elements above the m blocks are
nonzero? Explain.
(c) Does (9.1) still hold if the elements below the \( m \) blocks are nonzero? Explain.
(d) To which determinants, in Exercise 2, can (9.1) be applied? In each of those cases use (9.1) to evaluate \( \det \mathbf{A} \).

10. Deduce, from property D12, that if \( \mathbf{A}_1, \ldots, \mathbf{A}_k \) are \( n \times n \) matrices, then

\[
\det(\mathbf{A}_1 \mathbf{A}_2 \cdots \mathbf{A}_k) = (\det \mathbf{A}_1)(\det \mathbf{A}_2) \cdots (\det \mathbf{A}_k). \tag{10.1}
\]

11. (a) Derive the property D7. HINT: Write out the cofactor expansion about the row or column in question.
(b) Then, show that D8 follows from D7.

12. Prove property D6, using any of the other listed properties.

13. Prove property D9. HINT: Write out the cofactor expansion about the row (or column) \( \mathbf{a} \).

14. (Routh–Hurwitz criterion) First, review Section 3.4.5, on stability. According to the Routh–Hurwitz criterion, necessary and sufficient conditions for the stability of the system governed by equation (65) in Section 3.4.5 (i.e., for all the roots of its characteristic equation to have negative real parts) are that \( a_j > 0 \) for each \( j = 1, \ldots, n \), and that \( \Delta_j > 0 \) for each \( j = 1, \ldots, n \), where

\[
\Delta_j = \begin{vmatrix}
1 & 0 & 0 & \cdots & 0 \\
0 & a_2 & a_1 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & a_{2j-1} & a_{2j-2} & a_{2j-3} & a_{2j-4} & \cdots & a_j
\end{vmatrix},
\]

Zeros are entered for any \( a_k \)'s that are called for where \( k > n \). For example, if \( n = 3 \), then

\[
\Delta_1 = a_1, \quad \Delta_2 = \begin{vmatrix}
a_1 & 1 & 0 \\
a_3 & a_2 & a_1 \\
0 & 0 & a_3
\end{vmatrix}, \quad \Delta_3 = \begin{vmatrix}
a_1 & 1 & 0 \\
a_3 & a_2 & a_1 \\
0 & 0 & a_3
\end{vmatrix};
\]

expanding \( \Delta_3 \) about the third row yields the simplification \( \Delta_3 = a_3 \Delta_2 \). Here is the problem: Use the Routh–Hurwitz criterion to determine whether or not the systems associated with the following characteristic equations are stable.

(a) \( \lambda^4 + 6 \lambda^3 + 5 \lambda^2 + 4 \lambda + 1 = 0 \)
(b) \( \lambda^4 + 2 \lambda^3 + 7 \lambda^2 + 4 \lambda + 8 = 0 \)
(c) \( \lambda^4 + 2 \lambda^3 + 5 \lambda^2 + 8 \lambda + 12 = 0 \)
(d) \( \lambda^4 + \lambda^3 + 4 \lambda + 8 = 0 \)
(e) \( \lambda^5 + \lambda^4 + \lambda^3 + \lambda^2 + \lambda + 1 = 0 \)
(f) \( \lambda^5 + \lambda^4 + \lambda^3 + \lambda^2 + \lambda + 8 = 0 \)
(g) \( \lambda^5 + 2 \lambda^4 + 3 \lambda^3 + 4 \lambda^2 + 5 \lambda + 6 = 0 \)

15. It can be shown that the equations

\[
\begin{align*}
a_0 x^2 + a_1 x + a_2 &= 0, \\
b_0 x^2 + b_1 x + b_2 &= 0
\end{align*}
\]

have a common root if and only if

\[
\begin{vmatrix}
a_0 & a_1 & a_2 \\
b_0 & b_1 & b_2 \\
0 & 0 & 0
\end{vmatrix} = 0.
\]

[Similar results for two algebraic equations of degrees \( m \) and \( n \), say, were put forward by James Joseph Sylvester (1814–1897).] Use this result to determine whether or not the following equation pairs have any common roots.

(a) \( 3 x^2 + 2 x - 5 = 0 \)
(b) \( 3 x^2 + 2 x - 5 = 0 \)

16. (a) Suppose that the elements \( a_{ij} \) of an \( n \times n \) matrix \( \mathbf{A} \) are differentiable functions of some parameter \( t \). Regarding \( \det \mathbf{A} \) as a function of the \( n^2 \) variables \( a_{11}, a_{12}, \ldots, a_{nn} \), show that

\[
\frac{\partial}{\partial t} (\det \mathbf{A}) = A_{ij},
\]

where \( A_{ij} \) is the cofactor of \( a_{ij} \). Then use (16.1) and chain differentiation to show that

\[
\frac{d}{dt} (\det \mathbf{A}) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} \frac{da_{ij}}{dt},
\]

a formula first given by Carl Gustav Jacob Jacobi (1804–1851) in 1841. By the \( \sum \sum \) notation we mean

\[
\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} \equiv \sum_{i=1}^{n} \left( \sum_{j=1}^{n} c_{ij} \right).
\]

For example, if \( n = 2 \), then

\[
\sum_{i=1}^{2} \sum_{j=1}^{2} c_{ij} = \sum_{i=1}^{2} \left( \sum_{j=1}^{2} c_{ij} \right) = \sum_{i=1}^{2} (c_{11} + c_{12}) = c_{11} + c_{21} + c_{12} + c_{22}.
\]

Observe that (16.2) is equivalent to the statement

\[
\frac{d}{dt} (\det \mathbf{A}) = \begin{vmatrix}
da_{11} & \cdots & da_{1n} \\
\vdots & \ddots & \vdots \\
da_{n1} & \cdots & da_{nn}
\end{vmatrix},
\]
10.5. Rank; Application to Linear Dependence and to Existence and Uniqueness for $Ax = c$

With determinants defined, we can now introduce one more concept, the “rank” of a matrix, which concept will enable us to obtain important results regarding linear dependence, and also regarding the existence and uniqueness of solutions of the linear equation $Ax = c$.

10.5.1. Rank. First, we say that any matrix obtained from a given $m \times n$ matrix $A$ by deleting at most $m - 1$ rows and at most $n - 1$ columns from $A$
is a submatrix of $A$. For instance, the $2 \times 3$ matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

has 21 submatrices: one $2 \times 3$ ($A$ itself), three $2 \times 2$'s, two $1 \times 3$'s, three $2 \times 1$'s, six $1 \times 2$'s, and six $1 \times 1$'s. For instance,

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}, \quad \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \begin{bmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{bmatrix}, \quad \begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix}, \quad \begin{bmatrix} a_{21} \\ a_{22} \end{bmatrix},$$

and $[a_{13}]$ are all submatrices of $A$.

Then the rank of a matrix is defined as follows.

**DEFINITION 10.5.1 Rank**

A matrix $A$, not necessarily square, is of rank $r$, or $r(A)$, if it contains at least one $r \times r$ submatrix with nonzero determinant but no square submatrix larger than $r \times r$ with nonzero determinant. A matrix is of rank 0 if it is a zero matrix.

**EXAMPLE 1.** Let

$$A = \begin{bmatrix} 2 & -1 & 1 & 0 \\ 0 & 3 & 3 & 6 \\ 1 & 4 & 5 & 9 \end{bmatrix}.$$  \hfill (1)

Certainly, $r$ is at most 3 in this case since the largest possible square submatrix of $A$ is $3 \times 3$. (More generally, if $A$ is $m \times n$, then $r$ is at most equal to the smaller of $m$ and $n$.) However, upon calculation, we find that all four of the $3 \times 3$ submatrices have zero determinant so that $r$ is at most 2. In fact, there are a number of $2 \times 2$ submatrices with nonzero determinant such as

$$\begin{vmatrix} 2 & -1 \\ 0 & 3 \end{vmatrix} = 6 \neq 0,$$

but even if there were only one such submatrix that would still be all we need to conclude that $r(A) = 2$. \hfill \square

**EXAMPLE 2.** The ranks of

$$A = \begin{bmatrix} 5 \\ 6 \\ 0 \end{bmatrix}, \quad B = \begin{bmatrix} 5 & -2 \\ 6 & 3 \\ 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 5 & -2 & 0 \\ 6 & 3 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} 3 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 9 & 0 & 3 & 0 \end{bmatrix}$$

are 1, 2, 2, and 2, respectively. In $D$, for example, every $3 \times 3$ submatrix contains a column of zeros and hence has a vanishing determinant, but the $2 \times 2$ submatrix $\begin{bmatrix} 3 & 1 \\ 0 & 1 \end{bmatrix}$ has a
We may regard the rows of an \( m \times n \) matrix \( A = \{a_{ij}\} \) as \( n \)-dimensional vectors, which we call the row vectors of \( A \) and which we denote as \( r_1, \ldots, r_m \). Similarly, the columns are \( m \)-dimensional vectors, which we call the column vectors of \( A \) and which we denote as \( c_1, \ldots, c_n \). Further, we define the vector spaces \( \text{span} \{r_1, \ldots, r_m\} \) and \( \text{span} \{c_1, \ldots, c_n\} \) as the row and column spaces of \( A \), respectively. From the definition of dimension, the dimensions of the row and column spaces are equal to the number of \( \text{LI} \) row vectors and the number of \( \text{LI} \) column vectors, respectively.

It will be important to be able to calculate the rank of a given matrix efficiently. With that purpose in mind, we recall the elementary row operations defined in Section 8.3:

1. Addition of a multiple of one row to another
   Symbolically: \( r_j \rightarrow r_j + \alpha r_k \)
2. Multiplication of a row by a nonzero constant
   Symbolically: \( r_j \rightarrow \alpha r_j \)
3. Interchange of two rows
   Symbolically: \( r_j \leftrightarrow r_k \)

Furthermore, we defined (in Section 8.3) two matrices to be row equivalent if one can be obtained from the other by finitely many elementary row operations. The following theorem provides an efficient means of calculating the rank of a matrix.

**THEOREM 10.5.1** Elementary Row Operations and Rank

Row equivalent matrices have the same rank. That is, elementary row operations do not alter the rank of a matrix.

**Proof:** If matrices \( A \) and \( B \) are row equivalent, then \( B \) can be obtained from \( A \) by a finite number of elementary row operations. It follows that each row vector of \( B \) must be a linear combination of the row vectors of \( A \) so the row space of \( B \) must be a subspace of the row space of \( A \). Similarly, the row space of \( A \) must be a subspace of the row space of \( B \). Thus, the row space of \( A \) is identical to the row space of \( B \), and hence the dimension of the row space of \( A \) (which is the rank of \( B \)) must equal the dimension of the row space of \( B \) (which is the rank of \( B \)). ■

**EXAMPLE 3.** Let

\[
A = \begin{bmatrix} 2 & 1 & -3 & 4 \\ 2 & 4 & -2 & 5 \\ 0 & 3 & 1 & 3 \\ 2 & 1 & -3 & -2 \end{bmatrix}
\]
Using elementary row operations,
\[
A \rightarrow \begin{bmatrix} 2 & 1 & -3 & 4 \\ 0 & 3 & 1 & 1 \\ 0 & 3 & 1 & 3 \\ 0 & 0 & 0 & -6 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 1 & -3 & 4 \\ 0 & 3 & 1 & 1 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & -6 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 1 & -3 & 4 \\ 0 & 3 & 1 & 1 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix},
\]
where the operations were as follows: in the first step \( r_2 \rightarrow r_2 + (-1)r_1 \) and \( r_4 \rightarrow r_4 + (-1)r_1 \); in the second step \( r_3 \rightarrow r_3 + (-1)r_2 \); and in the final step \( r_4 \rightarrow r_4 + 3r_3 \).
Clearly, the rank of the final matrix is 3 because (deleting the fourth row and third column)
\[
\begin{bmatrix} 2 & 1 & 4 \\ 0 & 3 & 1 \\ 0 & 0 & 2 \end{bmatrix} = 12 \neq 0.
\]
Thus, by Theorem 10.5.1, \( r(A) = 3 \). \( \blacksquare \)

The idea, then, is to reduce a given matrix \( A \) to row echelon form by means of elementary row operations.* It can be seen that, in that form, the nonzero rows are LI. In Example 3, for instance, several conclusions follow from (3): \( r(A) = 3 \), the number of LI vectors among the rows of \( A \) is 3; the dimension of the row space of \( A \) is 3 and a basis for that row space is given by the vectors \([2, 1, -3, 4],[0, 3, 1, 1],[0, 0, 0, 2] \).

There is a connection between the rank of a matrix and the linear dependence of a set of vectors, studied in Chapter 9:

THEOREM 10.5.2 Rank and Linear Dependence
For any matrix \( A \), the number of LI row vectors is equal to the number of LI column vectors and these, in turn, equal the rank of \( A \).\(^1\)

Thus, if we wish to determine how many vectors in a given vector set \( \{u_1, \ldots, u_k\} \) are LI we can form a matrix \( A \) with \( u_1, \ldots, u_k \) as the rows (or columns) and then use elementary row operations to determine the rank of \( A \).

EXAMPLE 4. How many LI vectors are contained in \( \{u_1, u_2, u_3, u_4\} \), where
\[
\begin{align*}
u_1 &= \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix}, & u_2 &= \begin{bmatrix} 1 \\ 4 \\ 3 \end{bmatrix}, & u_3 &= \begin{bmatrix} -2 \\ -1 \\ -3 \end{bmatrix}, & u_4 &= \begin{bmatrix} 4 \\ 5 \\ 3 \end{bmatrix}.
\end{align*}
\]

\(^*\)If we scale the first three rows in the final matrix in (3) by \( \frac{1}{2}, \frac{1}{3}, \) and \( \frac{1}{2} \), respectively, so as to begin the nonzero rows with leading ones, then we would say that the resulting matrix is in reduced row echelon form.

\(^1\)For proof of the first part of this theorem, see Theorem 3.5.5 in Steven J. Leon's Linear Algebra, 3rd ed. (New York: Macmillan, 1990).
If we construct a matrix having these vectors as columns, then we have the $A$ matrix (2). Using elementary row operations, we saw, in Example 3, that $r(A) = 3$. Hence, there are three LI vectors in $\{u_1, u_2, u_3, u_4\}$ or, put differently, $\dim[\text{span} \{u_1, u_2, u_3, u_4\}] = 3$.

**COMMENT.** The ordering of the columns is immaterial. For instance, we could make $u_1$ the third column, $u_2$ the first, $u_3$ the fourth, and $u_4$ the second because rank depends upon the zeroness or nonzeroness of determinants whereas the interchanging of columns (or rows) merely changes the sign of a determinant.

**EXAMPLE 5. Application to Stoichiometry.** To model the combustion of gasoline in an automobile engine, one can begin by writing down a list of well over 100 simultaneous chemical reactions involving the various hydrocarbons, oxygen, nitrogen, and so on. In turn, these reactions can be modeled by ODE’s governing the amount of each chemical species as a function of time, and one can solve the resulting system of ODE’s by methods such as those discussed in Chapter 6. It is easy to appreciate that solving around 100 coupled ODE’s is a difficult undertaking. Thus, it is important to reduce the list of reactions insofar as possible, and we can do this by eliminating ones that are redundant. For instance, if $A + B \rightarrow C$ and $A + C \rightarrow D$, then a third statement, $2A + B + C \rightarrow C + D$, is redundant in that it is implied by the first two.

To illustrate the reduction process, consider the burning of a mixture of CO, $H_2$, and $CH_4$ in a furnace, producing CO, $CO_2$, and $H_2O$. Writing all possible reactions that we can think of gives the list

\[
\begin{align*}
CO + \frac{1}{2}O_2 & \rightarrow CO_2, \quad (4a) \\
H_2 + \frac{1}{2}O_2 & \rightarrow H_2O, \quad (4b) \\
CH_4 + \frac{3}{2}O_2 & \rightarrow CO + 2H_2O, \quad (4c) \\
CH_4 + 2O_2 & \rightarrow CO_2 + 2H_2O, \quad (4d)
\end{align*}
\]

where (4c) and (4d) represent the partial and complete combustion of $CH_4$, respectively. How many of these reactions are independent? It is convenient to re-express them symbolically in the equation format

\[
\begin{align*}
CO + \frac{1}{2}O_2 - CO_2 &= 0, \\
H_2 + \frac{1}{2}O_2 - H_2O &= 0, \\
CH_4 + \frac{3}{2}O_2 - CO - 2H_2O &= 0, \\
CH_4 + 2O_2 - CO_2 - 2H_2O &= 0,
\end{align*}
\]

where the elements of the coefficient matrix

---

*This example is discussed by Ben Noble in his book *Applications of Undergraduate Mathematics in Engineering* (New York: Macmillan, 1967). In turn, he notes that the problem was contributed by John Mahoney, Department of Chemical Engineering, West Virginia University, Morgantown, WV.*
are known as stoichiometric coefficients. To determine a minimum set of independent reactions we reduce $A$ by elementary row operations and obtain

$$
A = \begin{bmatrix}
\text{CO}_2 & \text{CO}_2 & \text{H}_2 & \text{H}_2\text{O} & \text{CH}_4 \\
1 & \frac{1}{2} & -1 & 0 & 0 \\
0 & \frac{1}{2} & 0 & 1 & -1 \\
-1 & \frac{3}{2} & 0 & 0 & -2 \\
0 & 2 & -1 & 0 & -2 \\
\end{bmatrix}
$$

the rank of which is three. Thus, there are three independent reactions such as the list

$$
\begin{align*}
\text{CO} + \frac{1}{2}\text{O}_2 - \text{CO}_2 &= 0, \\
\text{O}_2 + 2\text{H}_2 - 2\text{H}_2\text{O} &= 0, \\
-\text{CO}_2 - 4\text{H}_2 + 2\text{H}_2\text{O} + \text{CH}_4 &= 0, \\
\end{align*}
$$

implied by (7). That is, $\text{CO} + \frac{1}{2}\text{O}_2 \to \text{CO}_2$, and so on. 

10.5.2. Application of rank to the system $Ax = c$. In Chapter 8 we use the method of Gauss elimination both to solve systems of linear algebraic equations and to study the questions of the existence and uniqueness of solutions. Having developed vector and matrix concepts now, we can return to the important problem $Ax = c$ and bring these additional concepts to bear. In doing so, it is convenient to have a representative example to refer to.

EXAMPLE 6. Consider the system $Ax = c$ given by

$$
\begin{align*}
x_1 - x_2 + x_3 + 3x_4 + 2x_6 &= 4, \\
x_1 + 3x_3 + 3x_4 - x_5 + 6x_6 &= 3, \\
2x_1 - x_2 + 2x_3 + x_4 - x_5 + 7x_6 &= 9, \\
x_1 + 5x_3 + 8x_4 - x_5 + 7x_6 &= 1. \\
\end{align*}
$$

Carrying out Gauss elimination by applying elementary row operations to the augmented matrix

$$
A|c = \begin{bmatrix}
1 & -1 & 1 & 3 & 0 & 2 & 4 \\
1 & 0 & 3 & 3 & -1 & 6 & 3 \\
2 & -1 & 2 & 1 & -1 & 7 & 9 \\
1 & 0 & 5 & 8 & -1 & 7 & 1 \\
\end{bmatrix}
$$

(10)
10.5. Rank: Application to Linear Dependence and to Existence and Uniqueness for \( \mathbf{Ax} = \mathbf{c} \)  

gives the row echelon result

\[
\begin{bmatrix}
1 & -1 & 1 & 3 & 0 & 2 & 4 \\
0 & 1 & 2 & 0 & -1 & 4 & -1 \\
0 & 0 & 2 & 5 & 0 & 1 & -2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 
\end{bmatrix}
\]  \hspace{1cm} (11)

and hence the three-parameter family of solutions

\[
\begin{align*}
x_0 &= \alpha_1, \\
x_5 &= \alpha_2, \\
x_4 &= \alpha_3, \\
x_3 &= -1 - \frac{1}{3}\alpha_1 - \frac{5}{2}\alpha_3, \\
x_2 &= 1 - 3\alpha_1 + 5\alpha_2 + 5\alpha_3, \\
x_1 &= 6 - \frac{5}{2}\alpha_2 + \frac{9}{2}\alpha_3, \\
\end{align*}
\]  \hspace{1cm} (12)

where the parameters \( \alpha_1, \alpha_2, \alpha_3 \) are arbitrary.

It is illuminating to express (12) in vector form as

\[
\mathbf{x} = \begin{bmatrix}
6 - \frac{9}{2}\alpha_1 + \frac{9}{2}\alpha_2 + \frac{9}{2}\alpha_3 \\
-1 - \frac{1}{3}\alpha_1 - \frac{5}{2}\alpha_3 \\
\alpha_2 \\
\alpha_1 
\end{bmatrix} = \begin{bmatrix}
6 \\
-1 \\
0 \\
0 
\end{bmatrix} + \alpha_1 \begin{bmatrix}
-\frac{9}{2} \\
-\frac{1}{3} \\
0 \\
1 
\end{bmatrix} + \alpha_2 \begin{bmatrix}
1 \\
0 \\
0 \\
0 
\end{bmatrix} + \alpha_3 \begin{bmatrix}
\frac{9}{2} \\
1 \\
0 \\
0 
\end{bmatrix}
\]

\hspace{1cm} (13)

Observe that \( \mathbf{x}_0 \) is a particular solution of \( \mathbf{Ax} = \mathbf{c} \) (i.e., \( \mathbf{Ax}_0 = \mathbf{c} \)), and \( \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \) are homogeneous solutions (i.e., \( \mathbf{Ax}_1 = \mathbf{0}, \mathbf{Ax}_2 = \mathbf{0}, \mathbf{Ax}_3 = \mathbf{0} \)), by the following reasoning. Since (13) is a solution of \( \mathbf{Ax} = \mathbf{c} \) for any \( \alpha \)'s, we can set \( \alpha_1 = \alpha_2 = \alpha_3 = 0 \) and conclude that \( \mathbf{Ax}_0 = \mathbf{c} \). Next, put (13) into \( \mathbf{Ax} = \mathbf{c} \):

\[
\mathbf{A}(\mathbf{x}_0 + \alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \alpha_3 \mathbf{x}_3) = \mathbf{c},
\]  \hspace{1cm} (14)

hence, \( \mathbf{Ax}_0 + \alpha_1 \mathbf{Ax}_1 + \alpha_2 \mathbf{Ax}_2 + \alpha_3 \mathbf{Ax}_3 = \mathbf{c} \) or, since \( \mathbf{Ax}_0 = \mathbf{c} \),

\[
\alpha_1 \mathbf{Ax}_1 + \alpha_2 \mathbf{Ax}_2 + \alpha_3 \mathbf{Ax}_3 = \mathbf{0}.
\]  \hspace{1cm} (15)

The choice \( \alpha_1 = 1, \alpha_2 = \alpha_3 = 0 \) reveals that \( \mathbf{Ax}_1 = \mathbf{0} \); \( \alpha_2 = 1, \alpha_1 = \alpha_3 = 0 \) reveals that \( \mathbf{Ax}_2 = \mathbf{0} \); and \( \alpha_3 = 1, \alpha_1 = \alpha_2 = 0 \) reveals that \( \mathbf{Ax}_3 = \mathbf{0} \), as claimed.

Observe further that \( \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \) are L.I. for the rank of

\[
[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3] = \begin{bmatrix}
-\frac{9}{2} & 1 & \frac{9}{2} \\
-3 & 1 & 5 \\
-\frac{1}{2} & 0 & -\frac{5}{2} \\
0 & 0 & 1 \\
0 & 1 & 0 \\
1 & 0 & 0
\end{bmatrix}
\]  \hspace{1cm} (16)
is 3, as is readily seen from the bottom three rows. 

Generalizing the results of Example 6, suppose that a system

\[ Ax = c, \]

where \( A \) is \( m \times n \), has a \( p \)-parameter family of solutions

\[ x = x_0 + \alpha_1 x_1 + \cdots + \alpha_p x_p. \]

Then \( x_0 \) is necessarily a particular solution, and \( x_1, \ldots, x_p \) are necessarily LI homogeneous solutions. We call \( \text{span} \{ x_1, \ldots, x_p \} \) the \textbf{solution space} of the homogeneous equation \( Ax = 0 \), or the \textbf{null space} of \( A \). The dimension of that null space is called the \textbf{nullity} of \( A \).

It is helpful to see that if \( \{ c_1, \ldots, c_n \} \) denote the columns of \( A \), then \( Ax = c \) can be expressed as

\[ x_1 c_1 + x_2 c_2 + \cdots + x_n c_n = c, \]

from which we can see that (17) is consistent if and only if \( c \) happens to be in the column space of \( A \) [namely, \( \text{span} \{ c_1, \ldots, c_n \} \) ]. Or, in terms of rank, (17) is consistent if and only if the rank of the augmented matrix \( A|c \) equals the rank of the coefficient matrix \( A \) : \( r(A|c) = r(A) \).

In Example 6, for instance, we see from (11) that \( r(A|c) = 3 \) and (by covering up the last column, which is \( c \)) that \( r(A) = 3 \) as well. Thus, \( r(A|c) = r(A) \) and, sure enough, (9) is consistent, solutions being given by (12). However, if we modify (9) by changing the underlined 1 to a 2, say, then the underlined 0 in (11) becomes a 1. In that case \( r(A|c) = 4 \) and \( r(A) = 3 \) are unequal and there is no solution because the bottom row of (11) would then be equivalent to \( 0x_1 + \cdots + 0x_6 = 1 \), which cannot be satisfied by any combination of \( x_j \)'s.

Finally, what can be said about \( p \) in (18)? In Example 6, we see from (11) that \( p = 3 \) for there are three arbitrary \( x_j \) values (as seen from the third row), and that that value arises as the difference between the number of unknowns \( n = 6 \) and the rank \( r = 3 \).

Let us summarize, for any system \( Ax = c \).

\begin{thm}
Existence and Uniqueness for \( Ax = c \)
\end{thm}

Consider the linear system

\[ Ax = c, \]

where \( A \) is \( m \times n \). There is

1. no solution if and only if \( r(A|c) \neq r(A) \),
2. a unique solution if and only if \( r(A|c) = r(A) = n \),
3. an \( (n - r) \)-parameter family of solutions if and only if \( r(A|c) = r(A) \equiv r \) is less than \( n \).
The essential ideas required to prove these three results were developed above, so the proofs are left for the exercises.

Naturally, the homogeneous system

\[ \mathbf{Ax} = \mathbf{0} \]  \hspace{1cm} (21)

is but a special case of (20), hence, it is already covered by Theorem 10.5.3. However, it is such an important case that it deserves special attention. In (21), the augmented matrix \( r(\mathbf{A}|\mathbf{c}) \) is the \( \mathbf{A} \) matrix augmented by a column of zeros so it is surely true that \( r(\mathbf{A}|\mathbf{c}) = r(\mathbf{A}) \) and, according to Theorem 10.5.3, it must be true that (21) is consistent. That result is no great surprise since (21) always admits the trivial solution \( \mathbf{x} = \mathbf{0} \). Hence, the significant question about (21) is not whether or not it is consistent, but whether \( \mathbf{x} = \mathbf{0} \) is the only solution. That is, does (21) admit nontrivial solutions as well? That question is answered by parts 1 and 2 of Theorem 10.5.3 so we can state the following more specialized results.

**THEOREM 10.5.4 Homogeneous Case Where \( \mathbf{A} \) is \( m \times n \)**

If \( \mathbf{A} \) is \( m \times n \), then

\[ \mathbf{Ax} = \mathbf{0} \]  \hspace{1cm} (22)

1. is consistent,
2. admits the trivial solution \( \mathbf{x} = \mathbf{0} \),
3. admits the unique solution \( \mathbf{x} = \mathbf{0} \) if and only if \( r(\mathbf{A}) = n \),
4. admits an \( (n - r) \)-parameter family of nontrivial solutions, in addition to the trivial solution, if and only if \( r(\mathbf{A}) \equiv r < n \).

**THEOREM 10.5.5 Homogeneous Case Where \( \mathbf{A} \) is \( n \times n \)**

If \( \mathbf{A} \) is \( n \times n \), then

\[ \mathbf{Ax} = \mathbf{0} \]  \hspace{1cm} (23)

admits nontrivial solutions, besides the trivial solution \( \mathbf{x} = \mathbf{0} \), if and only if \( \det \mathbf{A} = 0 \).

As a final example, consider an interesting application of these concepts to “dimensional analysis.”

**EXAMPLE 7. Dimensional Analysis.** Consider a rectangular flat plate (i.e., a flat rectangle) in steady motion through otherwise-undisturbed air as shown in Fig. 1: \( V \) is the flight speed, \( \theta \) is the incidence or angle of attack of the airfoil, \( A \) is the chord length, and \( B \) is the span (the dimension normal to the paper). Equivalently, it is experimentally more convenient to keep the airfoil fixed (in a wind tunnel) and to blow air
past it, at a speed $V$. Imagine that our object is to conduct an experimental determination of the lift force $\ell$ generated on the airfoil, that is, to experimentally determine the functional dependence of $\ell$ on the various relevant quantities. What quantities are relevant? Surely $A$, $B$, $\theta$, $V$ are important, as well as the air density $\rho$ (for instance we expect a much greater lift in water than in air, and no lift at all in a vacuum). A reasonable list of the relevant variables is given in Table 1. Other variables come to mind, such as the ambient temperature,

<table>
<thead>
<tr>
<th>Table 1. Relevant variables.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>--------------------------</td>
</tr>
<tr>
<td>Chord</td>
</tr>
<tr>
<td>Span</td>
</tr>
<tr>
<td>Incidence</td>
</tr>
<tr>
<td>Flight velocity</td>
</tr>
<tr>
<td>Velocity of sound in air</td>
</tr>
<tr>
<td>Air density</td>
</tr>
<tr>
<td>Absolute viscosity</td>
</tr>
<tr>
<td>Lift</td>
</tr>
</tbody>
</table>

but if we expect $\ell$ to be only weakly dependent on them we can leave them out.

Major difficulties are now apparent. If $\ell$ depends upon the seven variables listed in Table 1, and we measure $\ell$ for five different values of each variable, then we need to conduct $5^7 = 78,125$ experimental runs, and then present the results (using graphs, tables, or whatever) in a user-friendly way. Furthermore, whereas some variables are easily varied (such as $A$, $B$, $\theta$, $V$) others are not (such as $\rho$, $\mu$). The principal object of the following "dimensional analysis" is to reduce the number of variables as much as possible.

To begin, we express each variable in terms of the fundamental units $M$ (mass), $L$ (length), and $T$ (time), in the right-hand column. We do not need to include $F$ (force) as a fundamental unit because, according to Newton's second law, $F = MLT^{-2}$, dimensionally speaking. Also, observe that $\theta$ is dimensionless.\textsuperscript{a}

Next, we seek all possible dimensionless products of the form

$$A^a B^b \theta^c V^d V_0^e \rho^f \mu^g \ell^h.$$  \hspace{1cm} (24)

That is, we seek the exponents $a, \ldots, h$ such that

$$L^a (L^b (M^0L^0T^0)^c (LT^{-1})^d (LT^{-1})^e (ML^{-3})^f (ML^{-1}T^{-1})^g (MLT^{-2})^h = M^0L^0T^0.$$  \hspace{1cm} (25)

Equating exponents of $L$, $T$, $M$ on both sides, we see that $a, \ldots, h$ must satisfy the hom-\textsuperscript{a} Recall that angle is defined by the formula $s = r\theta$, where $s$ is the arc length of a circular arc of radius $r$, subtended by an angle $\theta$, measured in radians. Thus, $\theta = s/r = \text{length}/\text{length} = \text{dimensionless}$.
geneous linear system

\[
\begin{align*}
  a + b + d + e - 3f - g + h &= 0, \\
  -d - e - g + 2h &= 0, \\
  f + g + h &= 0.
\end{align*}
\]  

(26)

Solving (26) by Gauss elimination gives the five-parameter family of solutions

\[
\begin{bmatrix}
  a \\ b \\ c \\ d \\ e \\ f \\ g \\ h
\end{bmatrix} = \alpha_1 \begin{bmatrix}
  -2 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1
\end{bmatrix} + \alpha_2 \begin{bmatrix}
  -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0
\end{bmatrix} + \alpha_3 \begin{bmatrix}
  -1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0
\end{bmatrix} + \alpha_4 \begin{bmatrix}
  0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0
\end{bmatrix} + \alpha_5 \begin{bmatrix}
  0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0
\end{bmatrix},
\]  

(27)

where \( \alpha_1, \ldots, \alpha_5 \) are arbitrary constants. With \( \alpha_1 = 1 \) and \( \alpha_2 = \cdots = \alpha_5 = 0 \), say, (27) gives \( a = -2, b = c = 0, \) etc., and hence the nondimensional parameter \( A^{-2}B^{0.7}V^{-2}V_0^{-1} \mu^{1.5} \), namely, the nondimensional lift \( \ell/(\rho V^2 A^2) \). Similarly, setting \( \alpha_2 = 1 \) and the other \( \alpha_j \)'s = 0 gives the nondimensional parameter \( \rho AV/\mu \), well known in fluid mechanics as the Reynolds number and denoted as \( \text{Re} \); setting \( \alpha_3 = -1 \) and the other \( \alpha_j \)'s = 0 gives the Mach number \( V/V_0 \), denoted as \( \mathcal{M} \); setting \( \alpha_4 = 1 \) and the other \( \alpha_j \)'s = 0 gives the incidence \( \theta \) (which was nondimensional to begin with); and setting \( \alpha_5 = 1 \) and the other \( \alpha_j \)'s = 0 gives \( B/A \), known as the aspect ratio, denoted typically as \( \mathcal{R} \).

The upshot is that rather than seek a functional relationship on the eight variables listed in the table, we can seek a relationship on the five nondimensional variables \( [\ell/(\rho V^2 A^2), \text{Re}, \mathcal{M}, \theta, \mathcal{R}] \). Or, singling out the nondimensional lift, we can express

\[
\frac{\ell}{\rho V^2 A^2} = f(\text{Re}, \mathcal{M}, \theta, \mathcal{R})
\]  

(28)

and determine \( f \) experimentally by measuring \( \ell/(\rho V^2 A^2) \) for various combinations of \( \text{Re}, \mathcal{M}, \theta, \) and \( \mathcal{R} \) values.

COMMENT 1. A trained fluid mechanicist could probably simplify the problem even further. For example, it is known (from the governing equations of fluid mechanics) that the effect of the Mach number \( \mathcal{M} \) will be negligible if \( \mathcal{M}^2 \ll 1 \). Thus, if we have flight speeds \( V \) less than 300 miles per hour in mind, then \( \mathcal{M} \) can, to a good approximation, be dropped from (28).\(^1\)

COMMENT 2. We mentioned, above, the practical difficulty in carrying out the experiment for a range of values of the fluid density. For instance, we could use air and water, but

\(^1\)A fluid mechanicist would probably change this to \( \ell/(\rho V^2 AB) \) (corresponding to \( \alpha_1 = 1, \alpha_2 = \alpha_3 = \alpha_4 = 0, \alpha_5 = -1 \)), or to \( \ell/(\frac{1}{2}\rho V^2 AB) \) because \( \frac{1}{2}\rho V^2 \) has physical significance as the stagnation pressure, and \( AB \) is the area of the airfoil.

\(^{1}\)At ground level, the speed of sound is 762 mph, so if \( V = 300 \) mph then \( \mathcal{M} = (300/762)^2 = 0.155 \) is indeed small compared to 1.
their densities are widely different and we would need both wind tunnel and water tunnel facilities. In the right-hand side of (28), however, \( \rho \) shows up only within the Reynolds number \( \text{Re} = \rho AV/\mu \), which can be varied readily by varying the wind speed \( V \).

**COMMENT 3.** Of course, (27) gives an infinite number of nondimensional parameters. However, there are only five independent ones, such as the ones named above. For instance, we could choose \( \alpha_3 = \alpha_5 = 1 \) and \( \alpha_4 = \alpha_4 = 0 \), but the resulting nondimensional parameter, \( V_0 B/(VA) \), is merely the aspect ratio divided by the Mach number.

**Closure.** The rank \( \text{r}(A) \) is defined as the size of the largest nonvanishing determinant within \( A \). Because the rank of a matrix is unaffected by elementary row operations, we can determine the rank of a given matrix efficiently by reducing it to row echelon form, in which form the rank can be seen by inspection. Principal applications of the concept of rank include the calculation of the number of LI vectors (\( n \)-tuple vectors, that is) within a given set, and the theory of the existence and uniqueness of solutions of systems of linear algebraic equations.

**Computer software.** Using Maple, we can evaluate the rank of a given matrix using the `rank(A)` command. For instance, to evaluate \( r(A) \) where the rows of \( A \) are \([1, 2, 3, 4], [2, 4, 6, 8], [1, 1, 1, 1]\), respectively, enter

```maple
with(linalg):
rank(array([[1, 2, 3, 4], [2, 4, 6, 8], [1, 1, 1, 1]]));
```

and return. The output is 2. Alternatively, the sequence

```maple
A := array([[1, 2, 3, 4], [2, 4, 6, 8], [1, 1, 1, 1]]);
rank(A);
```

gives the same result.

**EXERCISES 10.5**

1. Determine the rank, nullity, number of LI rows, and number of LI columns for the given matrix.

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( \begin{bmatrix} 0 &amp; 0 &amp; 2 &amp; 0 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 1 &amp; 2 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 3 &amp; 2 &amp; 1 \end{bmatrix} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \begin{bmatrix} 5 &amp; 7 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 6 &amp; 5 &amp; 2 &amp; 0 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 1 &amp; 3 \end{bmatrix} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \begin{bmatrix} 4 &amp; 9 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 0 &amp; 2 &amp; 3 &amp; 0 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 2 &amp; 6 \end{bmatrix} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \begin{bmatrix} 1 &amp; 2 &amp; 3 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 0 &amp; 0 &amp; 0 &amp; 6 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 3 &amp; 9 \end{bmatrix} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \begin{bmatrix} 4 &amp; 5 &amp; 6 \end{bmatrix} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \begin{bmatrix} 7 &amp; 8 &amp; 9 \end{bmatrix} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2. (a)–(n) Use computer software to determine the rank of the matrix given in the corresponding part of Exercise 1.

3. (a)–(n) Consider the problem $Ax = c$, where $A$ is the $m \times n$ matrix given in the corresponding part of Exercise 1. In each case let $c$ be the $m$-dimensional vector $[1, 1, \ldots, 1]^T$. Use Theorem 10.5.3, and suitable rank calculations, to determine whether or not the system is consistent. If consistent, determine whether it admits a unique solution or a $p$-parameter family of solutions. If the latter, determine $p$. Do not solve the system; merely use the concept of rank and Theorem 10.5.3.

4. If two matrices of the same form have the same rank, need they be row equivalent? Prove or disprove.

5. Show, by carrying out suitable row operations, that the following pairs of matrices are row equivalent.

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

6. Show that the following pairs of matrices are not row equivalent.

\begin{equation}
\begin{bmatrix}
1 & 3 & -3 & 0 \\
2 & 1 & 0 & 4
\end{bmatrix}
\quad \begin{bmatrix}
4 & 2 & 0 & 8 \\
3 & 4 & -3 & 4
\end{bmatrix}
\end{equation}

7. Exercises 5 and 6 are simple enough to be solvable by inspection. More generally, inspection may not suffice. Put forward a systematic procedure for determining whether or not two given matrices are row equivalent, and apply that procedure to the given matrices.

8. Determine whether the following set of vectors is LI or LD by computing the rank of a suitable matrix and invoking the relevant theorem.

\begin{equation}
\begin{bmatrix}
2 & 0 & 1 & -1 \\
0 & 3 & 0 & 3 \\
4 & 3 & 2 & 1
\end{bmatrix}
\end{equation}

9. Prove that $r(A^T) = r(A)$ for every $m \times n$ matrix using any results given in this section.

10. The property $\det(AB) = (\det A)(\det B)$, of determinants, where $A$ and $B$ are both $n \times n$, might seem to imply that $r(AB) = r(A)r(B)$. Is the latter true? Prove or disprove.

11. Is $r(A + B) = r(A) + r(B)$ true? Prove or disprove.

12. Prove that if $A$ is $m \times n$ and $B$ is $n \times p$, then $r(AB) \leq n$. HINT: Partition $B$ into rows, and write

\begin{equation}
AB = \begin{bmatrix}
\alpha_{11} & \cdots & \alpha_{1n} \\
\vdots & \ddots & \vdots \\
\alpha_{m1} & \cdots & \alpha_{mn}
\end{bmatrix}
\begin{bmatrix}
r_1 \\
\vdots \\
r_n
\end{bmatrix}
\end{equation}

13. (a) Below (18), we stated that $x_1, \ldots, x_p$ are necessarily LI. Prove that claim. HINT: Pattern your proof after the discussion in Example 6. (b) Show that $x_0$ cannot be in the span of $x_1, \ldots, x_p$.

14. Although we made a case for the truth of Theorem 10.5.3, we did not provide a detailed proof.

(a) Prove part 1. (b) Prove part 2. (c) Prove part 3.

15. This exercise refers to Example 6 and the discussion following that example. For definiteness, let $A$ be $3 \times 3$.

(a) Suppose that $Ax = c$ admits a one-parameter family of solutions
10.6 Inverse Matrix, Cramer’s Rule, Factorization

There exist important methods for solving a linear system \( Ax = c \) besides Gauss elimination. In this section we study three: the inverse matrix method, Cramer’s rule, and LU factorization.

10.6.1. Inverse matrix. Having introduced matrix notation so that a system of linear algebraic equations can be expressed compactly as

\[
Ax = c,
\]

the form of (1) itself suggests other solution strategies. For how would we solve the simple scalar equation \( 3x = 12 \)? We could divide both sides by 3 and obtain \( x = \frac{12}{3} = 4 \). However, if we try to carry that idea over to the matrix case (1), we obtain \( x = \frac{c}{A} \), and need to know how to divide one matrix into another. However, matrix division has not been (and will not be) defined. Alternatively, we can solve

\[
X = X_0 + \alpha_1 x_1.
\]  

(a) Explain, with the help of a labeled sketch, the geometrical significance of \( x_0, x_1 \) and \( x_0 + \alpha_1 x_1 \).

(b) Suppose that \( Ax = c \) admits a two-parameter family of solutions

\[
x = x_0 + \alpha_1 x_1 + \alpha_2 x_2.
\]  

17. (Dimensional analysis) In studying the drag force on a sphere moving beneath a water surface, the tabulated variables are deemed relevant. Proceeding along the same lines as in Example 7, obtain the following relevant dimensionless parameters: the dimensionless drag force \( D/(\rho V^2 R^2) \), the Reynolds number \( Re = \rho RV/\mu \), the Froude number \( V^2/(gR) \), and the two length ratios \( \lambda/R \) and \( d/R \).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Symbol</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Radius of sphere</td>
<td>( R )</td>
<td>( L )</td>
</tr>
<tr>
<td>Depth below water surface</td>
<td>( d )</td>
<td>( L )</td>
</tr>
<tr>
<td>Velocity of sphere</td>
<td>( V )</td>
<td>( LT^{-1} )</td>
</tr>
<tr>
<td>Water density</td>
<td>( \rho )</td>
<td>( ML^{-3} )</td>
</tr>
<tr>
<td>Absolute viscosity</td>
<td>( \mu )</td>
<td>( ML^{-1}T^{-1} )</td>
</tr>
<tr>
<td>Gravity</td>
<td>( g )</td>
<td>( LT^{-2} )</td>
</tr>
<tr>
<td>Wavelength of free surface waves</td>
<td>( \lambda )</td>
<td>( L )</td>
</tr>
<tr>
<td>Drag force</td>
<td>( D )</td>
<td>( MLT^{-2} )</td>
</tr>
</tbody>
</table>

(c) \( O_2 + CH_3 \rightarrow CH_3OO \),

\( CH_4 + CH_3OO \rightarrow CH_3 + CH_2OHOH \),

\( CH_3OOH \rightarrow CO + 2H_2 + O \),

\( CH_4 + O \rightarrow CH_3 + OH \),

\( CH_4 + O_2 \rightarrow CH_3OOH \)
3x = 12 by multiplying both sides by $\frac{1}{3}$, for that step gives $\frac{1}{3}3x = \frac{1}{3}12$, or $1x = 4$, and hence $x = 4$. That idea does carry over to (1) because matrix multiplication is defined.

The idea, then, is to seek a matrix "$A^{-1}$" having the property that $A^{-1}A = I$ for then

$$A^{-1}Ax = A^{-1}c$$

(2)

becomes

$$Ix = A^{-1}c$$

(3)

and since $Ix = x$, we have the solution

$$x = A^{-1}c$$

(4)

of (1). Note that $A^{-1}$ does not mean $1/A$ or $I/A$; it is simply the name of the matrix having the property

$$A^{-1}A = I,$$

(5)

if one exists. We call it the inverse of $A$, or "$A$-inverse" for brevity.

Consider an exploratory example.

**EXAMPLE 1.** Let $Ax = c$ be the system

\[
\begin{align*}
-x_1 - 2x_2 &= 1 \\
+ x_1 + x_2 &= 1 \\
\end{align*}
\]

or

\[
\begin{bmatrix}
-1 & -2 \\
1 & 1 \\
0 & 1 \\
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
\end{bmatrix}
= \begin{bmatrix}
1 \\
1 \\
1 \\
\end{bmatrix}
\]

(6)

We find, for example by trial and error, that

\[
\begin{bmatrix}
1 & 2 & 0 \\
1 & 1 & 2 \\
\end{bmatrix}
\begin{bmatrix}
-1 & -2 \\
1 & 1 \\
0 & 1 \\
\end{bmatrix}
= \begin{bmatrix}
1 & 0 \\
0 & 1 \\
\end{bmatrix}
= I,
\]

(7)

so that the pre-multiplication of (6) by the matrix

$\begin{bmatrix}
1 & 2 & 0 \\
1 & 1 & 2 \\
\end{bmatrix}$

yields

\[
\begin{bmatrix}
1 & 0 \\
0 & 1 \\
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
\end{bmatrix}
= \begin{bmatrix}
1 & 2 & 0 \\
1 & 1 & 2 \\
\end{bmatrix}
\begin{bmatrix}
1 \\
1 \\
1 \\
\end{bmatrix}
= \begin{bmatrix}
3 \\
4 \\
\end{bmatrix},
\]

(8)

and hence the (tentative) solution $x_1 = 3, x_2 = 4$. Yet the latter does not satisfy (6), so the method is in some way incorrect. [In fact, Gauss elimination reveals that (6) is inconsistent, has no solution.]  

In other words, we must proceed with caution. The idea is that the pre-multiplication of $Ax = c$ by a matrix $A^{-1}$ having the property that $A^{-1}A = I$ does not necessarily lead to an equivalent system.
However, we now show that if $A$ is square, say $n \times n$, and a matrix $A^{-1}$ can be found such that $A^{-1}A = I$, then the pre-multiplication of $Ax = c$ by $A^{-1}$ does lead to an equivalent system, namely, the unique solution $x = A^{-1}c$.

First, assuming that $A$ is $n \times n$, observe from (9) that $A^{-1}A = I$.

Second, it follows from (9) that
\[
\det(A^{-1}A) = \det I = 1
\]
or, since the determinant of a product equals the product of the determinants (property D12 in Section 10.4),
\[
(detA^{-1})(detA) = 1.
\]

Equation (10) cannot possibly be satisfied if $\det A = 0$. Hence, if a matrix $A^{-1}$ satisfying (9) is to exist, it is necessary that $\det A \neq 0$.

Assuming that that is the case, that $\det A \neq 0$, let us seek to determine $A^{-1}$.

Our starting point is the cofactor expansion from Section 10.4,
\[
\det A = \sum a_{jk}A_{jk},
\]
where $A_{jk}$ is the cofactor of the $a_{jk}$ element, and the sum is either on $j$ (for any fixed value of $k$) or on $k$ (for any fixed value of $j$). Let us take the sum to be on $j$.

Observe that
\[
\sum_j a_{jk}A_{ji} = \begin{cases} \det A & \text{if } i = k, \\ 0 & \text{if } i \neq k \end{cases}
\]
(12)
since if $i = k$, then (11) applies, and if $i \neq k$, then the left-hand side of (12) is again a cofactor expansion, this time an expansion about the $i$th column — but with the $i$th column replaced by the $k$th column; thus, it is a determinant containing two identical columns and according to property D5 in Section 10.4, it must therefore be zero. Rearranging (12) by dividing through by $\det A$ (which is permissible since we have assumed that $\det A \neq 0$) and using the Kronecker delta notation,*
\[
\sum_j \left( \frac{A_{ji}}{\det A} \right) a_{jk} = \delta_{ik},
\]
(13)
*Defined in Section 9.10.2, $\delta_{ik}$ is simply 1 if $i = k$ and 0 if $i \neq k$. Thus, a matrix $\{\delta_{ik}\}$ is an identity matrix $I$. 

Chapter 10. Matrices and Linear Equations
This scalar statement (which holds for \(1 \leq i \leq n\) and \(1 \leq j \leq n\)) is equivalent, according to the definition of matrix multiplication,\(^1\) to the matrix equation
\[
A^{-1}A = I,
\]
where the desired inverse matrix \(A^{-1}\) is\(^4\)
\[
A^{-1} = \left\{ \alpha_{ij} \right\} = \left\{ \frac{A_{ji}}{\det A} \right\}.
\]
Or, written out,
\[
A^{-1} = \frac{1}{\det A} \begin{bmatrix}
A_{11} & A_{21} & \cdots & A_{n1} \\
A_{12} & A_{22} & \cdots & A_{n2} \\
\vdots & \vdots & \ddots & \vdots \\
A_{1n} & A_{2n} & \cdots & A_{nn}
\end{bmatrix}.
\]
CAUTION: The \(A_{ji}\) in (15) is not a misprint; the \(j, i\) indices simply turn out to be in the reverse order of those in \(\alpha_{ij}\). For instance, the 1, 2 element of \(A^{-1}\) is \(A_{21}/\det A\), not \(A_{12}/\det A\), where \(A_{21}\) is the cofactor of the 2, 1 element in \(A\), not the 2, 1 element of \(A\) (which is \(a_{21}\)).

The matrix in (16) is called the adjoint of \(A\) and is denoted as \(\text{adj}A\) so
\[
A^{-1} = \frac{1}{\det A} \text{adj}A.
\]
To form the adjoint of a given square matrix \(A\) we replace each element by its cofactor, and take the transpose of the resulting matrix.

The upshot, then, is that if \(\det A \neq 0\), then \(A^{-1}\) exists and is given by (17). In that case we say that \(A\) is invertible. If \(\det A = 0\), then \(A^{-1}\) does not exist, and we say that \(A\) is singular.

EXAMPLE 2. Determine the inverse of
\[
A = \begin{bmatrix}
3 & 2 & -1 \\
0 & 1 & 4 \\
1 & 5 & -2
\end{bmatrix},
\]
if it exists. It does exist because \(\det A = -57 \neq 0\) and is given by (17). Since \(\text{adj}A\) is the transpose of the cofactor matrix, we have
\[
\begin{align*}
\text{adj}A &= \begin{bmatrix}
1 & 4 & -2 \\
5 & -2 & -1 \\
2 & -1 & 3
\end{bmatrix}^T \\
&= \begin{bmatrix}
1 & 0 & 0 \\
5 & -1 & 3 \\
2 & 1 & 0
\end{bmatrix}
\end{align*}
\]

\(^1\)Recall that if \(BC = D\), then \(d_{ij} = \sum \mathbf{b}_{ik}c_{kj}\) or, what is equivalent, \(d_{ik} = \sum \mathbf{b}_{ij}c_{jk}\).

\(^4\)It is tempting to let the \(i, j\) element of \(A^{-1}\) be denoted as \(a_{ij}^{-1}\), but this quantity could be misunderstood to be \(1/\alpha_{ij}\). Thus, let us use \(\alpha_{ij}\).
Chapter 10. Matrices and Linear Equations

\[
\begin{bmatrix}
-22 & 4 & -1 \\
-1 & -5 & -13 \\
9 & -12 & 3
\end{bmatrix}
= \begin{bmatrix}
-22 & -1 & 9 \\
4 & -5 & -12 \\
-1 & -13 & 3
\end{bmatrix}^{T}
\]

so

\[
A^{-1} = \frac{1}{\det A} \text{adj} A = -\frac{1}{57} \begin{bmatrix}
-22 & -1 & 9 \\
4 & -5 & -12 \\
-1 & -13 & 3
\end{bmatrix}
\]

It is readily verified from (20) and (18) that \( A^{-1}A = I \).

Besides having \( A^{-1} \) satisfy \( A^{-1}A = I \), as stated in (14), it is crucial, as we shall see, to have \( AA^{-1} = I \) as well. To show that \( AA^{-1} \) does equal \( I \), write

\[
AA^{-1} = \left\{ \sum_{k} a_{ik} \alpha_{jk} \right\} = \left\{ \sum_{k} a_{ik} A_{jk} \right\} = \frac{1}{\det A} \left\{ \sum_{k} a_{ik} A_{jk} \right\}
\]

where the first equality follows from the definition of matrix multiplication, and the second follows from (15). Now,

\[
\sum_{k} a_{ik} A_{jk} = \begin{cases} 
\det A, & i = j \\
0, & i \neq j
\end{cases}
\]

because if \( i = j \) then (11) applies, and if \( i \neq j \) then the left-hand side of (22) is again a cofactor expansion, this time an expansion about the \( j \)th row — but with the \( j \)th row replaced by the \( i \)th row. Thus, it is a determinant containing two identical rows and, according to property D5, it must therefore be zero. Hence (21) becomes

\[
AA^{-1} = \{ \delta_{ij} \} = I,
\]

as claimed. Thus,

\[
A^{-1}A = AA^{-1} = I.
\]

In view of the first equality in (24), we see that \( A^{-1} \) and \( A \) necessarily commute.

To understand the significance of (24), let us review the solution of \( Ax = c \) by the inverse matrix method. The steps (2) and (3) gave \( x = A^{-1}c \), provided that \( A^{-1}A = I \). To verify that \( x = A^{-1}c \) does indeed satisfy \( Ax = c \), let us put \( A^{-1}c \) into that equation in place of \( x \):

\[
A (A^{-1}c) = c
\]

or, by the associative property of matrix multiplication,

\[
(AA^{-1})c = c
\]
which is indeed true because $AA^{-1} = I$.

Let us pull these results together.

**Theorem 10.6.1 Inverse Matrix**

Let $A$ be $n \times n$. If $\det A \neq 0$, then there exists a unique matrix $A^{-1}$, also $n \times n$, called the inverse of $A$, such that

$$A^{-1}A = AA^{-1} = I. \quad (27)$$

$A$ is then said to be invertible, and its inverse is given by (17). If $\det A = 0$, then a matrix $A^{-1}$ satisfying (27) does not exist, and $A$ is said to be singular.

**Proof:** In the discussion preceding the theorem we proved all but the uniqueness of $A^{-1}$. To prove uniqueness, let $B$ and $C$ both be inverses of $A$. Then $BA = I$ and $CA = I$. Subtracting, $BA - CA = 0$ or $(B - C)A = 0$. And post-multiplying this last equation by $A^{-1}$ (which exists by assumption), we have $(B - C)AA^{-1} = 0A^{-1}$ or $(B - C)I = 0$. Thus, $B - C = 0$, and hence $B = C$. $\blacksquare$

Finally, we return to the application of $A^{-1}$ in the solution of $Ax = c$.

**Theorem 10.6.2 Solution of $Ax = c$**

If $A$ is $n \times n$ and $\det A \neq 0$, then $Ax = c$ admits the unique solution $x = A^{-1}c$.

**Proof:** That $x = A^{-1}c$ satisfies $Ax = c$ was shown just above Theorem 10.6.1. That the solution is unique follows from Theorem 10.5.3 because $\det A \neq 0$ implies that $r(A|c) = r(A) = n$. $\blacksquare$

There are several useful properties of inverse matrices.

---

*Now we can understand that the failure in Example 1 occurred because

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

*There exist interesting generalizations of the notion of the inverse matrix for matrices that are not strictly invertible (perhaps not even square). Such are the Moore–Penrose generalized inverse and the pseudoinverse. See, for example, Gilbert Strang, *Linear Algebra and Its Applications* (New York: Academic Press, 1976), Chap. 3.
PROPERTIES OF INVERSES

11. If $A$ and $B$ are of the same order, and invertible, then $AB$ is too, and

$$\text{(AB)}^{-1} = B^{-1}A^{-1}. \quad (28)$$

12. If $A$ is invertible, then

$$\text{(A}^{T}\text{)}^{-1} = \text{(A}^{-1}\text{)}^{T} \quad (29)$$

and

$$\det \text{(A}^{-1}\text{)} = \frac{1}{\det \text{A}}. \quad (30)$$

13. If $A$ is invertible, then $(A^{-1})^{-1} = A$ and $(A^{m})^{n} = A^{mn}$ for any (positive, negative, or zero) integers $m$ and $n$.

14. If $A$ is invertible, then $AB = AC$ implies that $B = C$. $BA = CA$ implies that $B = C$, $AB = 0$ implies that $B = 0$, and $BA = 0$ implies that $B = 0$.

Of these, let us prove (28) and (29) and leave the remaining proofs as exercises. First (28): Since $A$ and $B$ are invertible, by assumption, $\det A \neq 0$ and $\det B \neq 0$. Thus, $\det(AB) = (\det A)(\det B) \neq 0$ so $AB$ is invertible too. Let us denote $(AB)^{-1}$ as $C$. Then $ABC = I$, $A^{-1}ABC = A^{-1}I$, $BC = A^{-1}B^{-1}BC = B^{-1}A^{-1}$, hence, $C = B^{-1}A^{-1}$, as claimed. As a mnemonic device, note the resemblance of (28) to the transpose formula $(AB)^{T} = B^{T}A^{T}$.

To prove (29), begin with $(AA^{-1})^{T} = I^{T} = I$. But $(AA^{-1})^{T} = (A^{-1})^{T}A^{T}$. Hence, $(A^{-1})^{T} = (A^{T})^{-1}$.

10.6.2. Application to a mass-spring system. To illustrate a number of these ideas with a physical application, consider the arrangement of masses and springs shown in Fig. 1. The three masses are in static equilibrium under the action of prescribed applied forces $f_{1}, f_{2}, f_{3}$, and the $k$'s denote the stiffnesses of the various springs. For instance, $k_{12}$ denotes the stiffness of the spring connecting mass number 1 and mass number 2. Mass-spring systems are discussed in Section 1.3.

![Figure 1. Mass-spring system.](image-url)
10.6. Inverse Matrix, Cramer's Rule, Factorization

and in Example 3 of Section 3.9.1, which discussions should be reviewed if the following is not clear.

The free-body diagrams (i.e., the force diagrams) of the three masses are shown in Fig. 2, where it has been assumed, simply for definiteness, that \( x_1 > x_2 > x_3 > 0 \).

![Free-body diagrams for three masses](image)

**Figure 2.** Free-body diagrams.

From that assumption it follows that each spring is in compression except for the left-hand spring of stiffness \( k_1 \).

From Fig. 2 and Newton's second law, we obtain the equations of motion

\[
\begin{align*}
    m_1 \ddot{x}_1 &= f_1 - k_1 x_1 - k_{12}(x_1 - x_2) - k_{13}(x_1 - x_3), \\
    m_2 \ddot{x}_2 &= f_2 + k_{12}(x_1 - x_2) - k_{23}(x_2 - x_3), \\
    m_3 \ddot{x}_3 &= f_3 + k_{13}(x_1 - x_3) + k_{23}(x_2 - x_3) - k_3 x_3,
\end{align*}
\]

where primes denote differentiation with respect to the time \( t \). Since the system is in static equilibrium by assumption, \( \ddot{x}_1 = \ddot{x}_2 = \ddot{x}_3 = 0 \), and (31) becomes, in matrix form,

\[
\begin{bmatrix}
    (k_1 + k_{12} + k_{13}) & -k_{12} & -k_{13} \\
    -k_{12} & (k_{12} + k_{23}) & -k_{23} \\
    -k_{13} & -k_{23} & (k_{13} + k_{23} + k_3)
\end{bmatrix}
\begin{bmatrix}
    x_1 \\
    x_2 \\
    x_3
\end{bmatrix} =
\begin{bmatrix}
    f_1 \\
    f_2 \\
    f_3
\end{bmatrix}
\]

or

\[
Kx = f,
\]

where we will call \( K \) the *stiffness matrix*. We see that (33) is a matrix generalization of the simple Hooke's law \( f = kx \) for a single spring.

Is there a unique solution \( x = K^{-1}f \)? Experience and physical intuition probably tells us there is. Mathematically, everything hinges upon \( \det K \). If \( \det K \neq 0 \), there is a unique solution for \( x \), and if \( \det K = 0 \), then there is either no solution or an infinity of solutions. With five parameters within \( K \) (namely, \( k_1, k_{12}, k_{13}, k_{23}, k_3 \)) it is hard to imagine that we cannot have \( \det K = 0 \) for some choice(s) of those parameters. Let us see. Working out the determinant of \( K \), we find that

\[
\det K = k_1 (k_{12} k_{13} + k_{12} k_{23} + k_{12} k_3 + k_{23} k_{13} + k_{23} k_3) + k_3 (k_{12} k_{23} + k_{12} k_{13}).
\]
Since each sign is positive, and the $k$'s are positive, we see that $\det \mathbf{K} \neq 0$ so there is indeed a unique solution for $x$, namely, $x = \mathbf{K}^{-1} \mathbf{f}$.

However, suppose we degrade the system by removing one or more springs. We can see from (34) that even if we set any one $k$ value equal to zero (i.e., remove that spring), $\det \mathbf{K}$ is still positive. If we are willing to remove two springs, then we can obtain $\det \mathbf{K} = 0$ in either of two ways, by setting $k_1 = k_3 = 0$ or by setting $k_{12} = k_{23} = 0$. Let us consider the former, and leave the latter for the exercises.

With $k_1 = k_3 = 0$, $\mathbf{K}$ is singular (noninvertible) and (33) admits either no solution or an infinity of them. Which is it, and how is the result to be understood physically? Setting $k_1 = k_3 = 0$, (32) reduces to

$$
\begin{bmatrix}
(k_{12} + k_{13}) & -k_{12} & -k_{13} \\
-k_{12} & (k_{12} + k_{23}) & -k_{23} \\
-k_{13} & -k_{23} & (k_{13} + k_{23})
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix}
=
\begin{bmatrix}
f_1 \\
f_2 \\
f_3
\end{bmatrix}
$$

and, in augmented matrix form, Gauss elimination gives

$$
\begin{bmatrix}
1 & \frac{k_{23}}{k_{13}} & -1 - \frac{k_{23}}{k_{13}} & -\frac{f_3}{k_{13}} \\
0 & 1 + \frac{k_{23}}{k_{12}} & -1 - \frac{k_{23}}{k_{12}} - \frac{k_{23}}{k_{13}} & -\frac{f_3}{k_{12}} - \frac{f_2}{k_{12}} \\
0 & 0 & 0 & f_1 + f_2 + f_3
\end{bmatrix}
$$

which result reveals two possibilities.

(i) If $f_1 + f_2 + f_3 \neq 0$, then there is no solution. That mathematical result makes perfect sense physically, because with the end springs removed $f_1 + f_2 + f_3$ is the net lateral force on the three-mass system, and if that net force is nonzero, then the system cannot be in static equilibrium, as was assumed when we set $x_1' = x_2'' = x_3' = 0$ in (31)!

(ii) If $f_1 + f_2 + f_3 = 0$, then we see from (36) that there is an infinity of solutions, of the form

$$
x_3 = \alpha, \quad x_2 = \alpha + \text{etc.}, \quad x_1 = \alpha + \text{etc.},
$$

where $\alpha$ is an arbitrary constant and the two etc.'s involve the $f$'s and $k$'s. That is, the solution is nonunique because of the arbitrary translation $\alpha$. Again, that result makes sense physically because with $k_1 = k_3 = 0$ there are no end springs to restrain the three-mass system laterally.

Let us make one more important point. Observe from (32) that the $\mathbf{K}$ matrix is symmetric. Yet the system (Fig. 1) is not physically symmetric: that is, in general, $k_1 \neq k_3$ and $k_{12} \neq k_{23}$. Thus, the mathematical symmetry is somewhat unexpected and mysterious. In fact, we state without proof that for any number of masses interconnected with springs the resulting $\mathbf{K}$ matrix will be symmetric.

There is a striking consequence of the symmetry of $\mathbf{K}$, which we now explain. Property 12 gives $(\mathbf{K}^{-1})^T = (\mathbf{K}^T)^{-1} = \mathbf{K}^{-1}$ since $\mathbf{K}^T = \mathbf{K}$. Let us denote
\( K = \{a_{ij}\}, \) say, and let us compare the displacement \( x_3 \) of \( m_3 \) due to a unit load \( f_1 = 1 \) on \( m_1 \) (with \( f_2 = f_3 = 0 \)) with the displacement \( x_1 \) of \( m_1 \) due to a unit load \( f_3 = 1 \) on \( m_3 \) (with \( f_1 = f_2 = 0 \)). In the first case,

\[
\begin{bmatrix}
  x_1 \\
  x_2 \\
  x_3
\end{bmatrix} =
\begin{bmatrix}
  a_{11} & a_{12} & a_{13} \\
  a_{21} & a_{22} & a_{23} \\
  a_{31} & a_{32} & a_{33}
\end{bmatrix}
\begin{bmatrix}
  1 \\
  0 \\
  0
\end{bmatrix}
\]

gives \( x_3 = a_{31} \) and, in the second case,

\[
\begin{bmatrix}
  x_1 \\
  x_2 \\
  x_3
\end{bmatrix} =
\begin{bmatrix}
  a_{11} & a_{12} & a_{13} \\
  a_{21} & a_{22} & a_{23} \\
  a_{31} & a_{32} & a_{33}
\end{bmatrix}
\begin{bmatrix}
  0 \\
  0 \\
  1
\end{bmatrix}
\]

gives \( x_1 = a_{13} \). But these are the same \( (a_{31} = a_{13}) \) because \( K \) is symmetric. In this manner we find that

\[
\begin{pmatrix}
  \text{displacement } x_j \text{ of mass } m_j \\
  \text{due to unit load on mass } m_k
\end{pmatrix} =
\begin{pmatrix}
  \text{displacement } x_k \text{ of mass } m_k \\
  \text{due to unit load on mass } m_j
\end{pmatrix}. \tag{37}
\]

The latter "reciprocity" result can be generalized so as to apply to any linear elastic system and is known as Maxwell reciprocity. *

There is an electrical analog of the mechanical system shown in Fig. 1, a circuit containing resistors and voltage sources (such as batteries), and discussion of that case is left for the exercises.

10.6.3. Cramer's rule. We have seen that if \( A \) is \( n \times n \) and \( \det A \neq 0 \), then \( Ax = c \) has the unique solution

\[
x = A^{-1}c. \tag{38}
\]

To focus on the individual components of \( x \), rather than the entire \( x \) vector, let us write out (38):

\[
\begin{bmatrix}
  x_1 \\
  \vdots \\
  x_n
\end{bmatrix} =
\begin{bmatrix}
  a_{11} & a_{12} & \cdots & a_{1n} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{n1} & a_{n2} & \cdots & a_{nn}
\end{bmatrix}
\begin{bmatrix}
  c_1 \\
  \vdots \\
  c_n
\end{bmatrix} =
\begin{bmatrix}
  \sum_j a_{1j}c_j \\
  \vdots \\
  \sum_j a_{nj}c_j
\end{bmatrix}. \tag{39}
\]

Equating the \( i \)th component on the left with the \( i \)th component on the right, we have the scalar statement

\[
x_i = \sum_j a_{ij}c_j \tag{40}
\]

for any desired $i$ ($1 \leq i \leq n$). Or, recalling (17),

$$x_i = \sum_j \left( \frac{A_{ji}}{\det A} \right) c_j = \frac{\sum_j A_{ji}c_j}{\det A}. \quad (41)$$

Now, if the numerator on the right-hand side were $\sum_j A_{ji}a_{ji}$, instead, it would be recognizable as the determinant of $A$, namely, the cofactor expansion about the $i$th column. But the $a_{ji}$'s are replaced, in (41), by the $c_j$'s, so the numerator of (41) amounts to a determinant but not the determinant of $A$; rather, it is the determinant of the $A$ matrix with its $i$th column replaced by the column of $c_j$'s (or the $c$ vector, if you like).

The result, known as Cramer's rule, after Gabriel Cramer (1704–1752), is as follows.

**THEOREM 10.6.3 Cramer's Rule**

If $Ax = c$ where $A$ is invertible, then each component $x_i$ of $x$ may be computed as the ratio of two determinants; the denominator is $\det A$, and the numerator is also the determinant of the $A$ matrix but with the $i$th column replaced by $c$.

**EXAMPLE 3.** Let us solve the system

$$\begin{bmatrix} 1 & 3 & 0 \\ -2 & 3 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 5 \\ 1 \\ -2 \end{bmatrix} \quad (42)$$

for $x_1$ and $x_2$, say, using Cramer's rule. In this case $\det A = 8 \neq 0$ so the method is, first of all, applicable. Thus,

$$x_1 = \frac{\begin{vmatrix} 5 & 3 & 0 \\ 1 & 3 & 1 \\ -2 & 1 & 1 \end{vmatrix}}{8} \quad \text{and} \quad x_2 = \frac{\begin{vmatrix} 1 & 5 & 0 \\ -2 & 1 & 1 \\ 0 & -2 & 1 \end{vmatrix}}{8} = \frac{13}{8}, \quad (43)$$

where we have printed the "replacement columns" as boldface, for emphasis. \[\square\]

Cramer's rule, like the inverse matrix solution (38) from which it comes, has the advantage of being an explicit formula, rather than a method. It is also useful in that it permits us to focus on any single component of $x$ without having to compute the entire $x$ vector.

**10.6.4. Evaluation of $A^{-1}$ by elementary row operations.** Equation (16) gives $A^{-1}$ in terms of $\det A$ and $n^2$ cofactors, each of which is $\pm 1$ times an $(n - 1) \times
(n - 1) minor determinant. Each of these determinants can be evaluated by the cofactor expansion definition or, especially if n is large, by a faster method—such as triangularization. Alternatively, we can bypass (16) altogether, and determine $A^{-1}$ efficiently as follows.

Whether we are seeking $A^{-1}$ in order to solve a system $Ax = c$, or whether we are simply seeking the inverse of a given matrix $A$, observe that if we solve a system $Ax = c$ of $n$ equations in $n$ unknowns, or equivalently $Ax = Ic$, by Gauss-Jordan reduction, the result is the form $x = A^{-1}c$, or equivalently $Ix = A^{-1}c$. Symbolically, then, the sequence of elementary row operations effects the following transformation:

$$Ax = Ic$$
$$Ix = A^{-1}c.$$  

That is, at the same time that the row operations are transforming $A$ to $I$ they are also transforming $I$ to $A^{-1}$. Thus, we can skip $x$ and $c$ altogether, put $A$ and $I$ "side by side" as an augmented matrix $A|I$, and carry out elementary row operations on $A|I$ so as to reduce $A$, on the left, to $I$. When that has been accomplished, the matrix on the right will be $A^{-1}$.

**Example 4.** To illustrate, let us find the inverse of

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

Then

$$A|I = \begin{bmatrix} 1 & 3 & 0 & 1 & 0 & 0 \\ -2 & 3 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 3 & 0 & 1 & 0 & 0 \\ 0 & 9 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1/9 & -2/3 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 3 & 0 & 1 & 0 & 0 \\ 0 & 9 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1/9 & -2/3 & 1 & 0 \end{bmatrix}.$$  

$$\rightarrow \begin{bmatrix} 1 & 3 & 0 & 1 & 0 & 0 \\ 0 & 9 & 0 & 9/8 & 9/8 & 9/8 \\ 0 & 0 & 1 & -1/4 & -1/8 & -1/8 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 3 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1/4 & 1/8 & -1/8 \\ 0 & 0 & 1 & -1/4 & -1/8 & -1/8 \end{bmatrix}.$$  

$$\begin{bmatrix} 1 & 0 & 0 & 1/4 & -3/8 & 3/8 \\ 0 & 1 & 0 & 1/4 & 1/8 & -1/8 \\ 0 & 0 & 1 & -1/4 & -1/8 & -1/8 \end{bmatrix} (46)$$

so

$$A^{-1} = \begin{bmatrix} 1/4 & -3/8 & 3/8 \\ 1/4 & 1/8 & -1/8 \\ -1/4 & -1/8 & 3/8 \end{bmatrix}. (47)$$
Chapter 10. Matrices and Linear Equations

10.6.5. LU-factorization. This final subsection is not really about the inverse matrix or about the inverse matrix method of solving $A \mathbf{x} = \mathbf{c}$. Rather, it is about an alternative method of solution that is based upon the factorization of an $n \times n$ matrix $A$ as a lower triangular matrix $L$ times an upper triangular matrix $U$:

$$ A = LU = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}, \quad (48) $$

where we have taken $n = 3$ simply for compactness. If we carry out the multiplication on the right and equate the nine elements of $LU$ to the corresponding elements of $A$ we obtain nine equations in the 12 unknown $l_{ij}$'s and $u_{ij}$'s. Since we have more unknowns than equations, there is some flexibility in implementing the idea. Hence there are various versions of LU-factorization.

According to Doolittle's method we can set each $l_{jj} = 1$ in $L$ (i.e., the diagonal elements) and solve uniquely for the remaining $l_{ij}$'s and the $u_{ij}$'s. With $L$ and $U$ determined, we then solve $A \mathbf{x} = LU \mathbf{x} = \mathbf{c}$ by setting $U \mathbf{x} = \mathbf{y}$ so that $L(U \mathbf{x}) = \mathbf{c}$ breaks into the two problems:

$$ L \mathbf{y} = \mathbf{c}, \quad (49a) $$
$$ U \mathbf{x} = \mathbf{y}, \quad (49b) $$

each of which is simple because $L$ and $U$ are triangular. We solve (49a) for $\mathbf{y}$, put that $\mathbf{y}$ into (49b), and then solve (49b) for $\mathbf{x}$. Let us illustrate the procedure.

EXAMPLE 5. To solve

$$ \begin{bmatrix} 2 & -3 & 3 \\ 6 & -8 & 7 \\ -2 & 6 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -2 \\ -3 \\ 3 \end{bmatrix} \quad (50) $$

by the Doolittle LU-factorization method, we first need to determine $L$ and $U$ by equating

$$ \begin{bmatrix} 2 & -3 & 3 \\ 6 & -8 & 7 \\ -2 & 6 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ l_{21} & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix}. \quad (51) $$

Matching $a_{11}, a_{12}, a_{13}, a_{21}, \ldots, a_{32}, a_{33}$ to the corresponding terms on the right gives a sequence of equations for $u_{11}, u_{12}, u_{13}, l_{21}, u_{22}, u_{23}, l_{31}, l_{32},$ and $u_{33}$ (i.e., the underlined
The beauty of the method is that the $l_{ij}$'s and $u_{ij}$'s are found not by solving simultaneous coupled equations but by solving a sequence of linear equations in only one unknown [as illustrated in (52)]. With $L$ and $U$ thus determined, the solution of (49a) and (49b) is likewise simple since $L$ and $U$ are triangular. In fact, the method is around twice as fast as Gauss–Jordan elimination.

**Closure.** The inverse of a matrix $A$, denoted as $A^{-1}$, exists if and only if $A$ is square ($n \times n$) and $\det A \neq 0$. If it exists it is given uniquely as

$$A^{-1} = \frac{1}{\det A} \text{adj} A,$$

where the matrix $\text{adj} A$ is the adjoint of $A$ (the transpose of the cofactor matrix), and is such that

$$A^{-1} A = AA^{-1} = I. \quad (54)$$

The case where $A$ is not invertible (i.e., is singular) is the exceptional case; in the generic case a given $n \times n$ matrix is invertible. Besides (54), several useful properties of inverses are given as 11–14 in Section 10.6.1.

If $A$ is invertible, then $Ax = c$ admits the unique solution

$$x = A^{-1} c = \frac{1}{\det A} [\text{adj} A] c, \quad (55)$$
which result gives us Cramer’s rule, whereby each component of \( \mathbf{x} \) is expressed as the ratio of \( n \times n \) determinants.

Notice that the equation \( a\mathbf{x} = \mathbf{c} \) has a unique solution if and only if \( a \neq 0 \). For \( A\mathbf{x} = \mathbf{c} \), where \( A \) is \( n \times n \), that condition generalizes not to \( A \neq 0 \) but to \( \text{det}A \neq 0 \).

In contrast with Gauss–Jordan reduction and LU-factorization, which are solution methods, (55) and Cramer’s rule are explicit formulas for the solution (when a unique solution does exist).

Finally, we urge you to be careful with the sequencing of matrices because of the general absence of commutativity under multiplication. For instance, \( A\mathbf{x} = \mathbf{c} \) implies \( \mathbf{x} = A^{-1}\mathbf{c} \) (if \( A \) is invertible), NOT \( \mathbf{c} = A\mathbf{x}^{-1} \). Indeed, the product \( cA^{-1} \) is not even defined (unless \( n = 1 \)) since \( \mathbf{c} \) is \( n \times 1 \) and \( A^{-1} \) is \( n \times n \).

**Computer software.** Using *Maple*, the relevant command is `inverse(A)`, within the `linalg` package. For instance, to invert

\[
A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix},
\]

enter

\[
\text{with(linalg):}
\]

and return. Then the steps

\[
A := \text{array}([[1, 2], [3, 4]]):
\]

and

\[
\text{inverse}(A);
\]

give the result

\[
\begin{bmatrix} -2 & 1 \\ 3 & -1/2 \end{bmatrix}
\]

**EXERCISES 10.6**

1. Use (17) to evaluate the inverse matrix. If the matrix is not invertible, state that.

(a) \[
\begin{bmatrix} a & b \\ c & d \end{bmatrix}, \text{ where } ad - bc \neq 0
\]

(b) \[
\begin{bmatrix} 5 & 4 \\ 3 & 2 \end{bmatrix}
\]

(c) \[
\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}
\]

(d) \[
\begin{bmatrix} 2 & 1 \\ 2 & -1 \end{bmatrix}
\]

(e) \[
\begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}
\]

(f) \[
\begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 3 \\ 0 & 3 & -1 \end{bmatrix}
\]

(g) \[
\begin{bmatrix} 0 & 1 & 0 \\ 2 & 0 & 5 \\ 0 & 0 & 3 \end{bmatrix}
\]

(h) \[
\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}
\]

(i) \[
\begin{bmatrix} 0 & 0 & 2 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix}
\]

(j) \[
\begin{bmatrix} 3 & 1 & -2 \\ 1 & 2 & 1 \\ 1 & -3 & -3 \end{bmatrix}
\]

(k) \[
\begin{bmatrix} 7 & 1 & 3 \\ 2 & -1 & 1 \\ 0 & 1 & 4 \end{bmatrix}
\]
2. (c)-(o) Evaluate the inverse of the matrix given in the corresponding part of Exercise 1 using elementary row operations, as we did in Example 4.

3. (c)-(o) Evaluate the inverse of the matrix given in the corresponding part of Exercise 1 using computer software.

4. (Block-diagonal matrices) If an $n \times n$ matrix $A$ can be partitioned as

$$A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_k \end{bmatrix} \quad (k \leq n)$$

it is said to be block diagonal. All of the $A_j$ submatrices need to be square, although not necessarily of equal order, with their main diagonals coinciding with the main diagonal of $A$. For instance,

$$A = \begin{bmatrix} 2 & 1 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 3 & 0 \\ 0 & 0 & 2 & 5 & 1 & 0 \\ 0 & 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 \end{bmatrix} = \begin{bmatrix} A_1 & 0 & 0 \\ 0 & A_2 & 0 \\ 0 & 0 & A_3 \end{bmatrix}$$

is block diagonal. Such matrices exhibit essentially the same simple features as diagonal matrices.

(a) Show that $A$ is invertible if and only if $A_1, \ldots, A_k$ are.

HINT: Recall equation (9.1) in Exercise 9, Section 10.4.

(b) Assuming that $A_1, \ldots, A_k$ are invertible, verify that

$$A^{-1} = \begin{bmatrix} A_1^{-1} & 0 & \cdots & 0 \\ 0 & A_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & A_k^{-1} \end{bmatrix}$$

(c) Use the latter result to evaluate $A^{-1}$, where $A$ is given in (4.1).

5. Solve for $x_1$ and $x_2$ by Cramer’s rule.

(a) $x_1 + 4x_2 = 0$

(b) $ax_1 + bx_2 = c$

(c) $x_1 - 2x_2 + x_3 = 4$

(d) $x_1 + 2x_2 + 3x_3 = 9$

(e) $2x_1 + x_2 = 1$

(f) $x_1 + x_2 + x_3 = 1$

6. (a) Given a certain $3 \times 3$ matrix $A$, we find its inverse to be

$$A^{-1} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 0 \end{bmatrix}.$$ Can that result be correct? Explain.

(b) Same as (a), for

$$A^{-1} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 0 & 0 \end{bmatrix}.$$  

7. If $A^{-1}$ is the given matrix, find $A$. 

(a) $\begin{bmatrix} 3 & -1 \\ 3 & -2 \end{bmatrix}$ 

(b) $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ 

(c) $\begin{bmatrix} 1 & 2 & 1 \\ 0 & 3 & 1 \\ 1 & 1 & 0 \end{bmatrix}$ 

(d) $\begin{bmatrix} 2 & 1 & 1 \\ 4 & 1 & 1 \\ 6 & 2 & 1 \end{bmatrix}$. 

10.6. Inverse Matrix, Cramer’s Rule, Factorization 523
8. Suppose that $A x = c$ is a linear system of order 3, and that to the $c$ vectors

$$c = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

there correspond unique solutions

$$x = \begin{bmatrix} 2 \\ 5 \\ 1 \\ 1 \\ 3 \\ 0 \\ 0 \\ 4 \end{bmatrix},$$

respectively. Then what is the solution $x$ corresponding to $c = [4, 3, -1]^T$? Is it unique? Explain.

9. Suppose that Gauss elimination gives the solution of a linear system $A x = c$ as $x = x_0 + \alpha_1 x_1 + \alpha_2 x_2$, where $A$ is $6 \times 6$ and $\alpha_1$ and $\alpha_2$ are arbitrary. Is $A$ invertible? Explain.

10. (Nilpotent matrices) If there is some positive integer $p$ such that $A^p = 0$, then $A$ said to be nilpotent (i.e., potentially nil).

(a) Show that a nilpotent matrix is necessarily singular.

(b) If $A$ is nilpotent, with $A^p = 0$, then

$$(I - A)^{-1} = I + A + A^2 + \cdots + A^{p-1}. \quad (10.1)$$

11. First, read Exercise 10. Use (10.1) to find the inverse of the given matrix. HINT: You will need to identify $A$.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 8 \\ 0 & 0 & 1 \end{bmatrix}$$

(a) \( \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 2 & 7 & 1 \end{bmatrix} \)

(b) \( \begin{bmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ 0 & 3 & 1 \end{bmatrix} \)

(c) \( \begin{bmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ 2 & 1 & 6 \end{bmatrix} \)

12. First, read Exercise 10. Use (10.1) to find the inverse of the given matrix.

$$A = \begin{bmatrix} 2 & 8 & 10 \\ 0 & 3 & 12 \\ 0 & 0 & 4 \end{bmatrix}$$

(a) \( \begin{bmatrix} 2 & 8 & 10 \\ 0 & 3 & 12 \\ 0 & 0 & 4 \end{bmatrix}^{-1} \)

\( = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{bmatrix}^{-1} \)

\( = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \)

\( = \begin{bmatrix} 1 & 4 & 5 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}^{-1} \)

\( = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix} \)

13. (Ill-conditioned systems) Consider the system

$$\begin{bmatrix} 3 & 0 & 0 \\ 4 & -2 & 0 \\ 10 & 0 & 2 \end{bmatrix} x = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \quad (a)$$

or $A x = c.$ (13.1)

(a) Evaluate $A^{-1}$, by any analytical means that you wish, and show that $x = A^{-1} c = [-3, -12, 30]^T$.

(b) To simulate the effects of roundoff error, consider in place of (13.1) the rounded off system

$$\begin{bmatrix} 1 & 0.5 & 0.33 \\ 0.5 & 0.33 & 0.25 \\ 0.33 & 0.25 & 0.2 \end{bmatrix} x = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}. \quad (13.2)$$

Solving (13.2) by any means you wish (computer software being the easiest), show that the solution of (13.2) is $x \approx [11.1, -84.1, 96.8]^T$. NOTE: In this example we see that only a slight error in $A$ leads to a disproportionately large error in the solution. Hence, (13.1) is said to be ill-conditioned. (Ill-conditioned systems are also mentioned in Exercise 13, Section 8.3.) In applications it is important to know if a given system is ill-conditioned so that steps can be taken to obtain a sufficiently accurate solution. According to one criterion in the literature, an $n \times n$ matrix $A$ may be considered as ill-conditioned if

$$|\text{det}A| \left( \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}}{n} \right) \ll 1. \quad (13.3)$$

For the Hilbert matrix in (13.1), the left-hand side of (13.3) is 0.00033, which is indeed much smaller than unity.

(c) In place of (13.2), use the more accurate rounded off system

$$\begin{bmatrix} 1 & 0.5 & 0.333 \\ 0.5 & 0.333 & 0.25 \\ 0.333 & 0.25 & 0.2 \end{bmatrix} x = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}. \quad (13.4)$$

and see how much closer the solution of (13.4) comes to the exact solution $x = [-3, -12, 30]^T$ of (13.1).
16. In Section 10.6.2 we stated that if we are willing to remove two springs then we can have \( \det K = 0 \) either by setting \( k_1 = k_3 = 0 \) or by setting \( k_{12} = k_{23} = 0 \). We discuss the former choice, \( k_1 = k_3 = 0 \), both mathematically and physically as well. Do the same for the latter choice, \( k_{12} = k_{23} = 0 \).
17. (A dc circuit) Application of Kirchhoff’s laws to the circuit shown in Exercise 10 of Section 8.3 produced the five equations (10.1) on the currents \( i_1, i_2, i_3 \). Of these equations, the second is the same as the first, and the fifth is the fourth minus the third. Thus, deleting those two redundant equations leaves the system

\[
\begin{align*}
  i_1 - i_2 - i_3 &= 0 \\
  R_3i_2 - R_3i_3 &= 0 \\
  R_1i_1 + R_2i_2 &= E.
\end{align*}
\]

(a) Show that the determinant of the coefficient matrix in (17.1) is necessarily nonzero, so that the system \( Ri = e \) given by (17.1) necessarily admits the unique solution \( i = R^{-1}e \).

NOTE: \( R_1 > 0, R_2 > 0, \) and \( R_3 > 0 \).

(b) Solve for \( i \) by the inverse matrix method. Also, solve for \( i_1, i_2, i_3 \) by Cramer’s rule and verify that the results are the same.

(c) Suppose, instead, that we allow one or more of the resistances to be zero so that \( R_1 \geq 0, R_2 \geq 0, R_3 \geq 0 \). Show that if any two, or all three, of the resistances are zero, then the determinant does vanish so that equations (17.1) admit either no solution or an infinity of solutions. For each of these four “singular” cases determine whether there is no solution or an infinity of solutions by applying Gauss elimination. State the physical significance of each of these results insofar as possible.

18. (Circuit analog) The electrical circuit analog of the mass-spring system shown in Fig. 1 is shown below.

(a) Applying Kirchhoff’s voltage law, show that

\[
\begin{bmatrix}
  (R_1 + R_{12} + R_{13}) & -R_{12} & -R_{13} \\
  -R_{12} & (R_{12} + R_{23}) & -R_{23} \\
  -R_{13} & -R_{23} & (R_{13} + R_{23} + R_3)
\end{bmatrix}
\begin{bmatrix}
  i_1 \\
  i_2 \\
  i_3
\end{bmatrix}
= \begin{bmatrix}
  E_1 \\
  E_2 \\
  E_3
\end{bmatrix},
\]

which is the analog of (32) under the correspondence \( R_{ij} \leftrightarrow k_{ij}, i_j \leftrightarrow x_j, E_j \leftrightarrow f_j \).

(b) Discuss the existence and uniqueness of solutions of (18.1) in the same way that we did that for the mass-spring system in Section 10.6.2, including a reciprocity result analogous to (37).


(a) \[
\begin{bmatrix}
  2 & 3 \\
  8 & -1
\end{bmatrix}
\begin{bmatrix}
  x_1 \\
  x_2
\end{bmatrix}
= \begin{bmatrix}
  4 \\
  0
\end{bmatrix}
\]

(b) \[
\begin{bmatrix}
  2 & -1 \\
  2 & 1
\end{bmatrix}
\begin{bmatrix}
  x_1 \\
  x_2
\end{bmatrix}
= \begin{bmatrix}
  7 \\
  13
\end{bmatrix}
\]

(c) \[
\begin{bmatrix}
  2 & 5 & 1 \\
  2 & 8 & 0 \\
  8 & 2 & 2
\end{bmatrix}
\begin{bmatrix}
  x_1 \\
  x_2 \\
  x_3
\end{bmatrix}
= \begin{bmatrix}
  0 \\
  -7 \\
  10
\end{bmatrix}
\]

(d) \[
\begin{bmatrix}
  1 & 3 & -1 \\
  2 & 2 & 0 \\
  3 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
  x_1 \\
  x_2 \\
  x_3
\end{bmatrix}
= \begin{bmatrix}
  0 \\
  6 \\
  12
\end{bmatrix}
\]

20. (Dual or reciprocal set) In Exercise 13 of Section 9.9 we introduced the concept of a set of dual or reciprocal vectors \( \{e_1^*, \ldots, e_n^*\} \) corresponding to a basis \( \{e_1, \ldots, e_n\} \) that is not necessarily orthogonal, or ON. Having learned more about the solution of systems of linear algebraic equations in Sections 10.5 and 10.6, we can now return to that exercise and prove the claim made in part (a). Specifically, prove the dual set exists, is unique, and is itself a basis for \( \mathbb{R}^n \).
10.7 Change of Basis (Optional)

In a given problem one selects what appears to be the most convenient basis, but at some stage of the analysis it may be desirable to switch to some other basis. For example, in studying the aerodynamics of a propeller it is generally most convenient to carry out the analysis (of the propeller-induced pressure field, for example) with respect to a propeller-fixed basis, one that rotates with the propeller, although eventually we may wish to relate quantities back to a stationary (nonrotating) basis. How do the coordinates (i.e., the components) of a given vector change as we change the basis? It is that question which we address in this section.

Let 

\[ B = \{ e_1, \ldots, e_n \} \]

be a given basis for the vector space \( V \) under consideration so that any given vector \( x \) in \( V \) can be expanded as

\[ x = x_1 e_1 + \cdots + x_n e_n. \]  

(1)

If we switch to some other basis \( B' = \{ e'_1, \ldots, e'_n \} \), then we may, similarly, expand the same vector \( x \) as

\[ x = x'_1 e'_1 + \cdots + x'_n e'_n. \]  

(2)

How are the \( x'_j \) coordinates related to the \( x_j \) coordinates? Since \( B' \) is a basis, we may expand each of the \( e_j \)'s in terms of \( B' \):

\[ e_1 = q_{11} e'_1 + \cdots + q_{1n} e'_n, \]

\[ \vdots \]

\[ e_n = q_{n1} e'_1 + \cdots + q_{nn} e'_n. \]  

(3)

Putting (3) into (1) gives

\[ x = x_1 (q_{11} e'_1 + \cdots + q_{1n} e'_n) + \cdots + x_n (q_{n1} e'_1 + \cdots + q_{nn} e'_n) \]

\[ = (x_1 q_{11} + \cdots + x_n q_{n1}) e'_1 + \cdots + (x_1 q_{n1} + \cdots + x_n q_{nn}) e'_n, \]  

(4)

and a comparison of (2) and (4) gives the desired relations

\[ x'_1 = q_{11} x_1 + \cdots + q_{1n} x_n, \]

\[ \vdots \]

\[ x'_n = q_{n1} x_1 + \cdots + q_{nn} x_n \]  

(5)

or, in matrix notation,

\[ [x]_{B'} = Q [x]_{B}, \]  

(6)

where

\[ Q = \begin{bmatrix} q_{11} & \cdots & q_{1n} \\ \vdots & \ddots & \vdots \\ q_{n1} & \cdots & q_{nn} \end{bmatrix} \]  

(7)
and
\[
[x]_B = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad [x]_{B'} = \begin{bmatrix} x'_1 \\ \vdots \\ x'_n \end{bmatrix}.
\] (8)

We call $[x]_B$ the coordinate vector of the vector $x$ with respect to the ordered basis $B$, and similarly for $[x]_{B'}$, and we call $Q$ the coordinate transformation matrix from $B$ to $B'$.

Thus far, our results apply whether the bases are orthogonal or not. In the remainder of this section we assume that both bases, $B$ and $B'$, are ON. Thus, let us rewrite (3) as
\[
\hat{e}_1 = q_{11} \hat{e}'_1 + \cdots + q_{1n} \hat{e}'_n, \\
\vdots \\
\hat{e}_n = q_{n1} \hat{e}'_1 + \cdots + q_{nn} \hat{e}'_n,
\] (9)

where the carets denote unit vectors, as usual. If we dot $\hat{e}'_1$ into both sides of the first equation in (9), and remember that $B'$ is ON, we obtain $q_{11} = \hat{e}'_1 \cdot \hat{e}_1$. Dotting $\hat{e}'_2$ gives $q_{21} = \hat{e}'_2 \cdot \hat{e}_1$, . . . , dotting $\hat{e}'_n$ gives $q_{n1} = \hat{e}'_n \cdot \hat{e}_1$, and similarly for the second through $n$th equation in (9). The result is the formula
\[
q_{ij} = \hat{e}'_i \cdot \hat{e}_j.
\] (10)

which tells us how to compute the transformation matrix $Q$.

There are two properties of the $Q$ matrix to address before turning to an example. To obtain the first of these, observe that
\[
Q^T Q = \begin{bmatrix} q_{11} & \cdots & q_{1n} \\ q_{21} & \cdots & q_{2n} \\ \vdots & \vdots & \vdots \\ q_{n1} & \cdots & q_{nn} \end{bmatrix} \begin{bmatrix} q_{11} \\ q_{12} \\ \vdots \\ q_{nn} \end{bmatrix} = \begin{bmatrix} \hat{e}_1 \cdot \hat{e}_1 & \cdots & \hat{e}_1 \cdot \hat{e}_n \\ \vdots & \vdots & \vdots \\ \hat{e}_n \cdot \hat{e}_1 & \cdots & \hat{e}_n \cdot \hat{e}_n \end{bmatrix} = I
\] (11)

so that
\[
Q^{-1} = Q^T.
\] (12)

It was useful to partition the $Q^T$ and $Q$ matrices in (11) because we can see from (9) that the columns of $Q$ (and hence the rows of $Q^T$) are actually the $\hat{e}_1, \ldots, \hat{e}_n$ vectors (in $n$-tuple form). Thus, from the way matrix multiplication is defined, we can see that the elements of the product matrix are dot products. Specifically, the $i, j$ element of $Q^T Q$ is $\hat{e}_i \cdot \hat{e}_j$, which is the Kronecker delta $\delta_{ij}$. Hence, $Q^T Q$ equals the identity matrix $I$, so $Q^T$ must be the inverse of $Q$, as stated in (12).
That result makes it easy for us to reverse equation (6) – that is, to solve for $[x]_B$ in terms of $[x]_{B'}$ for then $[x]_B = Q^{-1}[x]_{B'} = Q^T [x]_{B'}$. In other words, we do not need to face up to the evaluation of $Q^{-1}$ since $Q^{-1}$ is merely $Q^T$.

Any matrix with the useful property (12) is known as an **orthogonal matrix** because it follows from (12) that the column vectors in $Q$ are orthonormal.

As the second property of $Q$, observe that it also follows from (12) that

$$\det Q = \pm 1,$$

(13)

that is, either $+1$ or $-1$ since $Q^TQ = I$ implies that $\det(Q^TQ) = \det I = 1$. But $\det(Q^TQ) = (\det Q^T)(\det Q) = (\det Q)(\det Q) = (\det Q)^2$. Hence, $\det Q$ must be $+1$ or $-1$.

**EXAMPLE 1.** *Rotation in the Plane.* Consider the vector space $\mathbb{R}^2$, with the ON bases $B = \{\hat{e}_1, \hat{e}_2\}$ and $B' = \{\hat{e}_1', \hat{e}_2'\}$ shown in Fig. 1. $B'$ is obtained from $B$ by a counterclockwise rotation through an angle $\theta$ (or clockwise if $\theta$ is negative). From the figure,

$$q_{11} = \hat{e}_1' \cdot \hat{e}_1 = (1)(1) \cos \theta = \cos \theta,$$

$$q_{12} = \hat{e}_1' \cdot \hat{e}_2 = (1)(1) \cos \left( \frac{\pi}{2} - \theta \right) = \sin \theta,$$

$$q_{21} = \hat{e}_2' \cdot \hat{e}_1 = (1)(1) \cos \left( \frac{\pi}{2} + \theta \right) = -\sin \theta,$$

$$q_{22} = \hat{e}_2' \cdot \hat{e}_2 = (1)(1) \cos \theta = \cos \theta,$$

so that the coordinate transformation matrix is

$$Q = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}. $$

(15)

Hence,

$$\begin{bmatrix} x'_1 \\ x'_2 \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. $$

(16)

Or, the other way around,

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}^{-1} \begin{bmatrix} x'_1 \\ x'_2 \end{bmatrix} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}^T \begin{bmatrix} x'_1 \\ x'_2 \end{bmatrix}$$

(17)

so

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x'_1 \\ x'_2 \end{bmatrix}. $$

COMMENT 1. It is easy enough to check (16) and (17) for one or two special cases. For example, if $\theta = 0$ then the two bases coincide so we should have $x'_1 = x_1$ and $x'_2 = x_2$, and that is what (16) and (17) give. Also, if $\theta = \pi/2$, say, we should have $x'_1 = x_2$ and $x'_2 = -x_1$ and, again, that is what (16) and (17) give.

COMMENT 2. Two ON bases in a plane are not necessarily related through a rotation. In this example, for instance, if we reverse the direction of $\hat{e}_2'$, then $\{\hat{e}_1', \hat{e}_2'\}$ is still ON.
Given \( \{ e_1, e_2 \} \) by means of a rotation alone. Rather, we
need a rotation and a reflection, a counterclockwise rotation through an angle \( \theta \), and then
a reflection about \( AA \) (or, first a reflection about the \( e_1 \) axis and then a counterclockwise
rotation through an angle \( \theta \)). In this case

\[
\begin{align*}
q_{11} &= \hat{e}_1' \cdot \hat{e}_1 = \cos \theta, \\
q_{12} &= \hat{e}_1' \cdot \hat{e}_2 = \cos \left( \frac{\pi}{2} - \theta \right) = \sin \theta, \\
q_{21} &= \hat{e}_2' \cdot \hat{e}_1 = \cos \left( \frac{\pi}{2} - \theta \right) = \sin \theta, \\
q_{22} &= \hat{e}_2' \cdot \hat{e}_2 = \cos (\pi - \theta) = -\cos \theta
\end{align*}
\]

so that \( Q \) is the orthogonal matrix

\[
Q = \begin{bmatrix}
\cos \theta & \sin \theta \\
\sin \theta & -\cos \theta
\end{bmatrix},
\]

Recall from (13) that \( \det Q \) is either +1 or −1. For the case where \( B \) and \( B' \) are
related through a pure rotation \([Q \text{ given by (15)} \det Q = +1 \) and for the case where they
are related through a reflection and a rotation \([Q \text{ given by (18)} \det Q = -1 \). \]

**Closure.** In this brief section we study the relationship between the components,
or coordinates, of any given vector \( x \) expanded in terms of two different bases \( B \)
and \( B' \). We find the linear relationship (6), where the \( q_{ij} \) elements of the coordinate
transformation matrix \( Q \) are the expansion coefficients of \( e_i \) in terms of \( e_1', \ldots, e_n' \),
as indicated by (3).

If \( B \) and \( B' \) are ON, then the \( q_{ij} \)’s are computed, simply, from (10), and \( Q \)
adopts the properties that \( Q^T = Q^{-1} \) and that \( \det Q = +1 \) or −1. *Any matrix \( Q \)
having the property \( Q^T = Q^{-1} \) has ON column vectors and is called an orthogonal
matrix.*

**EXERCISES 10.7**

1. Given that \( e_2 = e_1' + 2e_2' \) and \( e_2 = e_1' - e_2' \), find the
coordinate transformation matrix \( Q \). Is \( Q \) orthogonal? If
\( [x]_B = [5, -1]^T \), find \( [x]_{B'} \). If \( [x]_{B'} = [2, 3]^T \), find \( [x]_B \).
2. Given that \( e_1 = e_1' + e_2' - e_3' \), \( e_2 = e_1' - e_2' + e_3' \), and
\( e_3 = -e_1' + e_2' + e_3' \), find the coordinate transformation ma-
trix \( Q \). Is \( Q \) orthogonal? If \( [x]_B = [4, 1, -2]^T \), find \( [x]_{B'} \). If
\( [x]_{B'} = [1, 0, 2]^T \), find \( [x]_B \).
3. Let \( \hat{e}_1 = [1, 0, 0]^T \), \( \hat{e}_2 = [0, 1, 0]^T \), and \( \hat{e}_1' = \frac{1}{\sqrt{5}}[2, 1, 1]^T \),
\( \hat{e}_2' = \frac{1}{\sqrt{5}}[1, -2, 1]^T \).
4. Let \( \hat{e}_1 = [1, 0, 0]^T \), \( \hat{e}_2 = [0, 1, 0]^T \), \( \hat{e}_3 = [0, 0, 1]^T \), \( \hat{e}_4 = [1, 0, 0]^T \), \( \hat{e}_5 = \frac{1}{\sqrt{2}}[1, 1, 0]^T \), \( \hat{e}_6 = \frac{1}{\sqrt{2}}[1, 1, 0]^T \), \( \hat{e}_7 = \frac{1}{\sqrt{6}}[1, -1, 0]^T \), \( \hat{e}_8 = \frac{1}{\sqrt{6}}[1, -1, 0]^T \).
(a) Find the coordinate transformation matrix \( Q \). Is \( Q \) orthog-
onal?
(b) If \( [x]_B = [1, 1, 2, 5]^T \), find \( [x]_{B'} \).
(c) If \([x]_{u'} = [1, 1, 2, 5]^T\), find \([x]_u\).

5. Show whether or not these matrices are orthogonal.

\[
\begin{align*}
(a) & \; \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \\
(b) & \; \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \\
(c) & \; \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \\
(d) & \; \begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \\
(e) & \; \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \\
(f) & \; \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix}
\end{align*}
\]

6. For the case of rotation in a plane, the transformation \(Q\) corresponded to a counterclockwise rotation of the basis through an angle \(\theta\). Does \(Q^{-1}\) correspond to the reverse of this, a clockwise rotation \(\theta\)? Prove or disprove.

7. (Rotation and reflection) (a) Show that every orthogonal coordinate transformation matrix of order 2 is of one of the following two types:

\[
Q_1 = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}, \quad Q_2 = \begin{bmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{bmatrix},
\]

i.e., as given by the pure rotation (15) or by the rotation plus reflection (18).

(b) Show that these two cases can be distinguished by the sign of the determinant, specifically, that \(\det Q = +1\) if \(Q\) corresponds to pure rotation, and that \(\det Q = -1\) if \(Q\) corresponds to rotation plus reflection.

8. (a) Prove that if \(Q\) is orthogonal, then so is \(Q^T\).

(b) Prove that if \(Q\) is orthogonal, then so is \(Q^{-1}\).

9. Evaluate \[\begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}^n\].

---

10.8 Vector Transformation (Optional)

Recall that a real-valued function \(f\) of a real variable \(x\) is a rule that assigns a uniquely determined value \(f(x)\) to each specified value \(x\) as illustrated in Fig. 1. Thus, \(f\) is a transformation, or mapping, from points on an \(x\) axis to points on an \(f\) axis, and we view \(x\) as the "input" and \(f(x)\) as the "output." [A more familiar graphical display of \(f\), called the graph of \(f\), can be obtained if, following Descartes (1596–1650), we arrange the \(x\) and \(f\) axes at right angles to each other and plot the set of points \(x, f(x)\) as illustrated in Fig. 2.]

In this section we reconsider vectors and matrices from this transformation point of view. Specifically, we consider vector-valued functions \(F\) of a vector variable \(x\). That is, the "input" is now a vector \(x\) from some vector space \(V\), and the function \(F\) assigns a uniquely determined "output" vector \(F(x)\) in some vector space \(W\). We call \(F\) a transformation, or mapping, from \(V\) into \(W\), and denote it as

\[F : V \to W.\]

We call \(V\) the domain of \(F\) and \(W\) the range of definition of \(F\). \(W\) may, but need not, be identical to \(V\). If it is identical, then \(F : V \to V\) is called an operator on \(V\).
EXAMPLE 1. To illustrate, consider the transformation \( F : \mathbb{R}^4 \rightarrow \mathbb{R}^3 \) defined by

\[
F(x) = \begin{bmatrix}
  x_1 - 2x_2 + x_4 \\
  x_1 + x_3 \\
  x_1 + 2x_2 + 2x_3 - x_4
\end{bmatrix}.
\] (1)

Here \( V = \mathbb{R}^4 \), \( W = \mathbb{R}^3 \), and the input vector \( x \) and the output vector \( F(x) \) are

\[
x = \begin{bmatrix}
  x_1 \\
  x_2 \\
  x_3 \\
  x_4
\end{bmatrix}
\] in \( \mathbb{R}^4 \), and

\[
F(x) = \begin{bmatrix}
  x_1 - 2x_2 + x_4 \\
  x_1 + x_3 \\
  x_1 + 2x_2 + 2x_3 - x_4
\end{bmatrix}
\] in \( \mathbb{R}^3 \). (2)

For example, if \( x = [2, 3, 6, -1]^T \), then \( F(x) = [-5, 8, 21]^T \). This transformation is not an operator since \( W = \mathbb{R}^5 \), whereas \( V = \mathbb{R}^4 \).

We say that the vector \( F(x) \) in \( W \) is the image of the vector \( x \) in \( V \) under the transformation \( F \), and that \( x \) is the inverse image of \( F(x) \).

Since \( V \) and \( W \) are vector spaces, each must contain a zero vector. We will denote these zero vectors as \( 0_V \) and \( 0_W \), respectively. Finally, we define the image of \( V \) in \( W \) as the range \( R \) of \( F \), and we define the inverse image of \( 0_W \) in \( V \) as the nullspace or kernel \( K \) of \( F \). That is, the kernel \( K \) is the part of \( V \) that maps to the zero vector \( 0_W \) in \( W \).

EXAMPLE 2. Let us find the range and kernel of the transformation \( F \) given in Example 1. First, the range. The range of \( F \) is the set of all vectors \( c \) in \( W \) for which the equation \( F(x) = c \) is consistent, that is, has at least one solution \( x \) in \( V \). In the present case \( F(x) = c \) is

\[
\begin{bmatrix}
  x_1 - 2x_2 + x_4 \\
  x_1 + x_3 \\
  x_1 + 2x_2 + 2x_3 - x_4
\end{bmatrix} = \begin{bmatrix}
  c_1 \\
  c_2 \\
  c_3
\end{bmatrix}
\] (3)

or, in scalar form,

\[
x_1 - 2x_2 + x_4 = c_1,
\]
\[
x_1 + x_3 = c_2,
\]
\[
x_1 + 2x_2 + 2x_3 - x_4 = c_3.
\] (4)

Applying elementary operations to (4), we obtain the equivalent system

\[
x_1 - 2x_2 + x_4 = c_1,
\]
\[
2x_2 + x_3 - x_4 = c_2 - c_1,
\]
\[
0 = c_3 - 2c_2 + c_1.
\] (5a,b,c)

This system is consistent if and only if \( c \) lies in the plane (through the origin) defined by \( c_3 - 2c_2 + c_1 = 0 \). That plane is a two-dimensional subspace of \( \mathbb{R}^3 \) and is the range \( R \) of \( F \).
Turning to the kernel $K$ of $F$, that is, the inverse image of $[0,0,0]^T$ in $\mathbb{R}^4$, we need merely set $c_1 = c_2 = c_3 = 0$ in (5) and solve for $x$. That solution gives

$$x = \alpha_1 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 2 \end{bmatrix} + \alpha_2 \begin{bmatrix} -2 \\ -1 \\ 2 \\ 0 \end{bmatrix} = \alpha_1 x_1 + \alpha_2 x_2,$$

where $\alpha_1, \alpha_2$ are arbitrary constants. Then $K = \text{span} \{x_1, x_2\}$ and $\dim K = 2$. These results are summarized, schematically, in Fig. 3.

![Figure 3. The transformation $F$.](image)

Just as linearity, or the absence of it, was crucial in the theory of ordinary differential equations, it is likewise crucial here. We distinguish transformations as linear or nonlinear as follows.

**DEFINITION 10.8.1 Linear Transformation**

We say that $F : V \rightarrow W$ is a **linear** transformation if

$$F(\alpha u + \beta v) = \alpha F(u) + \beta F(v)$$

for every choice of vectors $u, v$ in $V$, and scalars $\alpha, \beta$; otherwise, $F$ is said to be **nonlinear**.
EXAMPLE 3. To illustrate, consider the transformation $F : \mathbb{R}^3 \to \mathbb{R}^2$, where
\[
F(x) = \begin{bmatrix}
x_1 - 2x_2 + x_3 \\
x_2 + 5x_3
\end{bmatrix}.
\] (7)
Let us see if $F$ is linear.
\[
F(\alpha u + \beta v) = \begin{bmatrix}
(\alpha u_1 + \beta v_1) - 2(\alpha u_2 + \beta v_2) + (\alpha u_3 + \beta v_3) \\
(\alpha u_2 + \beta v_2) + 5(\alpha u_3 + \beta v_3)
\end{bmatrix}
= \alpha \begin{bmatrix}
u_1 - 2u_2 + u_3 \\
u_2 + 5u_3
\end{bmatrix} + \beta \begin{bmatrix}
v_1 - 2v_2 + v_3 \\
v_2 + 5v_3
\end{bmatrix}
= \alpha F(u) + \beta F(v)
\] (8)
for every choice of $u, v, \alpha, \beta$, so $F$ is indeed linear. Notice that the key step in (8), the second equality, follows from the definitions of the addition and scalar multiplication of matrices. \(\blacksquare\)

EXAMPLE 4. Consider $F : \mathbb{R}^2 \to \mathbb{R}^2$, where
\[
F(x) = \begin{bmatrix}
x_1^2 \\
x_1 + 2x_2
\end{bmatrix}.
\] (9)
Then
\[
F(\alpha u + \beta v) = \begin{bmatrix}
(\alpha u_1 + \beta v_1)^2 \\
(\alpha u_1 + \beta v_1) + 2(\alpha u_2 + \beta v_2)
\end{bmatrix}
= \alpha \begin{bmatrix}
u_1^2 \\
u_1 + 2u_2
\end{bmatrix} + \beta \begin{bmatrix}
v_1^2 \\
v_1 + 2v_2
\end{bmatrix} + \begin{bmatrix}
(\alpha^2 - \alpha)u_1^2 + (\beta^2 - \beta)v_1^2 + 2\alpha\beta u_1v_1 \\
0
\end{bmatrix}
= \alpha F(u) + \beta F(v) + \text{deviation},
\] (10)
where the "deviation."
\[
F(\alpha u + \beta v) - \alpha F(u) - \beta F(v) = \begin{bmatrix}
(\alpha^2 - \alpha)u_1^2 + (\beta^2 - \beta)v_1^2 + 2\alpha\beta u_1v_1 \\
0
\end{bmatrix}
\] (11)
is obviously not zero for all choices of $\alpha, \beta, u, v$. For instance, if $\alpha = \beta = u_1 = v_1 = 1$, then (for any $u_2$ and $v_2$) the deviation vector is $[2, 0]^T$. Thus, (6) does not hold for all choices of $\alpha, \beta, u, v$, so $F$ is nonlinear. \(\blacksquare\)

If $F$ is linear, then besides (6) we have
\[
F(\alpha u + \beta v + \gamma w) = F(\alpha u + (\beta v + \gamma w))
= \alpha F(u) + F(\beta v + \gamma w) = \alpha F(u) + \beta F(v) + \gamma F(w)
\]
or, more generally,
\[
F(\alpha_1 u_1 + \cdots + \alpha_n u_n) = \alpha_1 F(u_1) + \cdots + \alpha_n F(u_n).
\] (12)
Observe that (7) can be expressed in matrix notation as
\[
F(x) = \begin{bmatrix} 1 & -2 & 1 \\ 0 & 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.
\] (13)
That is, the action of \( F \) on \( x \) is equivalent to multiplication by \( A \), where \( A \) is the \( 2 \times 3 \) matrix in (13),
\[
F(x) = Ax.
\] (14)
Note that we do not say that "the transformation \( F \) is the matrix \( A \)." Rather, we say that \( F \) is the transformation \textit{multiplication} by \( A \). For that reason we call \( F \) a matrix transformation. The point that we wish to make here is that the linear transformation given in Example 3 happens to be a matrix transformation. We now show that that correspondence is no accident.

**THEOREM 10.8.1 Matrix Transformation**

A transformation \( F : \mathbb{R}^n \to \mathbb{R}^m \) is linear if and only if it is a matrix transformation.

\[ Proof: \] First, we show that if \( F \) is a matrix transformation [i.e., if there is an \( m \times n \) matrix \( A \) such that \( F(x) = Ax \) for each \( x \) in \( \mathbb{R}^n \)], then \( F \) is linear. That is easy since
\[
F(\alpha u + \beta v) = A(\alpha u + \beta v) = \alpha Au + \beta Av = \alpha F(u) + \beta F(v)
\]
for all \( u, v \) in \( \mathbb{R}^n \) and for all scalars \( \alpha, \beta \). To prove the converse, let \( \{i_1, \ldots, i_n\} \) and \( \{j_1, \ldots, j_m\} \) be bases for \( \mathbb{R}^n \) and \( \mathbb{R}^m \), respectively. We may express
\[
x = \sum_{j=1}^{n} x_j i_j \quad \text{and} \quad y = \sum_{k=1}^{m} y_k j_k.
\]
Then
\[
F(x) = F \left( \sum_{j=1}^{n} x_j i_j \right) = \sum_{j=1}^{n} x_j F(i_j)
\]
by (12). Since \( F(i_j) \) is in \( \mathbb{R}^m \), it can be expressed in the form
\[
F(i_j) = \sum_{k=1}^{m} a_{kj} j_k
\]
so
\[
F(x) = \sum_{j=1}^{n} x_j \sum_{k=1}^{m} a_{kj} j_k = \sum_{k=1}^{m} \left( \sum_{j=1}^{n} a_{kj} x_j \right) j_k.
\] (15)
But we also have
\[
F(x) = y = \sum_{k=1}^{m} y_k j_k,
\] (16)
and it follows from (15) and (16), and the linear independence of the $j_k$'s, that

$$y_k = \sum_{j=1}^{n} a_{kj}x_j$$

or $y = Ax$, where $A = \{a_{kj}\}$ is $m \times n$, as was to be proved. ■

We now introduce some additional terminology. First, recall that $F$ is understood to be single valued. That is, to each vector $x$ in the domain $V$ of $F$ there corresponds a unique image $F(x)$ in the range $R$ of $F$. If, in addition to each vector in $R$ there corresponds a unique inverse image in $V$, then $F$ is said to be one-to-one. Notice that we do not say “to each vector in $W$ there corresponds a unique inverse image in $V$” since $R$ may not be all of $W$, in which case those vectors which are in $W$ but not in $R$ have no inverse image at all. If $R$ does turn out to be all of $W$, then $F$ is said to be onto; that is, $F$ maps $V$ “onto” $W$ rather than “into” $W$. Finally, if $F$ is both one-to-one and onto, it is said to be invertible for then every vector in $W$ has a unique image in $V$. This inverse transformation, from $W$ onto $V$, is called the inverse of $F$ and is denoted as $F^{-1}$.

**EXAMPLE 5.** Consider the matrix transformation $F$ in Example 2. There, $F(x) = Ax$, with

$$A = \begin{bmatrix} 1 & -2 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 2 & 2 & -1 \end{bmatrix}.$$  \hspace{1cm} (17)

From (5c) we see that $R$ is only the two-dimensional subspace of $W (= \mathbb{R}^3)$ consisting of the plane $c_3 = -2c_2 + c_1 = 0$ so that $F$ is not onto. That result is illustrated schematically in Fig. 3, where $R$ is shown to be only a part of $W$. Furthermore, if $c$ is in $R$ [i.e., if (5c) is satisfied], then (5) yields a nonunique solution for $x$, so $F$ is not one-to-one. Summarizing, $F$ is neither one-to-one nor onto and is therefore not invertible. ■

**EXAMPLE 6.** Consider the matrix transformation $F : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ associated with the matrix

$$A = \begin{bmatrix} 1 & -2 \\ 1 & 1 \\ 2 & -1 \end{bmatrix}.$$  \hspace{1cm} (18)

Applying elementary operations to the system $Ax = c$, namely, to

$$\begin{bmatrix} 1 & -2 \\ 1 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix},$$  \hspace{1cm} (19)

we obtain the equivalent system

$$x_1 - 2x_2 = c_1,$$
$$3x_2 = c_2 - c_1,$$
$$0 = c_3 - c_2 - c_1,$$  \hspace{1cm} (20a,b,c)
from which it is seen that $F$ is not onto (why not?), although it is one-to-one (why?). Thus $F$ is not invertible. 

**Example 7.** Consider the matrix transformation $F : \mathbb{R}^3 \to \mathbb{R}^2$ associated with the matrix

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 0 & -3 \end{bmatrix}. \quad (21)$$

Applying elementary operations to the system $Ax = c$, namely, to

$$\begin{bmatrix} 1 & 1 & 2 \\ 2 & 0 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}, \quad (22)$$

we obtain the equivalent system

$$x_1 + x_2 + 2x_3 = c_1,$$

$$2x_2 + 7x_3 = 2c_1 - c_2,$$

from which it is seen that $F$ is onto (why?), although not one-to-one (why not?). Thus, $F$ is not invertible. 

**Example 8.** Consider the matrix transformation $F : \mathbb{R}^3 \to \mathbb{R}^3$ associated with the matrix

$$A = \begin{bmatrix} 2 & -1 & 1 \\ 0 & 3 & 1 \\ 1 & 1 & 0 \end{bmatrix}. \quad (24)$$

Applying elementary operations to the system $Ax = c$, we obtain

$$2x_1 - x_2 + x_3 = c_1,$$

$$3x_2 + x_3 = c_2,$$

$$-2x_3 = 2c_3 - c_2 - c_1,$$

from which it is seen that $F$ is both onto (so that $F$ is an operator) and one-to-one, and is therefore invertible; completion of the Gauss–Jordan reduction of (25) reveals that the inverse operator $F^{-1}$ is the matrix operator associated with the matrix

$$\begin{bmatrix} 6 & -1 & 2 \\ -6 & 6 & 3 \\ 1 & 1 & -1 \end{bmatrix}, \quad (26)$$

which, of course, is the inverse of the $A$ matrix, $A^{-1}$. 

**Closure.** Recall that we develop $n$–space in Chapter 9, and then generalize the vector space concept in Section 9.6. The role of the present section is analogous in
that here we have generalized the concept of matrix, developed earlier in this chapter, to that of transformations on vector spaces.

Besides establishing the concept, together with the standard mathematical terminology, the key result is given in Theorem 10.8.1, that a transformation \( F : \mathbb{R}^n \rightarrow \mathbb{R}^m \) is linear if and only if it is a matrix transformation.

**EXERCISES 10.8**

1. In general, the effect of a transformation on the input vector varies from one input vector to another. For example, let \( F : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) be a matrix transformation \( F(x) = Ax \), where

\[
A = \begin{bmatrix} 3 & 2 \\ 0 & 1 \end{bmatrix}.
\]

In 2-space the "effect" of a transformation on a nonzero vector \( x \) amounts to the resulting rotation in the plane, and the dilation (i.e., \( ||Ax||/||x|| \)). For the transformation \( F \) given above, show that these effects are as follows for the given input vectors, and notice that the effect of \( F \) varies from one \( x \) to another.

<table>
<thead>
<tr>
<th>Input</th>
<th>( E )ffect of ( F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) ( x = \begin{bmatrix} 1 \ 0 \end{bmatrix} )</td>
<td>rotation = 0 radians, dilation = 3</td>
</tr>
<tr>
<td>(b) ( x = \begin{bmatrix} 0 \ 1 \end{bmatrix} )</td>
<td>rotation ( \approx ) 1.1 radians, dilation ( \approx \sqrt{5} )</td>
</tr>
<tr>
<td>(c) ( e = \begin{bmatrix} 1 \ -1 \end{bmatrix} )</td>
<td>rotation = 0 radians, dilation = 1</td>
</tr>
</tbody>
</table>

2. Determine whether \( F \) is linear or nonlinear by determining whether or not the deviation \( F(\alpha u + \beta v) - \alpha F(u) - \beta F(u) \) is necessarily zero.

(a) \( F : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \), \( F(x) = \begin{bmatrix} x_1^2 \\ x_1 + x_2 \end{bmatrix} \)
(b) \( F : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \), \( F(x) = \begin{bmatrix} 3x_1 \\ x_1 + x_2 \end{bmatrix} \)
(c) \( F : \mathbb{R}^3 \rightarrow \mathbb{R}^2 \), \( F(x) = \begin{bmatrix} x_1x_2 \\ x_3 \end{bmatrix} \)
(d) \( F : \mathbb{R}^2 \rightarrow \mathbb{R}^3 \), \( F(x) = \begin{bmatrix} \sin x_2 \\ 4x_1 \end{bmatrix} \)
(e) \( F : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \), \( F(x) = \begin{bmatrix} x_1 + 1 \\ x_2 + 1 \end{bmatrix} \)

3. (Identity operator) We say that \( I : V \rightarrow V \) is an identity operator if \( I(x) = x \) for each \( x \) in \( V \).

(a) Suppose that \( F : V \rightarrow V \) is linear and that \( \{v_1, \ldots, v_n\} \) is a basis for \( V \). Show that if \( F(v_1) = v_1, \ldots, F(v_n) = v_n \), then \( F = I \).
(b) Determine the matrix \( A \) corresponding to the identity operator \( I : \mathbb{R}^n \rightarrow \mathbb{R}^n \), i.e., such that \( I(x) = Ax \).

4. (Zero transformation) We say that \( \Phi : V \rightarrow W \) is a zero transformation if \( \Phi(x) = 0 \) for all \( x \)'s in \( V \).

(a) Suppose that \( F : V \rightarrow W \) is linear and that \( \{v_1, \ldots, v_n\} \) is a basis for \( V \). Show that if \( F(v_1) = 0, \ldots, F(v_n) = 0 \), then \( F = \Phi \).
(b) Determine the matrix \( A \) corresponding to the zero transformation \( \Phi : \mathbb{R}^n \rightarrow \mathbb{R}^m \), i.e., such that \( \Phi(x) = Ax \).

5. In each case \( F : \mathbb{R}^n \rightarrow \mathbb{R}^m \) is the matrix transformation corresponding to the given \( m \times n \) matrix \( A \). Determine \( \dim R, \dim K, \) and \( \dim(V) \). Is \( F \) onto? One-to-one? Invertible? Explain. Put forward any basis for \( K \) and any basis for \( R \) (if, indeed, they have bases; see Exercise 7 in Section 9.9).

(a) \( \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \) (b) \( \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \) (c) \( \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 2 \end{bmatrix} \) (d) \( \begin{bmatrix} 2 & -1 & 1 \\ 1 & 0 & 3 \end{bmatrix} \) (e) \( \begin{bmatrix} 2 & 1 & 1 \\ 1 & 0 & 3 \end{bmatrix} \) (f) \( \begin{bmatrix} 2 & 1 & 1 \\ 1 & 0 & 3 \end{bmatrix} \)

6. Make up any example of a matrix transformation \( F : \mathbb{R}^n \rightarrow \mathbb{R}^m \) that is one-to-one but not onto, one that is onto but not one-to-one, one that is neither one-to-one nor onto, and one that is both one-to-one and onto. If such an example is impossible, explain why that is so.
Here is a representation of the given text:

7. (Projection operators) Let \( F : \mathbb{R}^3 \rightarrow \mathbb{R}^3 \) be the transformation

\[
F(x) = (x \cdot \hat{v}) \hat{v},
\]

(7.1)

where \( \hat{v} = [v_1, v_2, v_3]^T \) is a prescribed unit vector. In geometric terms, \( F(x) \) is the vector orthogonal projection of \( x \) onto the line of action of \( \hat{v} \), as illustrated below. Hence, \( F \) in (7.1) is known as a projection operator. We now state the problem: Show that \( F \) is linear, so that one can express \( F(x) \) as \( Ax \). Then determine the nine elements of the \( A \) matrix. (They will depend on \( v_1, v_2, v_3 \).)

8. (More about projection operators) Let \( F : \mathbb{R}^3 \rightarrow \mathbb{R}^3 \) be the transformation

\[
F(x) = (x \cdot \hat{v}_1) \hat{v}_1 + (x \cdot \hat{v}_2) \hat{v}_2.
\]

(8.1)

where \( \hat{v}_1, \hat{v}_2 \) are prescribed ON vectors. Then \( F(x) \) is the vector orthogonal projection of \( x \) onto the plane spanned by \( \{\hat{v}_1, \hat{v}_2\} \).

(a) Given that \( \hat{v}_1 = [1, 0, 0]^T \) and \( \hat{v}_2 = [0, 1, 0]^T \), work out \( F(x) \) for \( x = [2, 3, 4]^T \), and draw an informative, labeled picture, analogous to the one shown in Exercise 7.

(b) Show that \( F \) is linear so that one can express \( F(x) \) as \( Ax \). Then determine the nine elements of the \( A \) matrix, in terms of the components \( v_{11}, v_{12}, v_{13} \) of \( \hat{v}_1 \) and \( v_{21}, v_{22}, v_{23} \) of \( \hat{v}_2 \).

9. Show that \( F(\alpha u + \beta v) = \alpha F(u) + \beta F(v) \), in Definition 10.8.1, is equivalent to the two conditions \( F(u + v) = F(u) + F(v) \) and \( F(\alpha u) = \alpha F(u) \).

10. (Reflection about a line) Let \( F : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) reflect any given vector \( x \) about the line \( L \), as shown in the accompanying figure.

\[
F(x) = x + \Delta x,
\]

(13.1)

where \( \Delta x = [x, y, z]^T \), is the position vector to the point and \( \Delta X = [\Delta x, \Delta y, \Delta z]^T \) is the translation. However, whereas it is convenient, in the computer software, to express all translations and rotations as matrix transformations, the operator \( F : \mathbb{R}^3 \rightarrow \mathbb{R}^3 \) is not linear (Exercise 12), and hence not expressible as a matrix transformation. To circumvent this difficulty we define \( X = [x, y, z, 1]^T \), instead, where the fourth
component, unity, is included for convenience. Then we can express
\[ F(X) = TX \]
\[
\begin{bmatrix}
1 & 0 & 0 & \Delta x \\
0 & 1 & 0 & \Delta y \\
0 & 0 & 1 & \Delta z \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
z \\
1
\end{bmatrix}
= \begin{bmatrix} x + \Delta x \\ y + \Delta y \\ z + \Delta z \\ 1 \end{bmatrix},
\]
which, if we pay attention to only the first three components, effects the translation by multiplication by \( T \), where \( T \) is the \( 4 \times 4 \) matrix in (13.2).

(a) Show that
\[
F(X) = R_zX = \begin{bmatrix}
c_z & -s_z & 0 & 0 \\
s_z & c_z & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
z \\
1
\end{bmatrix}
\]
effects a rotation about the \( z \) axis through an angle \( \theta_z \), taken according to the right-hand rule, where \( c_z, s_z \) are shorthand for \( \cos \theta_z, \sin \theta_z \), respectively. HINT: Letting \( x = r \cos \theta, y = r \sin \theta \), show that
\[
F(X) = R_zX = \begin{bmatrix}
r \cos (\theta + \theta_z) \\
r \sin (\theta + \theta_z) \\
0 \\
1
\end{bmatrix}.
\]

NOTE: Similarly, rotations about the \( x \) axis through an angle \( \theta_x \), and about the \( y \) axis through an angle \( \theta_y \), are effected by
\[
F(X) = R_yX = \begin{bmatrix}
c_y & 0 & -s_y & 0 \\
0 & 1 & 0 & 0 \\
s_y & 0 & c_y & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
z \\
1
\end{bmatrix}
\]
and
\[
F(X) = R_xX = \begin{bmatrix}
c_x & -s_x & 0 & 0 \\
s_x & c_x & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
z \\
1
\end{bmatrix}
\]
respectively, where \( c_x, s_x, c_y, s_y \) denote \( \cos \theta_x, \sin \theta_x, \cos \theta_y, \sin \theta_y \), respectively.

(b) Show that a rotation about the \( z \) axis, followed by a translation, is effected by the composite transformation (see Exercise 11)
\[
F(X) = TR_zX = \begin{bmatrix}
c_z & -s_z & 0 & \Delta x \\
s_z & c_z & 0 & \Delta y \\
0 & 0 & 1 & \Delta z \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
z \\
1
\end{bmatrix}.
\]

Does the order of the operations matter? That is, is \( TR_z = R_zT \)?

(c) Compute \( F(X) = TR_zR_yR_xX \) for \( X = [1, 1, 0, 1]^T \),
\[
\theta_x = -\pi/4, \theta_y = \pi/2, \theta_z = \pi, \Delta x = 2, \Delta y = 1, \Delta z = -3.
\]
Verify the result by drawing the coordinate axes, identifying the initial point \( X \), and then carrying out each rotation and translation graphically in a neat sketch.

(d) Let the point and eraser of a pencil be located, initially, by \( X_p = [0, 1, 1, 1]^T \) and \( X_e = [0, 1, 0, 1]^T \), respectively. Locate the point and the eraser following the composite transformation
\[
F(X) = TR_zR_yR_xX,
\]
where \( \theta_x = 0.2, \theta_y = 0.3, \theta_z = -0.6, \Delta x = 1, \Delta y = 3, \Delta z = 1 \). In a neat sketch, show the pencil in its initial and final configurations, and verify that its length has remained the same.

(e) Repeat part (d), with \( \theta_x = \pi/2, \theta_y = 0, \theta_z = -\pi, \Delta x = \Delta y = \Delta z = 1 \).

(f) Repeat part (d), with \( \theta_x = -\pi/2, \theta_y = \theta_z = \pi/2, \Delta x = \Delta y = 1, \Delta z = 0 \).

---

Chapter 10 Review

The following review is limited to a number of isolated results and formulas that should be both understood and memorized.
Matrix multiplication:
\[ AB \neq BA \quad \text{(in general)} \]

Transpose:
\[ (AB)^T = B^T A^T \]

Determinants:
\[
\begin{align*}
\det(\alpha A + \beta B) &\neq \alpha \det A + \beta \det B \quad \text{(in general)} \\
\det(AB) &\neq (\det A)(\det B)
\end{align*}
\]

Rank:
\[ r(A) = \text{number of LI columns in } A = \text{number of LI rows in } A \]

Systems of linear algebraic equations, \(Ax = c\) (where \(A\) is \(m \times n\)):

Inconsistent:
No solution if \(r(A|c) \neq r(A)\)

Consistent:
Unique solution if \(r(A|c) = r(A) = n\)

\((n - r)\)-parameter family of solutions if \(r(A|c) = r(A) = r < n\)

The case where \(m = n\):
Unique solution \(x = A^{-1}c\) if and only if \(\det A \neq 0\) \([\text{i.e., } r(A) = n]\)

Inverse matrix, \(A^{-1}\):
Exists, and is unique, if and only if \(\det A \neq 0\).
\[
A^{-1}A = AA^{-1} = I \\
(AB)^{-1} = B^{-1}A^{-1} \\
(A^{-1})^T = (A^T)^{-1}
\]

Orthogonal matrices:
A matrix \(Q\) is orthogonal if it is square and its columns are ON.
Chapter 11

The Eigenvalue Problem

11.1 Introduction

In this chapter we study the problem

$$\mathbf{Ax} = \lambda \mathbf{x},$$

(1)

where \( \mathbf{A} \) is a given \( n \times n \) matrix, \( \mathbf{x} \) is an unknown \( n \times 1 \) vector, and \( \lambda \) is an unknown scalar. If we re-express (1) as \( \mathbf{Ax} = \lambda \mathbf{Ix} \) (where \( \mathbf{I} \) is an \( n \times n \) identity matrix), then subtraction of \( \lambda \mathbf{Ix} \) from both sides gives the equivalent equation*

$$ (\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0}, $$

(2)

which is a homogeneous system of \( n \) equations in the \( n \) unknown \( x_j \)'s, where the coefficient matrix \( \mathbf{A} - \lambda \mathbf{I} \) contains the parameter \( \lambda \).

To be sure that (1) and (2) are clear, let us write them out in scalar form, for \( n = 3 \), for example. Then (1) is the system

\[
\begin{align*}
  a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= \lambda x_1, \\
  a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= \lambda x_2, \\
  a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= \lambda x_3.
\end{align*}
\]

Subtracting the terms on the right from those on the left gives

\[
\begin{align*}
  (a_{11} - \lambda)x_1 + a_{12}x_2 + a_{13}x_3 &= 0, \\
  a_{21}x_1 + (a_{22} - \lambda)x_2 + a_{23}x_3 &= 0, \\
  a_{31}x_1 + a_{32}x_2 + (a_{33} - \lambda)x_3 &= 0,
\end{align*}
\]

which, in matrix form, is equation (2).

*Of course we don't need to insert the \( \mathbf{I} \). We could re-express (1), correctly, as \( \mathbf{Ax} - \lambda \mathbf{x} = \mathbf{0} \), but it would not follow from the latter that \( (\mathbf{A} - \lambda)\mathbf{x} = \mathbf{0} \) because subtraction of a scalar \( (\lambda) \) from a matrix \( (\mathbf{A}) \) is not defined. Hence the need to insert \( \mathbf{I} \).
Chapter 11. The Eigenvalue Problem

From Chapter 10, we know that (2) is consistent because it necessarily admits the "trivial" solution \( x = 0 \). However, our interest in (2) shall be in the search for nontrivial solutions, and we anticipate that whether or not nontrivial solutions exist will depend upon the value of \( \lambda \). Thus, the problem of interest is as follows: given the \( n \times n \) matrix \( A \), find the value(s) of \( \lambda \) (if any) such that (2) admits nontrivial solutions, and find those nontrivial solutions. The latter is called the eigenvalue problem and is the focus of this chapter. The \( \lambda \)'s that lead to nontrivial solutions for \( x \) are called the eigenvalues (or characteristic values), and the corresponding nontrivial solutions for \( x \) are called the eigenvectors (or characteristic vectors).

The eigenvalue problem (1) [or, equivalently, (2)] occurs in a wide variety of applications such as vibration theory, chemical kinetics, stability of equilibria, buckling of structures, convergence of iterative techniques, and systems of coupled ordinary differential equations. To place the eigenvalue problem (1) in perspective, recall that in Chapter 8 and 10 we studied the problem

\[
Ax = c
\]

of \( m \) linear algebraic equations in \( n \) unknowns (i.e., \( A \) was \( m \times n \)). In general, \( c \neq 0 \), in which case (3) was said to be nonhomogeneous. The eigenvalue problem is by no means unrelated to (3); it amounts to a special case, where \( c = 0 \) (i.e., it is homogeneous), where \( m = n \), and where the coefficient matrix "\( A \)" = \( A - \lambda I \) contains the parameter \( \lambda \). Thus, to solve the eigenvalue problem we will be able to use results already established in Chapter 10.

### 11.2 Solution Procedure and Applications

#### 11.2.1. Solution and Applications

The eigenvalue problem

\[
(A - \lambda I)x = 0
\]

has the unique trivial solution \( x = 0 \) if \( \det(A - \lambda I) \neq 0 \), and nontrivial solutions (in addition to the trivial solution) if and only if

\[
\det(A - \lambda I) = 0.
\]

The latter is not a vector or matrix equation: it is an algebraic equation in \( \lambda \), known as the characteristic equation corresponding to the matrix \( A \), and its left-hand side is an \( n \times n \) polynomial known as the characteristic polynomial. According to the fundamental theorem of algebra, such an equation has precisely \( n \) roots in the complex plane. Since one or more of these roots can be repeated, we can say that there is at least one eigenvalue \( \lambda \), and at most \( n \) distinct eigenvalues \( \lambda \), corresponding to any given \( n \times n \) matrix \( A \).

As in Chapter 10, we continue to consider only real matrices. However, even if \( A \) is real (so that the coefficients of the characteristic polynomial are too), the characteristic equation can still have complex roots. That case will not be very
important to us. Thus, we avoid it entirely in Chapter 11, and consider it briefly in
Chapter 12.
This is not the first time we have run into the need to solve polynomial equa-
tions. In Section 3.4 we sought solutions to linear, homogeneous, constant-coefficient
differential equations by seeking \( y(x) = e^{\lambda x} \). Putting that solution form into the
\( n \)th order differential equation gave an \( n \)th degree polynomial equation on \( \lambda \). In
fact, even the terminology was the same: the equation was called the characteristic
equation of the differential equation, and the \( n \)th degree polynomial was called the
characteristic polynomial. If \( n = 2 \) we can solve the characteristic equation by
the quadratic formula. For larger \( n \)'s we can, if necessary, use computer software such as the Maple
\texttt{fsolve} command discussed in Section 3.4. Thus, let us consider (2) to have been solved
for the eigenvalues, for the moment, and let us designate them as \( \lambda_1, \ldots, \lambda_k \) (\( 1 \leq k \leq n \)).
Next, set \( \lambda = \lambda_1 \) in (1). Since \( \det(A - \lambda_1 I) = 0 \), it is guaranteed that
\( (A - \lambda_1 I)x = 0 \) will have nontrivial solutions. We can find those solutions by Gauss
elimination, and we designate them as \( e_1 \), where the letter \( e \) is for eigenvector. The
\( e_1 \) solution space is called the eigenspace corresponding to the eigenvalue \( \lambda_1 \).
Next, we set \( \lambda = \lambda_2, \ldots, \lambda_k \) and repeat the process until the \( k \) eigenspaces have
been found.

**EXAMPLE 1.** Determine all eigenvalues and eigenspaces of

\[
A = \begin{bmatrix}
2 & 2 & 1 \\
1 & 3 & 1 \\
1 & 2 & 2
\end{bmatrix}.
\] (3)

The characteristic equation is

\[
det(A - \lambda I) = \begin{vmatrix}
2 - \lambda & 2 & 1 \\
1 & 3 - \lambda & 1 \\
1 & 2 & 2 - \lambda
\end{vmatrix} = \lambda^3 - 7\lambda^2 + 11\lambda - 5
\]

\[
= (\lambda - 5)(\lambda - 1)^2 = 0
\] (4)

so the eigenvalues of \( A \) are \( \lambda_1 = 5 \) and \( \lambda_2 = 1 \) (or vice versa since the order is immaterial),
with \( \lambda_2 \) called a repeated eigenvalue — specifically, an eigenvalue of multiplicity 2
because it is a double root of the characteristic equation (4).

Next, find the eigenspaces.

\( \lambda_1 = 5 \) : Then \( (A - \lambda_1 I)x = 0 \) becomes

\[
\begin{bmatrix}
2 - 5 & 2 & 1 \\
1 & 3 - 5 & 1 \\
1 & 2 & 2 - 5
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix}
= \begin{bmatrix}
-3 & 2 & 1 \\
1 & -2 & 1 \\
1 & 2 & -3
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix},
\] (5)

*Note that the eigenspace corresponding to an eigenvalue \( \lambda_j \) is not quite the same as the set
of eigenvectors corresponding to \( \lambda_j \); it is that set plus the trivial solution (which is NOT itself an
eigenvector). The reason we define the eigenspace corresponding to \( \lambda_j \) as the entire solution space
of \( (A - \lambda_j I)x = 0 \) (i.e., including the zero solution) is so that the eigenspace will be a vector space,
for recall that a vector space must contain a zero vector.
Gauss elimination of which gives
\[
\begin{bmatrix}
-3 & 2 & 1 \\
0 & 1 & -1 \\
0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix}.
\] (6)

The solution is \(x_3 = \alpha\) (arbitrary), \(x_2 = \alpha\), \(x_1 = \alpha\) so, using \(e\) in place of \(x\),
\[
e = \begin{bmatrix}
\alpha \\
\alpha \\
\alpha
\end{bmatrix} = \alpha \begin{bmatrix}
1 \\
1 \\
1
\end{bmatrix}.
\] (7)

Thus, the eigenspace corresponding to \(\lambda_1 = 5\) is \(\text{span}\{1, 1, 1\}^T\), the latter being a one-dimensional subspace of \(\mathbb{R}^3\), namely, the line through the origin given by (7).

\(\lambda_2 = 1\): Then \((A - \lambda_2 I)x = 0\) becomes
\[
\begin{bmatrix}
2 - 1 & 2 & 1 \\
1 & 3 - 1 & 1 \\
1 & 2 & 2 - 1
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix}
= \begin{bmatrix}
1 & 2 & 1 \\
1 & 2 & 1 \\
1 & 2 & 1
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix}.
\] (8)

Gauss elimination of which gives
\[
\begin{bmatrix}
1 & 2 & 1 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix}.
\] (9)

The solution is \(x_3 = \gamma\) (arbitrary), \(x_2 = \gamma\) (arbitrary), \(x_1 = -\beta - 2\gamma\) so
\[
e = \begin{bmatrix}
-\beta - 2\gamma \\
\gamma \\
\beta
\end{bmatrix} = \beta \begin{bmatrix}
-1 \\
0 \\
1
\end{bmatrix} + \gamma \begin{bmatrix}
-2 \\
1 \\
0
\end{bmatrix}.
\] (10)

Thus, the eigenspace corresponding to \(\lambda_2 = 1\) is \(\text{span}\{-1, 0, 1\}^T, [-2, 1, 0]^T\}, the latter being a two-dimensional subspace of \(\mathbb{R}^3\), namely, the plane through the origin, spanned by \([-1, 0, 1]^T\) and \([-2, 1, 0]^T\). In fact, the equation of that plane is seen, in (9), as \(x_1 + 2x_2 + x_3 = 0\).

COMMENT 1. We can determine eigenvectors only up to arbitrary scale factors (such as the \(\alpha\) in (7)) because, if a vector \(e\) satisfies \(Ae = \lambda e\), then so does any scalar multiple of \(e\). Along these lines, observe that it would be correct to write \(e = \beta[1, 0, -1]^T + \gamma[-2, 1, 0]^T\), say, since the scale factor of \(-1\) in the first vector can be absorbed by the arbitrary \(\beta\).

COMMENT 2. In the language of Section 10.5, the rank of the \(A - \lambda_1 I\) coefficient matrix in (6) is 2 so \(n - r = 3 - 2 = 1\), and (6) admits a one-parameter family of solutions. That is, the nullity of \(A - \lambda_1 I\) is 1 and the \(e_1\) eigenspace is one-dimensional. Similarly, the rank of \(A - \lambda_2 I\) in (9) is 1, so \(n - r = 3 - 1 = 2\), and (9) admits a two-parameter family of solutions. That is, the nullity of \(A - \lambda_2 I\) is 2, and the \(e_2\) eigenspace is two-dimensional. However, there is no reason to believe that the multiplicity of an eigenvalue necessarily
equals the dimension of the corresponding eigenspace, even though it happens to be true in this example.

EXAMPLE 2. The matrix
\[ A = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \] (11)

has the characteristic equation
\[ \det(A - \lambda I) = \begin{vmatrix} 1 - \lambda & 0 & 1 \\ 1 & 1 - \lambda & 0 \\ 0 & 0 & 1 - \lambda \end{vmatrix} = (1 - \lambda)^3 = 0 \] (12)

with roots \( \lambda = 1, 1, 1 \). That is, \( \lambda_1 = 1 \) is a root of multiplicity three. To find the eigenspace, write out \((A - \lambda_1 I)x = 0\) as
\[ \begin{bmatrix} 1 - 1 & 0 & 1 \\ 1 & 1 - 1 & 0 \\ 0 & 0 & 1 - 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \] (13)

The solution is \( x_3 = 0, x_1 = 0, x_2 = \alpha \) (arbitrary) so
\[ e = \begin{bmatrix} 0 \\ \alpha \\ 0 \end{bmatrix} = \alpha \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \] (14)

Thus, in this case an eigenvalue of multiplicity three gave rise to an eigenspace of dimension one.

With the "mechanics" of the eigenvalue problem explained in the first two examples, let us devote the next examples to applications.

EXAMPLE 3. Solution of Differential Equations. How can we solve the coupled differential equations
\[ x' = x + 4y, \]
\[ y' = x + y \] (15)
on \( x(t) \) and \( y(t) \)? We could use the method of elimination (Section 3.9) to uncouple them, or solve by the Laplace transform method (Chapter 5). Here, we pursue a different approach, that will lead to an eigenvalue problem.

Since (15) is linear, constant-coefficient, and homogeneous, we can find exponential solutions. Thus, seek \( x, y \) in the form
\[ x(t) = q_1 e^{rt}, \quad y(t) = q_2 e^{rt} \] (16)

where \( q_1, q_2, r \) are constants that are to be determined. Putting (16) into (15) gives
The Eigenvalue Problem

\[ rq_1e^{rt} = q_1e^{rt} + 4q_2e^{rt}, \]
\[ rq_2e^{rt} = q_1e^{rt} + q_2e^{rt}, \]

or, cancelling the \( e^{rt} \)'s (because they are nonzero) and expressing the result in matrix form, gives

\[
\begin{bmatrix}
1 & 4 \\
1 & 1
\end{bmatrix}
\begin{bmatrix}
q_1 \\
q_2
\end{bmatrix} = r
\begin{bmatrix}
q_1 \\
q_2
\end{bmatrix}
\]
or

\[ Aq = rq, \]

which is an eigenvalue problem with \( \lambda = r \).

We are not interested in the trivial solution \( q = 0 \) because it gives the trivial particular solution \( x(t) = y(t) = 0 \), of (15), whereas we seek the general solution.

Proceeding as above, we obtain these eigenvalues and eigenspaces:

\[ \lambda_1 = 3, \quad e_1 = \alpha \begin{bmatrix} 2 \\ 1 \end{bmatrix}; \quad \lambda_2 = -1, \quad e_2 = \beta \begin{bmatrix} -2 \\ 1 \end{bmatrix}. \]

Denoting \( x(t) = [x(t), y(t)]^T \), each "eigenpair" gives a solution of (15) as

\[ x(t) = \alpha \begin{bmatrix} 2 \\ 1 \end{bmatrix} e^{3t} \quad \text{and} \quad x(t) = \beta \begin{bmatrix} -2 \\ 1 \end{bmatrix} e^{-t}. \]

By the linearity of (15), we can superimpose these solutions and thereby obtain the general solution

\[ x(t) = \alpha \begin{bmatrix} 2 \\ 1 \end{bmatrix} e^{3t} + \beta \begin{bmatrix} -2 \\ 1 \end{bmatrix} e^{-t}. \]

or, in scalar form,

\[ x(t) = 2\alpha e^{3t} - 2\beta e^{-t}, \]
\[ y(t) = \alpha e^{3t} + \beta e^{-t}. \]

Of course, \( \alpha \) and \( \beta \) are the integration constants (usually denoted as \( A, B \) or \( C_1, C_2 \) in the ODE chapters).

COMMENT 1. Since we use \( q_1 \) and \( q_2 \) in (16), it would be natural to wonder why we don't also allow for different exponents, and seek \( x(t) = q_1 \exp (r_1 t) \) and \( y(t) = q_2 \exp (r_2 t) \). The reason is that unless \( r_1 = r_2 \) we obtain only the trivial solution \( q_1 = q_2 = 0 \) (Exercise 1).

COMMENT 2. The method illustrated in this example can be used for any system of coupled, linear, constant-coefficient homogeneous differential equations. However, it will fail to produce a general solution if \( A \) has a repeated eigenvalue of multiplicity \( k \) if the dimension of the corresponding eigenspace is less than \( k \).

**EXAMPLE 4. Markov Process.** Suppose that there is a population exchange between Delaware, Maryland, and Pennsylvania such that, each year, 20% of Delaware's residents move to Maryland and 8% move to Pennsylvania; 12% of Maryland's residents move to Pennsylvania every year. Each year, 30% of Pennsylvania's residents move to Maryland. If we define the states as:

- State 1: Delaware's residents
- State 2: Maryland's residents
- State 3: Pennsylvania's residents

Then the transition matrix \( A \) is given by:

\[
A = \begin{bmatrix}
-0.8 & 0.2 & 0.12 \\
-0.3 & 0.7 & 0.88 \\
0.3 & 0.2 & 0.8
\end{bmatrix}
\]

We can use the method described above to find the general solution for the population exchange over time.
11.2. Solution Procedure and Applications

Delaware and 10% to Pennsylvania; 10% of Pennsylvania’s residents move to Delaware and 3% move to Maryland. For simplicity, let us ignore gains in population due to births and losses due to deaths — or, equivalently, suppose that these effects are nonzero, but equal and opposite, so as to cancel. Further, let us suppose that the three states are a closed system; that is, they exchange populations only among themselves.

If we denote the populations at the end of the nth year, in DE, MD, and PA as \( x_n, y_n, z_n \), respectively, then (see Fig. 1)

\[
\begin{align*}
x_{n+1} &= x_n - (0.2 + 0.08)x_n + 0.12y_n + 0.1z_n, \\
y_{n+1} &= y_n + 0.2x_n - (0.12 + 0.1)y_n + 0.03z_n, \\
z_{n+1} &= z_n + 0.08x_n + 0.1y_n - (0.1 + 0.03)z_n,
\end{align*}
\]

(23)

which are coupled difference equations. Difference equations were studied in Section 6.5.3 within the context of differential equations. Here, however, let us consider (23) as a matrix equation,

\[
\begin{bmatrix}
x_{n+1} \\
y_{n+1} \\
z_{n+1}
\end{bmatrix} =
\begin{bmatrix}
0.72 & 0.12 & 0.1 \\
0.2 & 0.78 & 0.03 \\
0.08 & 0.1 & 0.87
\end{bmatrix}
\begin{bmatrix}
x_n \\
y_n \\
z_n
\end{bmatrix}
\]

(24)

or,

\[ p_{n+1} = A p_n, \]

(25)

where \( p_n = [x_n, y_n, z_n]^T \) is the “population vector.”

The first problem that we pose is to find the population \( p_n \) as a function of \( n \), \( n \) being essentially a discrete time variable, given some initial population \( p_0 \). That’s easy, because (25) gives \( p_1 = A p_0, p_2 = A p_1 = A(A p_0) = A^2 p_0, p_3 = A p_2 = A(A^2 p_0) = A^3 p_0 \), and so on.

\[ p_n = A^n p_0. \]

(26)

We wonder whether \( A^n p_0 \) keeps changing as \( n \) increases, or whether it settles down and approaches an equilibrium (or steady-state) vector, say \( P \). If there is such an equilibrium vector then, by definition of equilibrium, \( p_{n+1} = p_n = P \), so (25) becomes \( P = A P \). Surely, \( P = 0 \) satisfies

\[ A P = P, \]

(27)

but the interesting question is whether or not there exist nontrivial \( P \)’s. In fact, (27) is an eigenvalue problem with \( \lambda = 1 \), so we can say that nontrivial equilibrium vectors exist if and only if 1 is an eigenvalue of \( A \). As explained at the end of this section, we can use Maple to obtain the following eigenvalues and eigenspaces of \( A \):

\[
\begin{align*}
\lambda_1 &= 1, \quad e_1 = \alpha \\
&\begin{bmatrix}
-0.56 \\
-0.62 \\
-0.82
\end{bmatrix} ; \\
\lambda_2 &= 0.77, \quad e_2 = \beta \\
&\begin{bmatrix}
-0.09 \\
-0.70 \\
0.79
\end{bmatrix} ; \\
\lambda_3 &= 0.60, \quad e_3 = \gamma \\
&\begin{bmatrix}
-0.62 \\
0.70 \\
-0.07
\end{bmatrix} .
\end{align*}
\]

(28)

(Actually, Maple gave the \( \lambda_j \)'s and \( e_j \)'s to nine and ten significant figures, respectively, but we have rounded off for brevity.)
Sure enough, $\lambda = 1$ is among the eigenvalues of $A$ so there is an equilibrium population vector $P$ given by the corresponding eigenvector

$$P = \alpha \begin{bmatrix} 0.56 \\ 0.62 \\ 0.82 \end{bmatrix}, \quad (29)$$

where we have absorbed a factor of $-1$ into the scale factor $\alpha$, to avoid the appearance of negative populations. If desired, we can compute $\alpha$ by conserving the total population:

$$0.56\alpha + 0.62\alpha + 0.82\alpha = x_0 + y_0 + z_0,$$

so $\alpha = (x_0 + y_0 + z_0)/2.0.$

Finally, it is important to determine whether or not the equilibrium is stable for it will be observed only if it is stable, just as marbles are found in valleys but not on hilltops. To address the question of stability, let us use the set of LI eigenvectors $\{e_1, e_2, e_3\}$, with any nonzero values of $\alpha, \beta$, and $\gamma$, as a basis for $\mathbb{R}^3$, and expand the initial vector $p_0$ in terms of that basis as

$$p_0 = c_1 e_1 + c_2 e_2 + c_3 e_3. \quad (30)$$

Then (25) gives

$$p_1 = Ap_0 = c_1 A e_1 + c_2 A e_2 + c_3 A e_3$$

$$= c_1 \lambda_1 e_1 + c_2 \lambda_2 e_2 + c_3 \lambda_3 e_3, \quad (25)$$

$$p_2 = Ap_1 = c_1 \lambda_1 A e_1 + c_2 \lambda_2 A e_2 + c_3 \lambda_3 A e_3$$

$$= c_1 \lambda_1^2 e_1 + c_2 \lambda_2^2 e_2 + c_3 \lambda_3^2 e_3, \quad (31)$$

$$\vdots$$

$$p_n = Ap_{n-1} = c_1 \lambda_1^n e_1 + c_2 \lambda_2^n e_2 + c_3 \lambda_3^n e_3$$

$$= c_1 e_1 + c_2 (0.77)^n e_2 + c_3 (0.60)^n e_3$$

$$\sim c_1 e_1 \quad (31)$$

as $n \to \infty$, provided that $c_1 \neq 0$. In fact, $c_1$ cannot be zero because if it were zero then (31) would give $p_n \to 0$ as $n \to \infty$ and, since the total population is conserved, that could happen only in the uninteresting case where $p_0 = 0$. Thus, we see from (31) that $p_n$ inevitably tends to a multiple of $e_1$, namely, to the equilibrium vector $P$.

The upshot is that the population history is given by (26), and that $p_n$ inevitably tends to a unique steady state which is some scalar multiple of $e_1$, the multiple being fixed by the conservation of the total population.

**COMMENT.** This example incorporates a number of linear algebra concepts: matrix multiplication in expressing (23) compactly as (25) and in deriving the solution (26) for $p_n$; the eigenvalue problem in regard to the possibility of a steady-state solution $P$; and bases and expansions in assessing the stability of that steady state, in (30)-(31). Consider how effective are these linear algebra concepts and methods in providing a systematic approach to solving this problem, especially in determining the stability of the steady state. The same

---

*Recall that we built into (24) the assumption that any births and deaths cancel, in number, as revealed by adding the three scalar equations in (24), for that step gives $x_{n+1} + y_{n+1} + z_{n+1} = x_n + y_n + z_n$.**
approach applies whether the system includes only three states, as in the present example, or 30 states. 

The matrix \( A \) in Example 4 is an example of a “Markov” matrix. An \( n \times n \) matrix \( A = \{a_{ij}\} \) is called a Markov (or stochastic) matrix if \( a_{ij} \geq 0 \) for each \( i, j \), and if the elements of each column sum to unity, or if the elements of each row sum to unity. [In the case of the \( A \) given in (24) its columns sum to unity.] It was no coincidence that \( \lambda = 1 \) was an eigenvalue, in Example 4 since \( \lambda = 1 \) is an eigenvalue of every Markov matrix.

To prove that claim, suppose that \( A \) is a Markov matrix. The value \( \lambda = 1 \) will be among the eigenvalues of \( A \) if and only if \( Ax = \lambda x \) has nontrivial solutions for \( x \) or, equivalently, if the rows or columns of \( A - I \) are linearly dependent. Since \( A \) is a Markov matrix, either the elements of each of its columns sum to unity or the elements of each of its rows sum to unity. It follows that either the elements of each of the columns of \( A - I \) sum to zero (in which case the rows of \( A - I \) are linearly dependent) or the elements of each of the rows of \( A - I \) sum to zero (in which case the columns of \( A - I \) are linearly dependent), or both. Thus, \( A - I \) is singular and our claim is proved.

### 11.2.2. Application to elementary singularities in the phase plane.

If you studied Chapter 7, you will recall the fundamental role of the elementary singularities in the \( x, y \) phase plane, where \( x(t) \) and \( y(t) \) satisfy the linear ODE's

\[
\begin{align*}
x' &= ax + by \\
y' &= cx + dy.
\end{align*}
\]

For instance, the system (15) is of that form, so let us reconsider the result, given by (21) and (22), in terms of the \( x, y \) phase plane. If \( \beta = 0 \), then \( x(t) = 2\alpha e^{3t} \) and \( y(t) = \alpha e^{3t} \), so the phase trajectory is the line \( y = x/2 \), and if \( \alpha = 0 \) then \( x(t) = -2\beta e^{-t} \) and \( y(t) = \beta e^{-t} \), so the phase trajectory is the line \( y = -x/2 \). These are shown in Fig. 2. The directions of the arrows follow from the fact that \( e^{3t} \) increases with \( t \) and \( e^{-t} \) decreases (since the \( \lambda \)'s are of opposite sign), and they imply that (15) has a saddle at the origin.

More generally, observe that we can classify the singularity directly from the eigenvalues of the \[
\begin{bmatrix}
a & b \\
c & d
\end{bmatrix}
\]
matrix in (32). Let \( a, b, c, d \) be real. Then the eigenvalues are either real or they are complex conjugates. In general, then, we can write \( \lambda = \alpha \pm i\beta \), and these \( \lambda \)'s contribute solutions of the form \( e^{(\alpha \pm i\beta)t} \). We have these possibilities:

- \( \lambda \)'s real and of the same sign:
  - \( \lambda < 0 \) \( \Rightarrow \) stable node
  - \( \lambda > 0 \) \( \Rightarrow \) unstable node
- \( \lambda \)'s real and of opposite sign \( \Rightarrow \) saddle

![Figure 2. Significance of the eigenvectors.](image)
Chapter 11. The Eigenvalue Problem

\[ \lambda \text{'s complex } (\lambda = \alpha \pm i\beta) : \]
\[ \alpha < 0 \Rightarrow \text{stable focus} \]
\[ \alpha = 0 \Rightarrow \text{center} \]
\[ \alpha > 0 \Rightarrow \text{unstable focus} \]

In the case of an improper node or saddle, the eigenvectors give the stable and/or unstable manifolds, as mentioned above for the system (15), for which case the stable and unstable manifolds are shown in Fig. 2.

**Closure.** It is important to remember that the eigenvalue problem \( AX = \lambda X \) is homogeneous [since it is equivalent to \( (A - \lambda I)x = 0 \)] and that the whole point is to find nontrivial solutions. If, for a given eigenvalue \( \lambda \), you solve \( (A - \lambda I)x = 0 \) by Gauss elimination and obtain \( e = 0 \), then your calculations are incorrect: either your eigenvalue is incorrect and/or the Gauss elimination is incorrect.

Observe that our solution strategy uncouples the calculation of the eigenvalues and the eigenvectors: first we solve the characteristic polynomial equation \( \det(A - \lambda I) = 0 \) for the \( \lambda \)'s, and then for each \( \lambda \) we solve \( (A - \lambda I)x = 0 \), by Gauss elimination, for the corresponding eigenvectors.

**Computer software.** In Maple, the relevant commands are `eigenvals` and `eigenvects`, both of which are in the `linalg` package. The command `eigenvals` gives just the eigenvalues, and `eigenvects` gives the eigenvalues, their multiplicity, and a basis for each eigenspace. For instance, let \( A \) be the matrix in Example 1. First, enter

```maple
with(linalg):
```
and return. Then type

```maple
A := matrix(3, 3, [2, 2, 1, 1, 3, 1, 1, 2, 2]):
```
because \( A \) is \( 3 \times 3 \) and its rows are 2, 2, 1, and 1, 3, 1, and 1, 2, 2, in turn. Then

```maple
eigenvals(A);
```
gives the eigenvalues as

\[ 5, 1, 1 \]
and

```maple
eigenvects(A);
```
gives both the eigenvalues and the eigenvectors as

\[ [5, 1, \{[1, 1, 1]\}], [1, 2, \{[-2, 1, 0], [-1, 0, 1]\}] \]

In place of the `eigenvals` command, one can use `fsolve` to obtain the roots of the characteristic equation, but that is less convenient since to obtain the characteristic equation one needs to expand the \( n \times n \) determinant of \( A - \lambda I \).
Suppose we want the eigenvalues of $A^{20}$. Enter

$$A := \text{matrix } (3, 3, [2, 2, 1, 1, 3, 1, 1, 2, 2]):$$

then

$$\text{evalm}(A^{20});$$

and then

$$\text{eigenvals}();$$

The quotation mark saves us the trouble of entering the matrix $A^{20}$ that was calculated in the preceding step.

**EXERCISES 11.2**

1. (Example 3) (a) Derive the eigenvalues and eigenspaces given in (19).
   (b) Show that if we assume the forms $x(t) = q_1 \exp (r_1 t)$ and $y(t) = q_2 \exp (r_2 t)$, then we obtain only the trivial solution unless $r_1 = r_2$, as claimed in COMMENT 1.

2. (Example 4) To determine the stability of the equilibrium solution $P$, we expanded $p_0$ in terms of the basis consisting of the eigenvectors of $A$. Explain why that choice is particularly convenient. HINT: If necessary, you could try using a different basis, such as $\{[1, 0, 0]^T, [0, 1, 0]^T, [0, 0, 1]^T\}$, instead.

3. Find the eigenvalues and eigenspaces, as well as a basis for each eigenspace.

   (a) $\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$
   (b) $\begin{bmatrix} 1 & -3 \\ 0 & 0 \end{bmatrix}$
   (c) $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$
   (d) $\begin{bmatrix} -3 & 2 \\ 6 & -4 \end{bmatrix}$
   (e) $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix}$
   (f) $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix}$
   (g) $\begin{bmatrix} 0 & -5 & 0 \\ 0 & 0 & 4 \end{bmatrix}$
   (h) $\begin{bmatrix} 0 & -5 & 0 \\ 0 & 0 & 4 \end{bmatrix}$
   (i) $\begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$
   (j) $\begin{bmatrix} 2 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix}$
   (k) $\begin{bmatrix} 1 & 0 & 2 \\ 1 & 0 & 2 \\ 1 & 0 & 2 \end{bmatrix}$

4. (a)–(r) Use computer software to find the eigenvalues and eigenspaces for the matrix in the corresponding part of Exercise 3.

5. Is the following an eigenvector of the matrix $A$? Explain.

   $$A = \begin{bmatrix} 1 & 8 & 5 & 3 \\ 2 & 16 & 10 & 6 \\ -5 & -14 & -11 & -3 \\ -1 & -8 & -5 & -3 \end{bmatrix}$$

   (a) $\begin{bmatrix} 1, 2, -1, 3 \end{bmatrix}$
   (b) $\begin{bmatrix} 1, 2, -4, -1 \end{bmatrix}$
   (c) $\begin{bmatrix} 1, 2, 1 \end{bmatrix}$
   (d) $\begin{bmatrix} 1, 0, 1, -2 \end{bmatrix}$
   (e) $\begin{bmatrix} 1, 0, 1 \end{bmatrix}$
   (f) $\begin{bmatrix} 1, 1, 0 \end{bmatrix}$
   (g) $\begin{bmatrix} 1, 2, -1, -1 \end{bmatrix}$
   (h) $\begin{bmatrix} 2, 1, 0, 1 \end{bmatrix}$
   (i) $\begin{bmatrix} 2, 1, 1, -5 \end{bmatrix}$

6. The given matrix has $\lambda = 2$ among its eigenvalues. Find the eigenspace corresponding to that eigenvalue.

   $\begin{bmatrix} 3 & 2 & 2 & 2 \\ 2 & 3 & 1 & 2 \\ -1 & -1 & 2 & 0 \\ 2 & 4 & 3 & 5 \end{bmatrix}$
   $\begin{bmatrix} 3 & 1 & 2 & 1 \\ 2 & 3 & 1 & 2 \\ -1 & 3 & 1 & 2 \\ 0 & 2 & 5 & 3 \\ 2 & 4 & 3 & 5 \end{bmatrix}$
7. It is known that the $n \times n$ tridiagonal matrix

$$
A = \begin{bmatrix}
    b & c & 0 & 0 & \cdots & 0 \\
    a & b & c & 0 & \cdots & 0 \\
    0 & a & b & c & \cdots & 0 \\
    \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
    0 & \cdots & \cdots & \cdots & a & b \\
    \cdots & \cdots & \cdots & \cdots & \cdots & a & b \\
\end{bmatrix}
$$

has eigenvalues

$$
\lambda_j = b + 2\sqrt{ac} \cos \left( \frac{j\pi}{n+1} \right) \quad (7.1)
$$

for $j = 1, 2, \ldots, n$. ($A$ is called tridiagonal because all elements are zero except for those on the main diagonal and the two adjacent diagonals.)

(a) Verify (7.1) by calculating the eigenvalues for $n = 1$ and $n = 2$.

(b) Verify (7.1) by using computer software to determine the eigenvalues for $n = 1$ and $2$ and $3$.

(c) Verify (7.1) by using computer software to determine the eigenvalues for $n = 4$ and $a = 1, b = 2, c = 1$.

(d) Same as (c), for $n = 4$ and $a = 2, b = 3, c = -1$.

(e) Same as (c), for $n = 5$ and $a = 1, b = 5, c = 3$.

8. Is it possible for a matrix to have no eigenvalues? Explain.

9. We saw in Example 1 that a given eigenvalue can have more than one LI eigenvector. Can a given eigenvector correspond to more than one eigenvalue? Explain.

10. Let $x$ and $Ax$ be as shown. Is $x$ an eigenvector of $A$? If so, estimate the corresponding eigenvalue; if not, explain why not.

11. Show that the eigenvalues of $kA$, for any scalar $k$, are $k$ times those of $A$. Are the corresponding eigenspaces the same? Explain.

12. Show that the eigenvalues of $A^T$ are the same as those of $A$. Is the eigenspace corresponding to an eigenvalue $\lambda$ of $A$ the same as the eigenspace corresponding to the same eigenvalue $\lambda$ of $A^T$? Prove or disprove.

13. If $\lambda, \mathbf{e}$ are an eigenvalue and corresponding eigenvector of a matrix $A$, show that $\lambda = (\mathbf{e}^T A \mathbf{e}) / (\mathbf{e}^T \mathbf{e})$. HINT: Recall that the dot product of two column vectors $u$ and $v$ is $u \cdot v = u^T v$.

14. Show that if $A$ is triangular, its eigenvalues are simply the diagonal elements of $A$.

15. Show that if $\lambda$ is an eigenvalue of $A$, with a corresponding eigenvector $\mathbf{e}$, then $\lambda^n$ is an eigenvalue of $A^n$, with the same eigenvector $\mathbf{e}$, for any integer $n$. (Of course, if $n$ is negative, $A$ needs to be nonsingular if $A^n$ is to exist in the first place.) HINT: Pre-multiply $A^n \mathbf{e} = \lambda^n \mathbf{e}$ by $A$, $A^2$, ... .

16. Use the results stated in the preceding two exercises to determine the eigenvalues and eigenspaces of $A^{10}$ for each of the following $A$ matrices. Check your results by working out $A^{10}$ and its eigenvalues and eigenvectors.

\[
\begin{align*}
(\text{a}) & \quad \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ 1 & 1 & 3 \end{bmatrix} & \begin{bmatrix} 3 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 2 \end{bmatrix} \\
(\text{b}) & \quad \begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & -2 \end{bmatrix} & \begin{bmatrix} 1 & 1 & 2 \\ 0 & -1 & 2 \\ 0 & 0 & 2 \end{bmatrix}
\end{align*}
\]

17. For the given $A$ matrix, use computer software to determine its eigenvalues and eigenspaces. Then, use computer software to obtain $A^3$ and to determine its eigenvalues and eigenspaces. Then, verify the result stated in Exercise 15, for this case.

\[
\begin{align*}
(\text{a}) & \quad \begin{bmatrix} 2 & 2 & 0 \\ 0 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix} & \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix} \\
(\text{b}) & \quad \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \end{bmatrix}
\end{align*}
\]

18. (Similar matrices) (a) Suppose that $Ax = y$, where $A$ is square. Setting $x = Q \tilde{x}$ and $y = Q \tilde{y}$, where $Q$ is invertible, show that

$$
\tilde{A} \tilde{x} = \tilde{y}, \quad (18.1)
$$

where

$$
\tilde{A} = Q^{-1} AQ. \quad (18.2)
$$
Given any invertible matrix $Q$, matrices $A$ and $\tilde{A}$ related by (18.2) are said to be similar.

(b) Show that if $A$ and $\tilde{A}$ are similar, then they have the same characteristic polynomials and hence the same eigenvalues.

19. (The characteristic polynomial) Let us write the characteristic equation $\det(A - \lambda I) = 0$ in the standard form

$$
\det(A - \lambda I) = \begin{vmatrix}
\lambda - a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & \lambda - a_{22} & \cdots & a_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \cdots & \lambda - a_{nn}
\end{vmatrix}
$$

$$
= (-1)^n [\lambda^n - \beta_1 \lambda^{n-1} + \beta_2 \lambda^{n-2} - \cdots + (-1)^n \beta_n] = 0.
$$

(19.1)

If we denote the $n$ roots (which need not be real) as $\lambda_1, \lambda_2, \ldots, \lambda_n$, numbering repeated roots separately, we may factor

$$
\lambda^n - \beta_1 \lambda^{n-1} + \cdots + (-1)^n \beta_n
\quad = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n).
$$

(19.2)

Multiplying out the right-hand side of (19.2) and equating coefficients of like powers of $\lambda$, on both sides of the equation, yields the relations

$$
\beta_1 = \lambda_1 + \lambda_2 + \cdots + \lambda_n,
$$

$$
\beta_2 = \lambda_1 \lambda_2 + \lambda_1 \lambda_3 + \cdots + \lambda_{n-1} \lambda_n,
$$

$$
\beta_3 = \lambda_1 \lambda_2 \lambda_3 + \lambda_1 \lambda_2 \lambda_4 + \cdots + \lambda_{n-2} \lambda_n \lambda_1,
$$

$$
\vdots
$$

$$
\beta_n = \lambda_1 \lambda_2 \cdots \lambda_n.
$$

(19.3)

For example, if $n = 4$, then $\beta_2 = \lambda_1 \lambda_2 + \lambda_1 \lambda_3 + \lambda_1 \lambda_4 + \lambda_2 \lambda_3 + \lambda_2 \lambda_4 + \lambda_3 \lambda_4$ and $\beta_3 = \lambda_1 \lambda_2 \lambda_3 + \lambda_1 \lambda_2 \lambda_4 + \lambda_1 \lambda_3 \lambda_4 + \lambda_2 \lambda_3 \lambda_4$.

Alternatively, one may expand the determinant in (19.1) directly, and identify the coefficients of the various powers of $\lambda$ in terms of certain subdeterminants of $A$, and hence determine the $\beta_j$'s in terms of these subdeterminants. The result, we state without proof, is that $\beta_j$ (for $j = 1, \ldots, n$) is the sum of all the $j$th-order principal minors of $A$. (By a “principal minor” we mean the determinant of a submatrix of $A$, whose main diagonal lies along that of $A$.) Thus,

$$
\beta_1 = a_{11} + a_{22} + \cdots + a_{nn} \equiv \text{tr} A,
$$

$$
\vdots
$$

$$
\beta_n = \det A,
$$

(19.4)

where the sum $a_{11} + a_{22} + \cdots + a_{nn}$ is called the trace of $A$ and is denoted here as $\text{tr} A$.

(a) Verify equations (19.4) for $n = 2$ by expanding $\det (A - \lambda I)$.

(b) Verify equations (19.4) for $n = 3$ by expanding $\det (A - \lambda I)$; i.e., verify that

$$
\begin{align*}
\beta_1 &= a_{11} + a_{22} + a_{33}, \\
\beta_2 &= \begin{vmatrix}
\begin{vmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{vmatrix} + a_{11} & a_{13} \\
a_{21} & a_{23}
\end{vmatrix} + \begin{vmatrix}
\begin{vmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{vmatrix} & a_{22} & a_{23} \\
a_{31} & a_{32}
\end{vmatrix}, \\
\beta_3 &= \begin{vmatrix}
\begin{vmatrix}
\begin{vmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{vmatrix} & a_{13} \\
a_{31} & a_{32}
\end{vmatrix} & a_{23} \\
a_{31} & a_{32}
\end{vmatrix}.
\end{align*}
$$

(c) Finally, comparing equations (19.3) with equations (19.4), show that

$$
\lambda_1 + \lambda_2 + \cdots + \lambda_n = \text{tr} A,
$$

and hence that $A$ is singular if and only if at least one of its eigenvalues is zero.

20. Show that if two $n \times n$ matrices $A$ and $B$ have the same eigenvalues $\lambda_1, \ldots, \lambda_n$ and the same LI eigenvectors $e_1, \ldots, e_n$, then it must be true that $A = B$.

21. Can an $n \times n$ matrix have more than $n$ LI eigenvectors? Explain.

22. (Markov matrices) Recall the definition of a Markov matrix, and the fact every Markov matrix contains $\lambda = 1$ among its eigenvalues. You may use the result stated in Exercise 11, if you need it.

(g) Find the eigenvalues of

$$
A = \begin{bmatrix}
0 & 0.5 & 0.5 \\
0.5 & 0.5 & 0 \\
0 & 1 & 0
\end{bmatrix}.
$$

(b) Find one eigenvalue of

$$
A = \begin{bmatrix}
8 & 10 & 12 \\
9 & 10 & 11 \\
10 & 10 & 10
\end{bmatrix}.
$$

(c) Find one eigenvalue of

$$
A = \begin{bmatrix}
3 & 1 & 2 \\
2 & 2 & 1 \\
0 & 2 & 2
\end{bmatrix}.
(d) Determine the eigenvalues and a basis for each eigenspace for the $20 \times 20$ matrix $A$ having unity for each of its 400 elements.

(e) Determine the eigenvalues and a basis for each eigenspace for the $30 \times 30$ matrix $A$ having 2 for each of its 900 elements.

23. In each case, use the same method as in Example 3 to find the general solution of the given system of coupled differential equations, if possible. If your solution falls short of being a general solution, explain why that happened. Primes denote $d/dt$.

(a) $x' = x + 2y$

(b) $y' = 3x + 6y$

(c) $x'' = 2x + y$

(d) $y'' = 9x + 2y$

(e) $x' = 4x + y$

(f) $y' = -5x + z$

(g) $x' = 2x - y$

(h) $y' = -x + y$

24. (Cayley–Hamilton theorem) The Cayley–Hamilton theorem states that if the characteristic equation of any square matrix $A$ is $\lambda^n + \alpha_1 \lambda^{n-1} + \cdots + \alpha_n = 0$, then $A^n + \alpha_1 A^{n-1} + \cdots + \alpha_n A + \alpha_n I = 0$; i.e., $A$ satisfies its characteristic equation.

(a) Prove this theorem for the general $2 \times 2$ case, $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

(b) If $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, show that $A^2 - 4A + 3I = 0$ so that

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

25. (Generalized eigenvalue problem) If $B \neq I$, then $Ax = \lambda Bx$ is called a generalized eigenvalue problem. It should be easy to see that in this case the characteristic equation is $\det(A - \lambda B) = 0$, and that the eigenvectors then follow as the nontrivial solutions of $(A - \lambda B)x = 0$. Find the eigenvalues and eigenspaces in each case.

(a) $\begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} 8 & -4 \\ -4 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

(b) $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

(c) $\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \lambda \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

(d) $\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \lambda \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$

26. In seeking the eigenvalues and eigenvectors of a given matrix $A$, is it permissible first to simplify $A$ by means of some elementary row operations? (That is, are the eigenvalues and eigenvectors of $A$ invariant with respect to elementary row operations?) Explain.

27. In each case, given the values of $a, b, c, d$ in (32), use the eigenvalues to classify the singularity as an unstable node, stable node, saddle, stable focus, center, or unstable focus. If applicable, use the eigenvectors to give a labeled sketch of any stable and/or unstable manifolds.

(a) $a = 1, b = 1, c = 3, d = -1$

(b) $a = -1, b = -3, c = 1, d = 1$

(c) $a = 4, b = 1, c = 1, d = 4$

(d) $a = -3, b = 1, c = 1, d = -3$

(e) $a = 3, b = 1, c = -1, d = 3$

(f) $a = -2, b = -2, c = 2, d = -2$

(g) $a = 1, b = 2, c = 3, d = 4$

(h) $a = 5, b = 1, c = -8, d = 1$

11.3 Symmetric Matrices

In applications, symmetric matrices arise surprisingly often, and their symmetry leads to important results regarding their eigenvalues and eigenvectors. Thus, it is important to treat this case separately.

11.3.1. Eigenvalue problem $Ax = \lambda x$. We have the following three important
results. We shall prove the third, but defer proofs for the first two to the exercises.

**THEOREM 11.3.1 Real Eigenvalues**
If $A$ is symmetric ($A^T = A$), then all of its eigenvalues are real.

**THEOREM 11.3.2 Dimension of Eigenspace**
If an eigenvalue $\lambda$ of a symmetric matrix $A$ is of multiplicity $k$, then the eigenspace corresponding to $\lambda$ is of dimension $k$.

**THEOREM 11.3.3 Orthogonality of Eigenvectors**
If $A$ is symmetric, then eigenvectors corresponding to distinct eigenvalues are orthogonal.

*Proof:* Let $e_j$ and $e_k$ be eigenvectors corresponding to distinct eigenvalues $\lambda_j$ and $\lambda_k$, respectively. Thus,

$$ Ae_j = \lambda_j e_j \quad \text{and} \quad Ae_k = \lambda_k e_k. \quad (1a,b) $$

Next, recall that the dot product of $n$-dimensional column vectors $x$ and $y$ is $x \cdot y = x^T y$, and that $(AB)^T = B^T A^T$ for any matrices $A$ and $B$ that are conformable for multiplication. Then, if we dot $e_k$ into each side of (1a) and dot each side of (1b) into $e_j$ [i.e., we “pre-dot” (1a) with $e_k$ and “post-dot” (1b) with $e_j$], we obtain

$$
\begin{align*}
\langle e_k, (Ae_j) \rangle &= \langle e_k, (\lambda_j e_j) \rangle \\
\langle e_k^T A e_j \rangle &= \lambda_j \langle e_k^T e_j \rangle \\
\langle (Ae_k)^T e_j \rangle &= \lambda_k \langle e_k^T e_j \rangle \\
\langle e_k^T A^T e_j \rangle &= \lambda_k \langle e_k^T e_j \rangle.
\end{align*}
\]

But $A^T = A$ by assumption so if we subtract the bottom equations on the left and right of the vertical divider, we obtain

$$ 0 = \langle \lambda_j - \lambda_k \rangle \langle e_k^T e_j \rangle. \quad (3) $$

Finally, $\lambda_j - \lambda_k \neq 0$ since $\lambda_j$ and $\lambda_k$ were assumed to be distinct so it follows from (3) that $e_k^T e_j = 0$. Thus, $e_k \cdot e_j = 0$, as claimed. $\blacksquare$

As usual, be careful not to read converses into theorems. For instance, Theorem 11.3.1 says that if $A$ is symmetric, then its eigenvalues are real. It does not say that the eigenvalues of $A$ are real if and only if $A$ is symmetric. For instance, in
each of the four examples of Section 11.2 the $\lambda$'s are real, yet none of the matrices is symmetric.

**EXAMPLE 1.** For the symmetric matrix

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix},$$

we find (Exercise 1) the eigenvalues and eigenspaces

$$\lambda_1 = 4, \ e_1 = \alpha \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \lambda_2 = 1, \ e_2 = \beta \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} + \gamma \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix},$$

where $\lambda_2 = 1$ is of multiplicity two [i.e., the characteristic polynomial can be factored as $-(\lambda-4)(\lambda-1)^2$]. Since $A$ is symmetric, the theorems apply. In accordance with Theorem 11.3.1, the $\lambda$'s are real; in accordance with Theorem 11.3.2, $\lambda_1$ is of multiplicity 1 and its eigenspace is one-dimensional (namely, span{$[1,1,1]^T$}), and $\lambda_2$ is of multiplicity 2 and its eigenspace is two-dimensional (namely, span{$[-1,0,1]^T, [-1,1,0]^T$}); and in accordance with Theorem 11.3.3, $e_1 \cdot e_2 = 0$ for all choices of $\alpha, \beta, \gamma$. The eigenspace $e_2$ is the plane through the origin (in 3-space) that is spanned by $[-1,0,1]^T$ and $[-1,1,0]^T$, and the eigenspace $e_1$ is the line through the origin that is spanned by $[1,1,1]^T$ and is normal to the plane.

The vectors $[-1,0,1]^T$ and $[-1,1,0]^T$ in $e_2$ are L1 and a basis for $e_2$, but happen not to be orthogonal. Their lack of orthogonality does not violate Theorem 11.3.3 since they come from the same $\lambda$, not from distinct $\lambda$'s. Nonetheless, we can "trade those vectors in" for two within $e_2$ that are orthogonal (in a nonunique way, in fact, for there is an infinite number of pairs of orthogonal vectors within that plane). For instance, we can choose

$$e_2 = [-1,0,1]^T$$

[i.e., by setting $\beta = 1$ and $\gamma = 0$ in (5)] and seek

$$e_3 = \beta [-1,0,1]^T + \gamma [-1,1,0]^T$$

such that

$$e_2 \cdot e_3 = (1)(-\beta - \gamma) + (0)(\gamma) + (1)(\beta) = 2\beta + \gamma = 0.$$  

Choosing $\beta = 1$, say, then $\gamma = -2$, and (7) gives

$$e_3 = [1, -2, 1]^T,$$

and the vectors given by (6) and (9) constitute an orthogonal basis for the eigenspace corresponding to the eigenvalue 1.

And since (with $\alpha = 1$, say)

$$e_1 = [1,1,1]^T$$

556  Chapter 11. The Eigenvalue Problem
is orthogonal to each of those vectors, it follows that the eigenvectors given by (6), (9), and (10) constitute an orthogonal basis for 3-space. That is, among the eigenvectors of the $3 \times 3$ symmetric $A$ given by (4) we can find an orthogonal basis for 3-space.

COMMENT 1. The $A$ matrix in (4) happens to be symmetric about the other diagonal as well as about the main diagonal. That symmetry is irrelevant; by symmetry we always mean that $A^T = A$, which is symmetry about the main diagonal (from upper left to bottom right).

COMMENT 2. You might be thinking "Of course the eigenvalues are real, for the $A$ matrix is real." No, $A$ being real implies only that the coefficients are real in its characteristic equation, and a polynomial equation with real coefficients can have complex roots. For instance, the real but nonsymmetric matrix $A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ has the characteristic equation $\lambda^2 - 2\lambda + 2 = 0$ and the complex eigenvalues $\lambda = 1 \pm i$.

COMMENT 3. The procedure that we used to obtain the orthogonal set $\{e_2, e_3\}$ from the $L_1$ set $\{[-1,0,1]^T, [-1,1,0]^T\}$ is essentially the Gram–Schmidt orthogonalization procedure explained in Exercise 11 of Section 9.9. 

Generalization of the ideas contained in Example 1 yields the following theorem.

**THEOREM 11.3.4 Orthogonal Basis**

If an $n \times n$ matrix $A$ is symmetric, then its eigenvectors provide an orthogonal basis for $n$-space.

**Proof:** If all of $A$'s $n$ eigenvalues are distinct then, according to Theorem 11.3.3, the $n$ eigenspaces are orthogonal (each being a one-dimensional line in $n$-space) and therefore provide $n$ orthogonal vectors, which necessarily constitute a basis for $n$-space. What if the eigenvalues are not distinct? Suppose that all are distinct except for one, say $\lambda$, which is of multiplicity $k$. Then the $n - k$ eigenvectors corresponding to the other eigenvalues are orthogonal to each other and also to all vectors in the eigenspace corresponding to $\lambda$. Further, $\lambda$'s eigenspace is $k$-dimensional (Theorem 11.3.2) and hence contains $k$ orthogonal vectors. Altogether, then we have $(n - k) + k = n$ orthogonal eigenvectors and hence an orthogonal basis for $n$-space. A similar argument applies if there is more than one repeated eigenvalue.

**Example 2.** Free Vibration of a Two-Mass System. Consider the system of two masses subjected to forces $f_1(t)$ and $f_2(t)$ and restrained laterally by springs and supported vertically by a frictionless table as shown in Fig. 1. The equations of motion, already derived in Example 3 of Section 3.9.1, are

$$
\begin{align*}
    m_1 x_1'' + (k_1 + k_{12}) x_1 - k_{12} x_2 &= f_1(t), \\
    m_2 x_2'' - k_{12} x_1 + (k_2 + k_{12}) x_2 &= f_2(t).
\end{align*}
$$

Figure 1. Two-mass system.
Let \( m_1 = m_2 = k_1 = k_{12} = k_2 = 1 \), say, for definiteness, and consider the free vibration, where \( f_1(t) = f_2(t) = 0 \). Then (11) becomes

\[
\begin{align*}
x''_1 + 2x_1 - x_2 &= 0 \\
x''_2 + x_1 + 2x_2 &= 0.
\end{align*}
\] (12)

The system (12) is solved in Example 8 in Section 3.9.3, by the method of elimination, and we urge you to review that solution and to compare it with the following matrix eigenvalue problem approach. [We could also solve (12) by the Laplace transform method.]

Let us follow the same line of approach that was put forward in Example 2 of Section 11.2. Namely, seek

\[
\begin{align*}
x_1(t) &= q_1 e^{\lambda t}, \\
x_2(t) &= q_2 e^{\lambda t}.
\end{align*}
\] (13)

Actually, on physical grounds we expect the solution to be a vibration (i.e., the \( \lambda \)'s will turn out to be purely imaginary) so it seems more sensible to seek

\[
\begin{align*}
x_1(t) &= q_1 \sin (\omega t + \phi), \\
x_2(t) &= q_2 \sin (\omega t + \phi),
\end{align*}
\] (14)

where the \( q_1 \)'s are amplitudes, \( \omega \) is the frequency, and \( \phi \) is the phase angle.* Putting (14) into (12) and canceling the \( \sin (\omega t + \phi) \) factors gives

\[
\begin{align*}
-\omega^2 q_1 + 2q_1 - q_2 &= 0, \\
-\omega^2 q_2 - q_1 + 2q_2 &= 0
\end{align*}
\]
or, equivalently,

\[
\begin{bmatrix}
2 & -1 \\
-1 & 2
\end{bmatrix}
\begin{bmatrix}
q_1 \\
q_2
\end{bmatrix}
= \omega^2
\begin{bmatrix}
q_1 \\
q_2
\end{bmatrix},
\] (15)

which is a matrix eigenvalue problem

\[
Aq = \lambda q,
\] (16)

with \( \lambda = \omega^2 \) as the eigenvalue. Solving for the eigenvalues and eigenspaces as explained in Section 11.2 we obtain

\[
\lambda_1 = 1, \quad e_1 = \alpha \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \quad \lambda_2 = 3, \quad e_2 = \beta \begin{bmatrix} 1 \\ -1 \end{bmatrix}.
\] (17)

Each "eigenpair" gives us a solution of the form (14). The first gives \( \omega = \sqrt{\lambda_1} = 1 \), and\(^1\)

\[
\begin{align*}
x &= \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 1 \end{bmatrix} \sin (t + \phi_1),
\end{align*}
\] (18)

where \( \alpha \) and \( \phi_1 \) are arbitrary. The second gives \( \omega = \sqrt{\lambda_2} = \sqrt{3} \), and

\[
\begin{align*}
x &= \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \beta \begin{bmatrix} 1 \\ -1 \end{bmatrix} \sin (\sqrt{3} t + \phi_2),
\end{align*}
\] (19)

\*Along the lines of COMMENT 1 in Example 2 of Section 11.2, if we do not use the same \( \omega \)'s and the same \( \phi \)'s in (14), then we will obtain only the trivial solution \( x_1(t) = x_2(t) = 0 \).

\(^1\)The other root, \( \omega = -1 \), would yield no additional information.
where $\beta$ and $\phi_2$ are arbitrary, and there is no reason why $\phi_2$ should be the same as $\phi_1$. One can verify that (18) satisfies (12) for any $\alpha$ and any $\phi_1$ and that (19) also satisfies (12) for any $\beta$ and any $\phi_2$. Since (12) is linear and homogeneous, it follows that the linear combination

$$
\begin{bmatrix}
  x_1(t) \\
  x_2(t)
\end{bmatrix} = \alpha \begin{bmatrix}
  1 \\
  1
\end{bmatrix} \sin (t + \phi_1) + \beta \begin{bmatrix}
  -1 \\
  1
\end{bmatrix} \sin (\sqrt{3} t + \phi_2)
$$

is also a solution; indeed, it is a general solution of (12). Or, returning to scalar form, we have

$$
\begin{align*}
  x_1(t) &= \alpha \sin (t + \phi_1) + \beta \sin (\sqrt{3} t + \phi_2), \\
  x_2(t) &= \alpha \sin (t + \phi_1) - \beta \sin (\sqrt{3} t + \phi_2).
\end{align*}
$$

Here $\alpha, \beta, \phi_1, \phi_2$ are the constants of integration [just as $A, B$ are the constants of integration in the general solution $x(t) = Ae^{\omega t} + Be^{-\omega t}$ of $x'' - 9x = 0$], and are determined from the initial conditions $x_1(0), x_2(0), x'_1(0), x'_2(0)$.

COMMENT 1. Note the central and organizing role of the eigenvalue problem. Each eigenpair defines a vibrational “mode,” the eigenvalue gives the vibrational frequency $\omega = \sqrt{\lambda}$ and the eigenvector gives the mode shape or configuration. The frequencies are called the eigenfrequencies, or natural frequencies (natural in that they correspond to the free, unforced vibration). The two terms on the right-hand side of (20) are called the orthogonal modes of vibration, orthogonal because $[1, 1]^T [1, -1]^T = 0$, that orthogonality being a consequence of the symmetry of $A$. The first term is called the low mode because it occurs at the lower of the two natural frequencies, and the second term is called the high mode because it is at the higher of those two frequencies.

COMMENT 2. Depending upon the initial conditions, we can excite either one of those modes or both of them. For instance, the conditions $x_1(0) = x_2(0) = 0$ and $x'_1(0) = x'_2(0) = 1$ give $\beta = 0, \alpha = 1, \phi_1 = 0$ ($\phi_2$ is irrelevant because $\beta = 0$) and hence the low mode motion

$$
\begin{align*}
  x_1(t) &= \sin t, \\
  x_2(t) &= \sin t,
\end{align*}
$$

the conditions $x'_1(0) = x'_2(0) = 0$ and $x_1(0) = 1, x_2(0) = -1$ give $\alpha = 0, \beta = 1, \phi_2 = \pi/2$ ($\phi_1$ is irrelevant because $\alpha = 0$) and hence the high mode motion

$$
\begin{align*}
  x_1(t) &= \sin (\sqrt{3} t + \frac{\pi}{2}) = \cos \sqrt{3} t, \\
  x_2(t) &= -\sin (\sqrt{3} t + \frac{\pi}{2}) = -\cos \sqrt{3} t,
\end{align*}
$$

and the conditions $x_1(0) = 1, x_2(0) = x'_1(0) = x'_2(0) = 0$ give a motion containing both modes,

$$
\begin{align*}
  x_1(t) &= \frac{1}{2} \left( \sin (t + \frac{\pi}{2}) + \sin (\sqrt{3} t + \frac{\pi}{2}) \right) = \frac{1}{2} \cos t + \frac{1}{2} \cos \sqrt{3} t, \\
  x_2(t) &= \frac{1}{2} \left( \sin (t + \frac{\pi}{2}) - \sin (\sqrt{3} t + \frac{\pi}{2}) \right) = \frac{1}{2} \cos t - \frac{1}{2} \cos \sqrt{3} t.
\end{align*}
$$
Figure 2. Low mode (22), high mode (23), mixed modes (24).

These three motions are contrasted in Fig. 2. Observe that each of the individual modes is simple and "clean," but the mixed mode motion is not. In the low mode the masses vibrate in unison and at the low frequency \( \omega = 1 \), and in the high mode their motions are opposite and at the high frequency \( \omega = \sqrt{3} \).

To be sure it is clear how to apply the initial conditions, let us derive (24). We have the four equations

\[
\begin{align*}
    x_1(0) &= 1 = \alpha \sin \phi_1 + \beta \sin \phi_2, \\
    x_2(0) &= 0 = \alpha \sin \phi_1 - \beta \sin \phi_2, \\
    x'_1(0) &= 0 = \alpha \cos \phi_1 + \sqrt{3}\beta \cos \phi_2, \\
    x'_2(0) &= 0 = \alpha \cos \phi_1 - \sqrt{3}\beta \cos \phi_2,
\end{align*}
\]

in \( \alpha, \beta, \phi_1, \phi_2 \). They happen not to be linear algebraic equations [i.e., of the form \((\alpha + \beta + \phi_1 + \phi_2 = \), where the parentheses contain constants], but they are readily solved. For instance, adding the first two and last two gives

\[
\alpha \sin \phi_1 = \frac{1}{2}, \quad \alpha \cos \phi_1 = 0,
\]

and these give \( \phi_1 = \pi/2 \) and \( \alpha = 1/2 \).* Similarly, subtracting the second from the first and the fourth from the third results in \( \phi_2 = \pi/2 \) and \( \beta = 1/2 \).

COMMENT 3. In this example there were two masses and two "degrees of freedom," \( x_1(t) \) and \( x_2(t) \). Consequently, \( A \) was \( 2 \times 2 \) and the motion of each mass was found to consist of a linear combination of two eigenmodes. More generally, if there are \( n \) masses and \( n \) degrees of freedom \( x_1(t), \ldots, x_n(t) \), then the motion of each mass consists of a linear

*These values are not uniquely determined. For instance \( \phi_1 = 3\pi/2 \) and \( \alpha = -1/2 \) satisfy (26) too. However, such differences do not lead to different solutions \( x_1(t) \) and \( x_2(t) \).
11.3. Symmetric Matrices

11.3.2. Nonhomogeneous problem $Ax = \lambda x + c$. (Optional) In Chapters 8 and 10 we studied the general nonhomogeneous equation $Ax = c$, and in this chapter we have studied the eigenvalue problem, which is the homogeneous equation $Ax = \lambda x$. Next, we wish to show that eigenvalue problem concepts can be used to solve nonhomogeneous equations.

Specifically, consider the nonhomogeneous equation

$$Ax = \lambda x + c,$$  

(27)

where the scalar $\lambda$ is a parameter. We ask of $A$ that it be $n \times n$, and that its eigenvectors provide a basis (not necessarily orthogonal) for $\mathbb{R}^n$. That will surely be the case if $A$ is symmetric since then its eigenvectors provide an orthogonal basis for $\mathbb{R}^n$.

Of course, since $\lambda$ is considered as known we could subtract $\lambda x$ from $Ax$ and absorb $\lambda$ into the $A$ matrix. However, if $A$ is a design parameter and we wish to see the explicit effect of $\lambda$ on the solution, then it is best to leave the $\lambda x$ term intact.

The idea is that to solve (27) we first take a “time-out” and solve the eigenvalue problem for $A$ (i.e., $Ax = \lambda x$); that is, solve for the eigenvalues and eigenvectors of $A$, which we denote as $\lambda_1, \ldots, \lambda_n$ (not necessarily distinct) and $e_1, \ldots, e_n$. Next, expand both $x$ and $c$, in (27), in terms of the $\{e_1, \ldots, e_n\}$ basis:

$$x = \sum_{j=1}^{n} a_j e_j$$  

and  

$$c = \sum_{j=1}^{n} c_j e_j.$$  

(28a,b)

The $c_j$'s are known (i.e., they can be computed) since we know $c$ and the $e_j$ basis vectors so the $a_j$'s are our unknowns. To evaluate them, put (28) into (27):

$$A \sum_{j=1}^{n} a_j e_j = \lambda \sum_{j=1}^{n} a_j e_j + \sum_{j=1}^{n} c_j e_j.$$  

(29)

But

$$A \sum_{j=1}^{n} a_j e_j = \sum_{j=1}^{n} a_j A e_j = \sum_{j=1}^{n} a_j \lambda_j e_j,$$  

(30)

so that (29) can be re-expressed as

$$\sum_{j=1}^{n} (\lambda_j - \lambda) a_j e_j = \sum_{j=1}^{n} c_j e_j.$$  

(31)

Finally, since the $e_j$'s are LI (for they are a basis) it follows from (31) that

$$(\lambda_j - \lambda) a_j = c_j. \quad (j = 1, \ldots, n)$$  

(32)

At this point we need to be careful to distinguish these cases:

---

*By a parameter we mean a constant, the value of which we are free to specify. In the equation $x'' + 9x = F \sin \Omega t$, for instance, $F$ and $\Omega$ are parameters.*
(i) Suppose that none of the \( \lambda_j \)'s equals \( \Lambda \). Then we can divide both sides of (32) by \( \lambda_j - \Lambda \) and obtain \( a_j = c_j / (\lambda_j - \Lambda) \), and the unique solution
\[
x = \sum_{j=1}^{n} \frac{c_j}{\lambda_j - \Lambda} e_j
\]
(33)
of (27).

(ii) Next, suppose that \( \Lambda = \lambda_1 \), where \( \lambda_1 \) is an eigenvalue of multiplicity 1. Then (32) becomes \( (0)a_1 = c_1 \) for \( j = 1 \), and two possibilities exist: if \( c_1 \neq 0 \), then there exists no \( a_1 \) satisfying \( (0)a_1 = c_1 \), and there is no solution of (27); but if \( c_1 = 0 \), then \( a_1 \) is arbitrary, and (27) admits a nonunique solution, the one-parameter family of solutions
\[
x = a_1 e_1 + \sum_{j=2}^{n} \frac{c_j}{\lambda_j - \Lambda} e_j. \quad (a_1 \text{ arbitrary})
\]
(34)

(iii) Similarly if \( \Lambda = \lambda_1 \) is an eigenvalue of multiplicity \( p \); if \( c_1, \ldots, c_p \) are not all zero, then there is no solution of (27); and if \( c_1 = \cdots = c_p = 0 \), then there is a nonunique solution, the \( p \)-parameter family of solutions
\[
x = a_1 e_1 + \cdots + a_p e_p + \sum_{j=p+1}^{n} \frac{c_j}{\lambda_j - \Lambda} e_j. \quad (a_1, \ldots, a_p \text{ arbitrary})
\]
(35)

It is illuminating to compare (35) with the solution (18) in Section 10.5, to the problem \( Ax = c \) (i.e., \( \Lambda = 0 \) in that case). There, \( x_0 \) is a particular solution (i.e., \( Ax_0 = c \)), and \( x_1, \ldots, x_p \) were homogeneous solutions (i.e., \( Ax_1 = \cdots = Ax_p = 0 \)). In (35), the \( \sum_{j=p+1}^{n} \) term corresponds to \( x_0 \), and \( e_1, \ldots, e_p \) are homogeneous solutions (i.e., solutions of (27) with \( c \) removed)\(^*\).

**EXAMPLE 3.** Forced Vibration of the Two-Mass System. To illustrate, consider the mechanical system of Example 2 again, but this time with the forcing functions \( f_1(t) = F_1 \sin \Omega t \) and \( f_2(t) = F_2 \sin \Omega t \). Then in place of (12) we have
\[
x''_1 + 2x_1 - x_2 = F_1 \sin \Omega t, \\
x''_2 - x_1 + 2x_2 = F_2 \sin \Omega t.
\]
(36)

We already found the general solution of the homogeneous equations (12), so in this example let us seek only a particular solution (the total solution then being the sum of the two). Let us seek a particular solution in the form\(^1\)
\[
x_1(t) = q_1 \sin \Omega t, \\
x_2(t) = q_2 \sin \Omega t.
\]
(37)

\(^*\)Of course there is no reason why the \( e_1, \ldots, e_p \) need to equal \( x_1, \ldots, x_p \), but it is true that \( \text{span}\{e_1, \ldots, e_p\} \) and \( \text{span}\{x_1, \ldots, x_p\} \) are identical.

\(^1\)The form (37) is inspired by the method of undetermined coefficients. Actually, we might try \( x_1(t) = q_1 \sin \Omega t + r_1 \cos \Omega t \) and \( x_2(t) = q_2 \sin \Omega t + r_2 \cos \Omega t \), but we can anticipate that the \( \cos \Omega t \) terms will not be needed (i.e., we will find that \( r_1 = r_2 = 0 \)) because there are no \( x_1' \) or \( x_2' \) terms on the left-hand side of (36).
Putting (37) into (36), canceling the \( \sin \Omega t \) factors, and expressing the equations in matrix form, we have

\[
\begin{bmatrix}
2 & -1 \\
-1 & 2
\end{bmatrix}
\begin{bmatrix}
q_1 \\
q_2
\end{bmatrix} = \Omega^2
\begin{bmatrix}
q_1 \\
q_2
\end{bmatrix} +
\begin{bmatrix}
F_1 \\
F_2
\end{bmatrix}
\]

or

\[
A \mathbf{q} = \Lambda \mathbf{q} + \mathbf{c},
\]

where

\[
A = \begin{bmatrix}
2 & -1 \\
-1 & 2
\end{bmatrix}, \quad \Lambda = \Omega^2, \quad \mathbf{c} = \begin{bmatrix}
F_1 \\
F_2
\end{bmatrix}.
\]

Recall from (17) that the eigenvalues and eigenvectors of \( A \) are (with \( \alpha = \beta = 1 \), say)

\[
\lambda_1 = 1, \mathbf{e}_1 = \begin{bmatrix}
1 \\
1
\end{bmatrix}; \quad \lambda_2 = 3, \mathbf{e}_2 = \begin{bmatrix}
1 \\
-1
\end{bmatrix}.
\]

Suppose that \( \Lambda = \lambda_1 \) and \( \Lambda = \lambda_2 \); that is, the driving frequency \( \Omega \) does not equal either of the natural frequencies 1 and \( \sqrt{3} \). Then we have the unique solution (33), where "x" is \( \mathbf{q} \),

\[
\mathbf{e}_1 = (\mathbf{c} \cdot \mathbf{e}_1)/(\mathbf{e}_1 \cdot \mathbf{e}_1) = (F_1 + F_2)/2, \quad \mathbf{e}_2 = (\mathbf{c} \cdot \mathbf{e}_2)/(\mathbf{e}_2 \cdot \mathbf{e}_2) = (F_1 - F_2)/2, \quad \lambda_1 = 1, \lambda_2 = 3, \text{ and } \Lambda = \Omega^2.\]

Thus,

\[
\mathbf{q} = \frac{F_1 + F_2}{2(1 - \Omega^2)} \begin{bmatrix}
1 \\
1
\end{bmatrix} + \frac{F_1 - F_2}{2(3 - \Omega^2)} \begin{bmatrix}
1 \\
-1
\end{bmatrix} = \frac{1}{(1 - \Omega^2)(3 - \Omega^2)} \begin{bmatrix}
(2 - \Omega^2)F_1 + F_2 \\
F_1 + (2 - \Omega^2)F_2
\end{bmatrix}
\]

so the desired particular solution (37), of (36), is

\[
x_1(t) = \frac{(2 - \Omega^2)F_1 + F_2}{(1 - \Omega^2)(3 - \Omega^2)} \sin \Omega t,
\]

\[
x_2(t) = \frac{F_1 + (2 - \Omega^2)F_2}{(1 - \Omega^2)(3 - \Omega^2)} \sin \Omega t.
\]

In Section 3.8 we studied the forced vibration of a single mass, and stressed the importance to an engineer of the amplitude- and phase-response curves. Here too, let us plot the amplitude-response curves (Fig. 3) for the representative case where \( F_1 = 1 \) and \( F_2 = 0 \). The amplitudes \( A_1(\Omega) \) and \( A_2(\Omega) \) are, from (42a) and (42b),

\[
A_1(\Omega) = \left| \frac{(2 - \Omega^2)}{(1 - \Omega^2)(3 - \Omega^2)} \right| \quad \text{and} \quad A_2(\Omega) = \left| \frac{1}{(1 - \Omega^2)(3 - \Omega^2)} \right|,
\]

and we observe that these tend to infinity as \( \Omega \) tends to either of the natural frequencies, 1 and \( \sqrt{3} \).

What if \( \Omega \) equals 1 or \( \sqrt{3} \)? Then our derivation of (43) does not hold since it is based on the assumption that \( \Lambda \neq \lambda_1 \) and \( \Lambda \neq \lambda_2 \). If we do have \( \Lambda = \lambda_1 \) (i.e., \( \Omega = 1 \), say,

*If the formula \( c_j = (\mathbf{c} \cdot \mathbf{e}_j)/(\mathbf{e}_j \cdot \mathbf{e}_j) \) is unfamiliar to you, we urge you to review Section 9.9, especially equation (23) therein.*
The method presented here for solving (27) is known as the eigenvector expansion method. What is the advantage in using the eigenvectors of \( \mathbf{A} \) as our basis: why not use any basis for \( \mathbb{R}^n \)? The idea is that the vector \( \mathbf{A} \mathbf{e}_j \) in the middle of (30) is simply \( \lambda_j \mathbf{e}_j \) if \( \mathbf{e}_j \) is an eigenvector of \( \mathbf{A} \), whereas it would be a linear combination of all the base vectors if some other basis were used. In that case we would end up with a coupled system of linear algebraic equations for the \( a_j \)'s, rather than the uncoupled (and hence readily solved) system (32). This same comment applies to Example 4 in Section 11.2 where, to study the stability of the equilibrium population, we expanded the initial population vector \( \mathbf{p}_0 \) in terms of the eigenvectors of \( \mathbf{A} \).

Closure. Symmetric matrices arise frequently in applications, and their symmetry leads to several important results regarding their eigenvalues and eigenvectors: their eigenvalues are real, eigenvectors corresponding to distinct eigenvalues are orthogonal, and the eigenvectors of an \( n \times n \) symmetric matrix provide an orthogonal basis for \( \mathbb{R}^n \).

We discussed an important application to multimass mechanical systems, and found that the free oscillation can be represented as the superposition of orthogonal modes, with the eigenvalues giving the modal frequencies and the eigenvectors giving the modal configurations.

The special importance of symmetric matrices is further revealed in the remaining sections of this chapter.

In the second half of this section we return to the nonhomogeneous equation \( \mathbf{Ax} = \mathbf{c} \), actually to the form \( \mathbf{Ax} = \Lambda \mathbf{x} + \mathbf{c} \) where, \( \Lambda \) is a parameter, and develop a line of approach known as the eigenvector expansion method. The idea behind that method is to compute first the eigenvalues and eigenvectors of \( \mathbf{A} \). Assuming that we can obtain a basis of \( \mathbb{R}^n \) from the eigenvectors of \( \mathbf{A} \) (as is always possible for symmetric \( \mathbf{A} \)'s which, indeed, provide us with orthogonal bases), we use that basis, which is the most convenient or "natural" basis to expand \( \mathbf{x} \) and \( \mathbf{c} \). Finally, equating coefficients of the various base vectors on both sides of the equation gives uncoupled linear algebraic equations on the unknown coefficients in the expansion of \( \mathbf{x} \).

The eigenvector expansion method is applicable to other cases as well, such as ordinary and partial differential equations. We will meet it again when we come to
EXERCISES 11.3

1. From the eigenvectors of the given \( n \times n \) matrix obtain an orthogonal basis for \( \mathbb{R}^n \).

\[
\begin{align*}
\text{(a)} & \quad \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \\
\text{(b)} & \quad \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} \\
\text{(c)} & \quad \begin{bmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{bmatrix} \\
\text{(d)} & \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 4 \\ 0 & 4 & 0 \\ 4 & 0 & 0 \end{bmatrix} \\
\text{(e)} & \quad \begin{bmatrix} 6 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \\ 0 & 2 & 2 \end{bmatrix} \\
\text{(f)} & \quad \begin{bmatrix} 0 & 0 & 0 & 4 \\ 0 & 0 & 4 & 0 \\ 0 & 4 & 0 & 0 \\ 4 & 0 & 0 & 0 \end{bmatrix}
\end{align*}
\]

Thus,

\[
\begin{align*}
(\mathbf{Ae}) \cdot \bar{e} & = (\lambda e) \cdot \bar{e} \quad \text{and} \quad e \cdot (\mathbf{Ae}) = e \cdot (\bar{\lambda}e), \quad (3.3)
\end{align*}
\]

2. Determine whether or not the eigenvectors of the given non-symmetric \( 2 \times 2 \) matrix provide an orthogonal basis for \( \mathbb{R}^2 \).

\[
\begin{align*}
\text{(g)} & \quad \begin{bmatrix} 0 & 0 \\ 3 & 1 \end{bmatrix} \\
\text{(h)} & \quad \begin{bmatrix} 1 & 9 \\ 1 & 1 \end{bmatrix} \\
\text{(i)} & \quad \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \\
\text{(j)} & \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}
\end{align*}
\]

3. (Proof of Theorem 11.3.1) (a) Prove Theorem 11.3.1 for the simple case where \( A \) is merely \( 2 \times 2 \), i.e., of the form \[
\begin{bmatrix}
 a & b \\
 b & d
\end{bmatrix}
\]

(b) We will now supply the “skeleton” of a general proof of Theorem 11.3.1 and ask you to answer questions about the steps. Overhead bars will denote complex conjugates: if \( z = x + iy \), then \( \bar{z} = x - iy \). Proof:

\[
\mathbf{Ae} = \lambda e, \quad (3.1)
\]

so

\[
\mathbf{A} \bar{e} = \bar{\lambda} \bar{e}. \quad (3.2)
\]

Dot both sides of (3.1) into \( \bar{e} \), and dot \( e \) into both sides of (3.2):

\[
(\mathbf{Ae}) \cdot \bar{e} = (\lambda e) \cdot \bar{e} \quad \text{and} \quad e \cdot (\mathbf{Ae}) = e \cdot (\bar{\lambda}e). \quad (3.3)
\]

Thus,

\[
\begin{align*}
e^T A^T \bar{e} = \bar{\lambda} e^T \bar{e} \quad \text{and} \quad e^T A \bar{e} = \lambda e^T \bar{e}, \quad (3.4)
\end{align*}
\]

so

\[
(\lambda - \bar{\lambda}) e^T \bar{e} = 0, \quad (3.5)
\]

and hence

\[
\lambda = \bar{\lambda} \quad (3.6)
\]

so \( \lambda \) is real. Questions: How does (3.2) follow from (3.1)? (3.4) from (3.3)? (3.5) from (3.4)? (3.6) from (3.5)?

4. (Proof of Theorem 11.3.2) Prove Theorem 11.3.2 for the simple case where \( n = 2 \).

5. (Post-dotting versus pre-dotting) Observe that the top left equation in (2) was obtained by dotting \( e_k \) into both sides of (1a); i.e., we “pre-dotted” (1a) with \( e_k \). However, the top right equation in (2) was obtained by dotting both sides of (1b) into \( e_j \); i.e., we “post-dotted” \( e_j \) into (1b). Now, post-dotting or pre-dotting doesn’t matter, in the sense that the dot product is commutative: \( x \cdot y = y \cdot x \). Nevertheless, one may be more convenient than the other. Specifically, show that if we write the top left equation in (2) as \( (\mathbf{Ae}_j) \cdot e_k = (\lambda_j e_j) \cdot e_k \), instead of \( e_k \cdot (\mathbf{Ae}_j) = e_k \cdot (\bar{\lambda}_j e_j) \), then it is more difficult to obtain (3) and hence the desired result.

6. Use any theorem(s) from Chapter 3 to show that the solution to (12), with the initial conditions \( x_1(0), x_2(0), x'_1(0) \), and \( x'_2(0) \) specified, is unique.

7. Beginning with (21), complete the solution for the following initial conditions:

\[
\begin{align*}
\text{(a)} & \quad x_1(0) = 2, x_2(0) = 3, x'_1(0) = x'_2(0) = 0 \\
\text{(b)} & \quad x_1(0) = 1, x_2(0) = x'_1(0) = 0, x'_2(0) = -3 \\
\text{(c)} & \quad x_1(0) = x_2(0) = x'_1(0) = 0, x'_2(0) = 5 \\
\text{(d)} & \quad x_1(0) = -2, x_2(0) = x'_1(0) = 0, x'_2(0) = 3
\end{align*}
\]

8. Consider the three-mass system shown.
566  Chapter 11. The Eigenvalue Problem

Consider

The equations of motion for the free vibration, taking all masses and spring stiffnesses to be 1, say.

Following the same lines as in Example 2, find the orthogonal modes (i.e., the eigenvectors) and their corresponding natural frequencies, proceeding either by hand or by using computer software.

Give any set of initial conditions that will excite the low frequency mode only: the middle mode only; the high mode only.

Find the solution \( x_1(t), x_2(t), x_3(t) \) corresponding to the initial condition \( x_1(0) = 1, x_2(0) = x_3(0) = x_1'(0) = x_2'(0) = 0 \).

Consider a mass-spring system like the one shown in Exercise 8, but with five masses and six springs. If all the masses and spring stiffnesses are 1, then the equations of motion are these, in matrix form:

\[
\begin{bmatrix}
  x_1'' \\
  x_2'' \\
  x_3'' \\
  x_4'' \\
  x_5''
\end{bmatrix} + \begin{bmatrix}
  2 & -1 & 0 & 0 & 0 \\
  -1 & 2 & -1 & 0 & 0 \\
  0 & -1 & 2 & -1 & 0 \\
  0 & 0 & -1 & 2 & -1 \\
  0 & 0 & 0 & -1 & 2
\end{bmatrix} \begin{bmatrix}
  x_1 \\
  x_2 \\
  x_3 \\
  x_4 \\
  x_5
\end{bmatrix} = 0
\]

or \( x'' + Ax = 0 \). Observe that the \( A \) matrix is tridiagonal (i.e., all elements are zero except for the main diagonal and the two neighboring diagonals). Physically, this result corresponds to nearest-neighbor coupling whereby each mass feels only its immediate neighbors. Nearest-neighbor coupling occurs in other systems as well. For instance, in modeling single-lane traffic flow, each driver accelerates or decelerates according to the motion of the cars immediately ahead and immediately behind, so the resulting coupled differential equations exhibit nearest-neighbor coupling. The problem that we pose is for you to use computer software to determine the natural frequencies and corresponding mode shapes (eigenvectors).

Suppose further that \( x_1(0) = 1, x_2(0) = x_1'(0) = x_2'(0) = 0, \) proceed as in Example 2 and show that, for the free vibration,

\[
\begin{align*}
  x_1(t) &= \frac{1}{2} (\cos \sqrt{20} t + \cos \sqrt{22} t), \\
  x_2(t) &= \frac{1}{2} (\cos \sqrt{20} t - \cos \sqrt{22} t).
\end{align*}
\]

Next, use the trigonometric identities

\[
\begin{align*}
  \cos A + \cos B &= 2 \cos \frac{A+B}{2} \cos \frac{A-B}{2}, \\
  \cos A - \cos B &= -2 \sin \frac{A+B}{2} \sin \frac{A-B}{2}
\end{align*}
\]

to show that

\[
\begin{align*}
  x_1(t) &= \cos 4.58t \cos 0.11t, \\
  x_2(t) &= \sin 4.58t \sin 0.11t.
\end{align*}
\]

Use (10.3) to sketch \( x_1(t) \) and \( x_2(t) \) versus \( t \) in separate graphs, one below the other, labeling key values. Observe the slow transfer of energy back and forth between \( m_1 \) and \( m_2 \). In vibration theory this phenomenon is known as beats.

Let \( A \) be a symmetric \( n \times n \) matrix. Dotting any eigenvector \( e \) of \( A \) into both sides of \( Ae = \lambda e \) and solving for \( \lambda \), gives

\[
\lambda = \frac{e \cdot Ae}{e \cdot e} = \frac{e^T Ae}{e^T e}.
\]

More generally, if \( x \) is any vector, not necessarily an eigenvector of \( A \), then the number

\[
R(x) = \frac{x^T Ax}{x^T x}
\]

is known as Rayleigh’s quotient, after Lord Rayleigh (John William Strutt, 1842–1919). Imagine putting randomly chosen \( x \) vectors, one after another, into Rayleigh’s quotient. If \( x \) happens to coincides with an eigenvector, then, according to (11.1), \( R(x) \) gives the corresponding eigenvalue. In any case,

\[
|R(x)| = \left| \frac{x^T Ax}{x^T x} \right| \leq |\lambda_1|,
\]

where the eigenvalues are ordered so that \( |\lambda_1| \geq |\lambda_2| \geq ... \geq |\lambda_n| \). That is, \( |R(x)| \) provides a lower bound on the magnitude of the largest eigenvalue of \( A \), where \( x \) is any vector (i.e., any nonzero vector, since \( R(0) = 0/0 \) is undefined). Upper and lower bounds on eigenvalues are sometimes important, and Rayleigh’s quotient is used again in Exercise 12.

(a) Prove the inequality in (11.3). HINT: Since \( A \) is symmetric, it has \( n \) orthogonal eigenvectors \( e_1, ..., e_n \), corresponding to the eigenvalues \( \lambda_1, ..., \lambda_n \). Expand \( x \) in terms of the orthogonal basis \( \{e_1, ..., e_n\} \):
(b) Verify (11.3) for the matrix

\[
A = \begin{bmatrix}
2 & 0 & 0 \\
0 & -1 & -2 \\
0 & -2 & -1
\end{bmatrix}
\]  \hspace{1cm} (11.5)

by taking \( x = [1, 0, 0]^T, [0, 1, 0]^T, [0, 0, 1]^T, [1, 2, 3]^T \) and
\([3, -1, 4]^T\), say, verifying that \( |R(x)| \leq |\lambda_1| \) in each case.

(c) Evaluate \( R(x) \) for \( x = [0.4, 0.3, 0.2]^T, [0.6, 0.2, 0.2]^T, [0.8, 0.1, 0.1]^T, [0.96, 0.02, 0.02]^T, [1.0, 0]^T \), and for \( x = [0, 1, 1.4]^T, [0, 1, 1.1]^T, [0, 1, 1]^T \), and discuss your results in the light of (11.1) and (11.3).

12. (The power method) There exists a simple iterative procedure for calculating eigenvalues and eigenvectors, which is known as the power method. To begin, one selects any nonzero vector \( x^{(0)} \) and then computes \( x^{(1)} = Ax^{(0)} \), \( x^{(2)} = Ax^{(1)} \), and so on. That is,

\[
x^{(k+1)} = Ax^{(k)} \quad (k = 0, 1, 2, \ldots). \quad (12.1)
\]

Before analyzing the situation, let us apply (12.1) and see what happens. Let

\[
A = \begin{bmatrix}
1 & 1 & 1 \\
1 & 0 & 0 \\
1 & 0 & 0
\end{bmatrix}
\]  \hspace{1cm} (12.2)

for example. Choosing \( x^{(0)} = [1, 0, 0]^T \), say, successive application of (12.1) gives

\[
\begin{bmatrix}
1 & 1 & 1 \\
1 & 0 & 0 \\
1 & 0 & 0
\end{bmatrix} \begin{bmatrix}
x^{(0)} \\
x^{(1)} \\
x^{(2)}
\end{bmatrix} = \begin{bmatrix}
1 & 1 & 1 \\
0 & 0 & 1 \\
0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
3 & 5 & 11 \\
5 & 3 & 5 \\
5 & 3 & 5
\end{bmatrix} \begin{bmatrix}
21 \\
21 \\
21
\end{bmatrix},
\]

and so on. Now, observe, from the very nature of the eigenvalue problem \( Ax = \lambda x \), that if \( x^{(0)} \) were an eigenvector of \( A \), then \( Ax^{(0)} \) would be some scalar multiple of \( x^{(0)} \). But \( Ax^{(0)} \equiv x^{(1)} \), and \( x^{(1)} \) is seen from (12.3) not to be a scalar multiple of \( x^{(0)} \). Hence, \( x^{(0)} \) is not an eigenvector of \( A \). Similarly, \( x^{(1)} \) is not an eigenvector because \( x^{(2)} \) is not a multiple of \( x^{(1)} \), and so on. Nevertheless, we see that with each successive step \( x^{(k+1)} \) draws closer and closer to being a multiple of \( x^{(k)} \) so that the sequence \( x^{(k)} \) is evidently approaching an eigenvector of \( A \). In fact, \( x^{(5)} \) is very close to being a multiple of \( x^{(4)} \) so that there is an eigenvector

\[
x = \sum_{j=1}^{n} a_j e_j.
\]  \hspace{1cm} (11.4)

What is the corresponding \( \lambda \), \( \lambda \approx 21/11? \) \( \lambda \approx 11/5? \) An average of the two? We state without proof that one does well to use the Rayleigh quotient from Exercise 11:

\[
\lambda \approx \frac{x^{(4)}^T A x^{(4)}}{x^{(4)}^T x^{(4)}} = \frac{x^{(4)}^T x^{(5)}}{x^{(4)}^T x^{(4)}} = \frac{341}{171} = 1.994. \quad (12.5)
\]

How do (12.4) and (12.5) compare with exact values? The eigenvalues and eigenvectors of \( A \) are

\[
\lambda_1 = 2, \quad e_1 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \quad \lambda_2 = -1, \quad e_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix},
\]

\[
\lambda_3 = 0, \quad e_3 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}
\]

(12.6)

so the iteration (12.3) is evidently converging to \( e_1 \). [It is striking that whereas (12.4) is accurate to only around one part in 20, (12.5) is accurate to around one part in 200. This enhancement of accuracy, using Rayleigh's quotient to determine \( \lambda \), is not a coincidence and can be explained theoretically. See, for instance, Stephen H. Crandall, Engineering Analysis (New York: McGraw-Hill, 1956), Chap. 2.]

To see what is going on, suppose that \( A \) is a symmetric matrix of order \( n \) (although symmetry is more than we need; it would suffice for \( A \) to have \( n \) LI eigenvectors) and let its eigenvalues be ordered so that \( |\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n| \) as in (12.6). Since \( A \) is symmetric, its eigenvectors \( e_1, \ldots, e_n \) are on orthogonal basis for \( n \)-space. Hence, our initial vector \( x^{(0)} \) must be expressible as

\[
x^{(0)} = \sum_{j=1}^{n} a_j e_j. \quad (12.7)
\]

Of course, we cannot compute the \( a_j \)'s since we do not know the \( e_j \)'s yet, but that is no problem: it is the form of (12.7) that is important here. It follows from (12.1) and (12.7) that

\[
x^{(k)} = \sum_{j=1}^{n} a_j \lambda_j^k e_j \]

\[
= \lambda_1^k \left[ a_1 e_1 + a_2 \left( \frac{\lambda_2}{\lambda_1} \right)^k e_2 + \cdots + a_n \left( \frac{\lambda_n}{\lambda_1} \right)^k e_n \right].
\]
If \( \lambda_1 \) is, in fact, dominant, i.e., \( |\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n| \), then \( (\lambda_2/\lambda_1)^k, \ldots, (\lambda_n/\lambda_1)^k \) all tend to zero as \( k \to \infty \), so that \( x^{(k)} \sim \lambda_1^k a_1 e_1 \) as \( k \to \infty \) (provided that \( a_1 \neq 0 \), i.e., provided that \( x^{(0)} \) does not happen to be orthogonal to \( e_1 \)). Eigenvectors can be scaled arbitrarily so the \( \lambda_1^k a_1 \) factor is of little interest; the point is that \( x^{(k)} \) converges to \( e_1 \), the eigenvector corresponding to the dominant eigenvalue. That is precisely what was found in the preceding illustration, wherein the dominant eigenvalue is \( \lambda_1 = 2 \). NOTE: Once again we see that the most convenient basis to use is the basis provided by the \( A \) matrix itself.

(a) Show that (12.8) follows from (12.1) and (12.7).
(b) Determine the dominant eigenvalue and corresponding eigenvector by the power method for

\[
A = \begin{bmatrix}
0 & 0 & 1 \\
0 & 0 & 1 \\
1 & 1 & 1
\end{bmatrix}.
\]

Also, evaluate the eigenvalues and eigenvectors exactly, either by hand or by using computer software. NOTE: Observe that even though your iteration converges, you cannot be certain that the eigenvalue obtained is the dominant one, \( \lambda_1 \), since your chosen \( x^{(0)} \) may (without your knowing it) be orthogonal to \( e_1 \), as mentioned in the sentence following (12.8). Hence, we recommend that you carry out the iteration three times, once with \( x^{(0)} = [1, 0, 0]^T \), once with \( x^{(0)} = [0, 1, 0]^T \), and once with \( x^{(0)} = [0, 0, 1]^T \). since there is no way that all three of these \( x^{(0)} \)'s can be orthogonal to \( e_1 \). (Do you see why this is so?) Go as far as \( x^{(10)} \) in each case, and use the Rayleigh quotient to estimate \( \lambda \), as we did in (12.5).

13. The same as Exercise 12(b), for the given matrix

\[
\begin{bmatrix}
2 & 1 & -1 \\
1 & 4 & 3 \\
-1 & 3 & 4
\end{bmatrix}
\]

(g) A given in Exercise 13(a) \( (\lambda = 7, 3, 0), \Lambda = 2, \mathbf{c} = [1, 2, 3]^T \)
(b) A given in Exercise 13(a) \( (\lambda = 7, 3, 0), \Lambda = 3, \mathbf{c} = [2, 2, 0]^T \)
(c) A given in Exercise 13(a) \( (\lambda = 7, 3, 0), \Lambda = 3, \mathbf{c} = [1, 1, 3]^T \)
(d) A given in Exercise 13(g) \( (\lambda = 4, 2, 0, 0), \Lambda = 1, \mathbf{c} = [3, -1, 1, 0]^T \)
(e) A given in Exercise 13(g) \( (\lambda = 4, 2, 0, 0), \Lambda = 4, \mathbf{c} = [1, 2, 0, 3]^T \)
(f) A given in Exercise 13(g) \( (\lambda = 4, 2, 0, 0), \Lambda = 4, \mathbf{c} = [2, 0, 1, -2]^T \)
(g) A given in Exercise 13(g) \( (\lambda = 4, 2, 0, 0), \Lambda = 0, \mathbf{c} = [1, 3, 3, 1]^T \)
(h) A given in Exercise 13(g) \( (\lambda = 4, 2, 0, 0), \Lambda = 0, \mathbf{c} = [1, 2, 3, 4]^T \)

14. For the \( A \) matrix given in part (b) of Exercise 13, work out \( A^4 \). Apply the power method (Exercise 13) both to \( A \) and \( A^4 \), beginning with \( x^{(0)} = [0, 1, 0]^T \), say. Go as far as \( x^{(2)} \) and use the Rayleigh quotient to estimate \( \lambda \). Explain why the iteration converges more rapidly for \( A^4 \) than for \( A \), and show how to recover the eigenvalue and eigenvector of \( A \) from those of \( A^4 \).

15. Consider the problem \( A x = \lambda x + c \), where \( A, \lambda, c \) are given below (along with the eigenvalues of \( A \), for your convenience).

Solve for \( x \) by the eigenfunction expansion method; if no solution exists, state that. You may use any of equations (33)–(35) without deriving them.

(a) \( \mathbf{A} \) given in Exercise 13(a) \( (\lambda = 7, 3, 0), \Lambda = 2, \mathbf{c} = [1, 2, 3]^T \)
(b) \( \mathbf{A} \) given in Exercise 13(a) \( (\lambda = 7, 3, 0), \Lambda = 3, \mathbf{c} = [2, 2, 0]^T \)
(c) \( \mathbf{A} \) given in Exercise 13(a) \( (\lambda = 7, 3, 0), \Lambda = 3, \mathbf{c} = [1, 1, 3]^T \)
(d) \( \mathbf{A} \) given in Exercise 13(g) \( (\lambda = 4, 2, 0, 0), \Lambda = 1, \mathbf{c} = [3, -1, 1, 0]^T \)
(e) \( \mathbf{A} \) given in Exercise 13(g) \( (\lambda = 4, 2, 0, 0), \Lambda = 4, \mathbf{c} = [1, 2, 0, 3]^T \)
(f) \( \mathbf{A} \) given in Exercise 13(g) \( (\lambda = 4, 2, 0, 0), \Lambda = 4, \mathbf{c} = [2, 0, 1, -2]^T \)
(g) \( \mathbf{A} \) given in Exercise 13(g) \( (\lambda = 4, 2, 0, 0), \Lambda = 0, \mathbf{c} = [1, 3, 3, 1]^T \)
(h) \( \mathbf{A} \) given in Exercise 13(g) \( (\lambda = 4, 2, 0, 0), \Lambda = 0, \mathbf{c} = [1, 2, 3, 4]^T \)

16. To solve the nonhomogeneous equation (27), we first solve for the eigenvalues \( \lambda_j \) and eigenvectors \( \mathbf{e}_j \) of \( \mathbf{A} \), then we use those eigenvectors as a basis to expand \( \mathbf{x} \) and \( \mathbf{c} \). Why do we go to the extra trouble of solving for the eigenvalues and eigenvectors of \( \mathbf{A} \)?

17. (Generalized eigenvalue problem) With \( f_1(t) = f_2(t) = 0 \), \( x_1(t) = q_1 \sin (\omega t + \phi) \), and \( x_2(t) = q_2 \sin (\omega t + \phi) \), (11) becomes the eigenvalue problem

\[
\begin{bmatrix}
\frac{k_1 + k_{12}}{m_1} & -\frac{k_{12}}{m_1} \\
\frac{k_{12}}{m_2} & \frac{k_2 + k_{12}}{m_2}
\end{bmatrix}
\begin{bmatrix}
q_1 \\
q_2
\end{bmatrix} = \omega^2
\begin{bmatrix}
q_1 \\
q_2
\end{bmatrix}
\]

or \( \mathbf{A q} = \lambda \mathbf{q} \), where \( \lambda = \omega^2 \). In Example 2 we took \( m_1 = m_2 = 1 \) so \( \mathbf{A} \) is symmetric and the eigenmodes are orthogonal. In general, however, \( m_1 \neq m_2 \) and \( \mathbf{A} \) is not
symmetric. Nonetheless, observe that we can reexpress (16.1) as
\[
\begin{bmatrix}
  k_1 + k_{12} & -k_{12} \\
  -k_{12} & k_2 + k_{12}
\end{bmatrix}
\begin{bmatrix}
  q_1 \\
  q_2
\end{bmatrix}
= \omega^2
\begin{bmatrix}
  m_1 & 0 \\
  0 & m_2
\end{bmatrix}
\begin{bmatrix}
  q_1 \\
  q_2
\end{bmatrix}
\]
(17.2)
or
\[
Kq = \lambda Mq,
\]
(17.3)
where \( \lambda = \omega^2 \) again. Equation (17.3) is called a generalized eigenvalue problem, "generalized" because of the presence of \( M \). (The generalized eigenvalue problem is introduced in Exercise 20 of Section 11.2.) The eigenvalues are determined by \( \det(K - \lambda M) = 0 \), and then the corresponding eigenvectors follow as the nontrivial solutions of \((K - \lambda M)q = 0\).

Here is the problem:
(a) Let \( K \) and \( M \) be symmetric [as is the case in (17.2), although there is no need for \( M \) to be diagonal as well]. Show that if \( e_j \) and \( e_k \) are eigenvectors corresponding to distinct eigenvalues \( \lambda_j \) and \( \lambda_k \), respectively, then \( e_j \) and \( e_k \) satisfy the generalized orthogonality relation
\[
e_j \cdot (M e_k) = 0.
\]
(17.4)
That is, \( e_j \) is orthogonal to \( e_k \), "relative to \( M \."
(b) Verify the truth of (17.4) for the case
\[
\begin{bmatrix}
  1 & 1 \\
  1 & 1
\end{bmatrix}
\begin{bmatrix}
  q_1 \\
  q_2
\end{bmatrix}
= \lambda
\begin{bmatrix}
  3 & 0 \\
  0 & 2
\end{bmatrix}
\begin{bmatrix}
  q_1 \\
  q_2
\end{bmatrix}
\]

11.4 Diagonalization

We have seen that diagonal matrices are particularly straightforward. For instance, the solution of \( Ax = c \), where \( A \) is \( n \times n \), is generally tedious if \( n \) is large but is simple if \( A \) is diagonal, for then the scalar equations are not coupled. Similarly, raising \( A \) to the \( m \)th power is generally tedious if \( m \) is large but is simple if \( A \) is diagonal. Likewise, the solution of a system of differential equations
\[
x'(t) = Ax(t)
\]
(1)
is generally tedious but is simple if \( A \) is diagonal, for then the scalar equations are uncoupled.

To introduce the idea of diagonalization, let us focus on the application of diagonalization to the solution of the system of differential equations given by (1), where we assume that \( A \) is constant (i.e., its elements do not vary with \( t \)). We have already studied several methods for the solution of (1): the method of elimination (Section 3.9), which is essentially Gauss elimination but where coefficients are differential operators; the Laplace transform method (Section 5.4), which would reduce (1) to \( n \) linear algebraic equations in \( \tilde{x}_1(s), \ldots, \tilde{x}_n(s) \); and seeking \( x(t) = q e^{\lambda t} \) and obtaining an eigenvalue problem (Sections 11.2 and 11.3).

We begin the solution of (1) by diagonalization by making a linear change of variables from \( x_1, \ldots, x_n \) to \( \tilde{x}_1, \ldots, \tilde{x}_n \);
\[
x = Q \tilde{x},
\]
(2)
where \( Q \) is a constant matrix. Written out,

\[
\begin{bmatrix}
  x_1 \\
  \vdots \\
  x_n
\end{bmatrix} =
\begin{bmatrix}
  q_{11} & \cdots & q_{1n} \\
  \vdots & \ddots & \vdots \\
  q_{n1} & \cdots & q_{nn}
\end{bmatrix}
\begin{bmatrix}
  \bar{x}_1 \\
  \vdots \\
  \bar{x}_n
\end{bmatrix}.
\]

(3)

Putting (2) into (1) [and observing that \((Q\bar{x})' = Q\bar{x}'\) because the matrix \( Q \) is constant] gives

\[
Q\bar{x}' = AQ\bar{x}.
\]  

(4)

Since the choice of \( Q \) is ours, we can ask \( Q \) to be invertible. Then, multiplying (4) by \( Q^{-1} \) gives \( Q^{-1}Q\bar{x}' = Q^{-1}AQ\bar{x} \) or

\[
\bar{x}' = Q^{-1}AQ\bar{x}.
\]  

(5)

Given the \( A \) matrix, the idea is to try to find a \( Q \) matrix so that

\[
Q^{-1}AQ = D
\]

is diagonal because then the differential equations within (5) will be uncoupled. If there does exist such a \( Q \) we say that \( A \) is diagonalizable and that \( Q \) diagonalizes \( A \).

Two questions present themselves: given \( A \), does there exist such a \( Q \) and, if so, how do we find it? (There is also a question of uniqueness, but we are not especially interested in whether or not \( Q \) is unique; we’ll be happy to find any \( Q \) that diagonalizes \( A \).)

**THEOREM 11.4.1 Diagonalization**

Let \( A \) be \( n \times n \).

1. \( A \) is diagonalizable if and only if it has \( n \) LI eigenvectors.

2. If \( A \) has \( n \) LI eigenvectors \( e_1, e_2, \ldots, e_n \) and we make these the columns of \( Q \), so that \( Q = [e_1, \ldots, e_n] \), then \( Q^{-1}AQ = D \) is diagonal and the \( j \)th diagonal element of \( D \) is the \( j \)th eigenvalue of \( A \).

**Proof:** First, by \( Q = [e_1, \ldots, e_n] \) we mean that \( Q \) is partitioned into columns, the columns being \( e_1, \ldots, e_n \).

Let us prove that if \( A \) is diagonalizable, then it has \( n \) LI eigenvectors. If \( A \) is diagonalizable, then there is an invertible matrix \( Q \) such that

\[
Q^{-1}AQ = D =
\begin{bmatrix}
  d_1 & 0 & \cdots & 0 \\
  0 & d_2 & & \\
  \vdots & & \ddots & \\
  0 & \cdots & & d_n
\end{bmatrix}.
\]  

(7)
Pre-multiplying both sides of (7) by $Q$ gives $AQ = QD$:

$$AQ = \begin{bmatrix} q_{11} & \cdots & q_{1n} \\ \vdots & \ddots & \vdots \\ q_{n1} & \cdots & q_{nn} \end{bmatrix} \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & d_n \end{bmatrix} = \begin{bmatrix} d_1 q_{11} & \cdots & d_n q_{1n} \\ \vdots & \ddots & \vdots \\ d_1 q_{n1} & \cdots & d_n q_{nn} \end{bmatrix} = \begin{bmatrix} d_1 q_1, \ldots, d_n q_n \end{bmatrix}, \quad (8)$$

where the vector $q_j$ simply denotes the $j$th column of $Q$. Alternatively,

$$AQ = A[q_1, q_2, \ldots, q_n] = [AQ_1, AQ_2, \ldots, AQ_n] \quad (9)$$

and, comparing (8) and (9), we see that

$$AQ_1 = d_1 q_1, \ldots, AQ_n = d_n q_n. \quad (10)$$

Does (10) imply that the $q_j$'s are eigenvectors of $A$? Only if we can be sure that they are nonzero. Since we have assumed $A$ to be diagonalizable, $Q$ must be invertible. Hence, none of its columns $q_j$ can be $0$. Thus, the $d_j$'s and $q_j$'s are the eigenvalues $\lambda_j$ and eigenvectors $e_j$ of $A$. Furthermore, the rank of $Q$ must be $n$ since $Q$ is to be invertible, so (Theorem 10.5.2) its columns must be LI.

Thus far we have proved half of item 1, that if $A$ is diagonalizable, then it has $n$ LI eigenvectors. In doing so we have also proved item 2. It remains to prove the rest of 1, that if $A$ has $n$ LI eigenvectors, then it is diagonalizable. To do so, let us take $Q$ to be made up of columns which are the eigenvectors of $A$, so $Q = [e_1, \ldots, e_n]$. Then

$$AQ = [Ae_1, \ldots, Ae_n] = [\lambda_1 e_1, \ldots, \lambda_n e_n] = \begin{bmatrix} \lambda_1 e_{11} & \cdots & \lambda_n e_{n1} \\ \vdots & \ddots & \vdots \\ \lambda_1 e_{1n} & \cdots & \lambda_n e_{nn} \end{bmatrix} = QD. \quad (11)$$

Finally, $Q$ is invertible since its columns are LI, so pre-multiplying (11) by $Q^{-1}$ gives $Q^{-1}AQ = D$. Hence $A$ is diagonalizable, and the proof is complete. \hfill \blacksquare

Since the columns of $Q$ are LI eigenvectors of $A$, $Q$ is called a **modal matrix** of $A$. 

11.4. Diagonalization 571
Theorem 11.4.1 relates the diagonalizability of $A$ to the eigenvectors of $A$. With the help of Theorem 11.4.2, we will be able to relate the diagonalizability of $A$ to the eigenvalues of $A$ as well. When that is done we will turn to applications.

**Theorem 11.4.2** Distinct Eigenvalues, LI Eigenvectors

If an $n \times n$ matrix $A$ has distinct eigenvalues $\lambda_1, \ldots, \lambda_n$, then the corresponding eigenvectors $e_1, \ldots, e_n$ are LI.

Proof: We need to show that

$$c_1 e_1 + c_2 e_2 + \cdots + c_n e_n = 0$$  \hspace{1cm} (12)

holds only if $c_1 = c_2 = \cdots = c_n = 0$. Multiplying (12) by $A$ and noting that $A e_j = \lambda_j e_j$, we have

$$c_1 \lambda_1 e_1 + c_2 \lambda_2 e_2 + \cdots + c_n \lambda_n e_n = 0.$$  \hspace{1cm} (13)

Repeating the process gives

$$c_1 \lambda_1^2 e_1 + c_2 \lambda_2^2 e_2 + \cdots + c_n \lambda_n^2 e_n = 0,$$

$$\vdots$$

$$c_1 \lambda_1^{n-1} e_1 + c_2 \lambda_2^{n-1} e_2 + \cdots + c_n \lambda_n^{n-1} e_n = 0.$$  \hspace{1cm} (14)

Expressed in matrix form,

$$\begin{bmatrix}
1 & \cdots & 1 \\
\lambda_1 & \cdots & \lambda_n \\
\vdots & \cdots & \vdots \\
\lambda_1^{n-1} & \cdots & \lambda_n^{n-1}
\end{bmatrix}
\begin{bmatrix}
c_1 e_1 \\
c_2 e_2 \\
\vdots \\
c_n e_n
\end{bmatrix}
= 
\begin{bmatrix}
0 \\
0 \\
\vdots \\
0
\end{bmatrix}.$$  \hspace{1cm} (15)

The determinant of the coefficient matrix is a Vandermonde determinant, which (see Exercise 17 in Section 10.4) is nonzero if the $\lambda_j$'s are distinct. Since $\lambda_j$'s are indeed distinct by assumption, (15) admits the unique trivial solution $c_1 e_1 = 0, c_2 e_2 = 0, \ldots, c_n e_n = 0$, and since the $e_j$'s are nonzero (because they are eigenvectors) it follows that $c_1 = c_2 = \cdots = c_n = 0$, so $e_1, \ldots, e_n$ are LI.

From Theorems 11.4.1 and 11.4.2 we can draw the following conclusion.

**Theorem 11.4.3** Diagonalizability

If an $n \times n$ matrix has $n$ distinct eigenvalues, then it is diagonalizable.
As usual, be careful not to read converses into theorems when they are not stated. Specifically, Theorem 11.4.3 does not say that an \( n \times n \) matrix is diagonalizable if and only if it has \( n \) distinct eigenvalues.

Consider an application.

**EXAMPLE 1. A Problem in Chemical Kinetics.** We consider, here, a special class of chemical reactions known as *first-order reactions*. These reactions are governed by systems of linear, coupled, first-order ordinary differential equations. Specifically, suppose that \( X_1, \ldots, X_n \) are the chemical names of \( n \) reacting species (elements or molecules), that \( x_j(t) \) denotes the concentration of \( X_j \) (in suitable units) as a function of the time \( t \), and that the *rate constant* for the conversion of \( X_i \) to \( X_j \) is the positive constant \( k_{ij} \). For a two-component reaction, for example, denoted schematically in Fig. 1, this means that

\[
\begin{align*}
x'_1 &= -k_{21}x_1 + k_{12}x_2, \quad (16a) \\
 x'_2 &= k_{21}x_1 - k_{12}x_2. \quad (16b)
\end{align*}
\]

The first term on the right-hand side of (16a) accounts for the loss of \( X_1 \) due to the \( X_1 \rightarrow X_2 \) reaction; it is proportional to the concentration of \( X_1 \), namely \( x_1 \), and the constant of proportionality is the relevant rate constant \( k_{21} \). The second term on the right-hand side of (16a) accounts for the rate of gain of \( X_1 \) due to the reverse reaction \( X_2 \rightarrow X_1 \). A similar accounting holds for the terms in (16b).

The difficulty in solving (16), and similar systems for \( n \)-component reactions where \( n > 2 \), is due to the coupling. Equations (16) are coupled due to the \( k_{12}x_2 \) term in (16a) and the \( k_{21}x_1 \) term in (16b).

Let us solve (16) by diagonalization, if that is possible. In matrix form (16) is

\[
x' = Ax,
\]

where

\[
A = \begin{bmatrix} -k_{21} & k_{12} \\ k_{21} & -k_{12} \end{bmatrix}.
\]

The eigenvalues and eigenvectors of \( A \) are readily found to be

\[
\lambda_1 = 0, \quad \mathbf{e}_1 = \alpha \begin{bmatrix} k_{12} \\ k_{21} \end{bmatrix}; \quad \lambda_2 = -(k_{12} + k_{21}), \quad \mathbf{e}_2 = \beta \begin{bmatrix} 1 \\ -1 \end{bmatrix}.
\]

The \( \lambda_j \)'s are distinct, because \( k_{12} > 0 \) and \( k_{21} > 0 \), so Theorem 11.4.3 guarantees that \( A \) is diagonalizable. Alternatively, observe that the \( \mathbf{e}_j \)'s are necessarily LI because for them to be LD we would need \( k_{21} = -k_{12} \), which is impossible since \( k_{12} > 0 \) and \( k_{21} > 0 \). Their linear independence implies that \( A \) is diagonalizable, by Theorem 11.4.1.

Thus, if we set \( \mathbf{x} = \mathbf{Q}\tilde{\mathbf{x}} \), where (with \( \alpha = \beta = 1 \), say)

\[
\mathbf{Q} = [\mathbf{e}_1, \mathbf{e}_2] = \begin{bmatrix} k_{12} & 1 \\ k_{21} & -1 \end{bmatrix},
\]

then the preceding analysis assures us that

\[
\tilde{x}' = \mathbf{Q}^{-1}A\mathbf{Q}\tilde{x} = \mathbf{D}\tilde{x} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}\tilde{x}.
\]
Thus, we have the uncoupled system (which, of course, is our objective)

\[
\begin{align*}
\dot{x}_1' &= \lambda_1 x_1, \\
\dot{x}_2' &= \lambda_2 x_2,
\end{align*}
\]

the general solution of which is

\[
\begin{align*}
x_1 &= C_1 e^{\lambda_1 t} = C_1, \\
x_2 &= C_2 e^{\lambda_2 t} = C_2 e^{-(k_{12} + k_{21}) t}.
\end{align*}
\]

Finally, putting these expressions into \( x = Q\bar{x} \) gives

\[
\begin{bmatrix}
x_1 \\
x_2
\end{bmatrix} = \begin{bmatrix}
k_{12} & 1 \\
k_{21} & -1
\end{bmatrix} \begin{bmatrix}
C_1 \\
C_2 e^{-(k_{12} + k_{21}) t}
\end{bmatrix}
\]

or

\[
\begin{align*}
x_1(t) &= C_1 k_{12} + C_2 e^{-(k_{12} + k_{21}) t}, \\
x_2(t) &= C_1 k_{21} - C_2 e^{-(k_{12} + k_{21}) t}.
\end{align*}
\]

**COMMENT 1.** Here we have emphasized the mathematics rather than the chemistry and have assumed the rate constants to be known. A problem of importance to the chemist is the determination of those constants. Such determination normally involves a blend of the foregoing theory with suitable experiments.

**COMMENT 2.** The numbering of the eigenvalues and eigenvectors is immaterial. For instance, we could just as well take \( \lambda_1 = -(k_{12} + k_{21}) \) and \( \lambda_2 = 0 \). The final result, (24), would be the same. 

Theorem 11.4.3 revealed that diagonalizability is the typical case, the generic case, because an \( n \)th degree algebraic equation (namely, the characteristic equation of \( A \)) typically has distinct roots. Furthermore, every symmetric matrix is diagonalizable:

**THEOREM 11.4.4 Symmetric Matrices**

Every symmetric matrix is diagonalizable.

**Proof:** Theorem 11.4.1 states that \( A \) is diagonalizable if and only if it has \( n \) LI eigenvectors, and Theorem 11.3.4 assures us that every \( n \times n \) symmetric matrix has \( n \) orthogonal (and hence LI) eigenvectors.

Suppose that for a symmetric matrix \( A \) we use the normalized eigenvectors of \( A \) to form its modal matrix \( Q \) so that

\[
Q = [\hat{e}_1, \ldots, \hat{e}_n].
\]
Then, observe that

\[
Q^T Q = \begin{bmatrix}
\hat{e}_1^T \\
\vdots \\
\hat{e}_n^T
\end{bmatrix}
\begin{bmatrix}
\hat{e}_1 \\
\vdots \\
\hat{e}_n
\end{bmatrix}
= \begin{bmatrix}
\hat{e}_1^T \hat{e}_1 & \cdots & \hat{e}_1^T \hat{e}_n \\
\vdots & \ddots & \vdots \\
\hat{e}_n^T \hat{e}_1 & \cdots & \hat{e}_n^T \hat{e}_n
\end{bmatrix}

= \begin{bmatrix}
1 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & 1
\end{bmatrix} = I \tag{26}
\]

so that

\[
Q^{-1} = Q^T. \tag{27}
\]

Be sure to understand each step in (26). Q starts out as an \( n \times n \) matrix, but when we partition it into columns, as \([\hat{e}_1, \ldots, \hat{e}_n]\), it is then a \( 1 \times n \) matrix with elements \( \hat{e}_1, \ldots, \hat{e}_n \). To form \( Q^T \), we make the \( j \)th column of \( Q \), namely \( \hat{e}_j \), the \( j \)th row of \( Q^T \), and to put it into row format we need to write it as \( \hat{e}_j^T \) rather than \( \hat{e}_j \). Thus, working with the partitioned \( Q \) and \( Q^T \) matrices, the product to the right of the first equal sign in (26) is an \( n \times 1 \) matrix times a \( 1 \times n \) matrix, which product gives the \( n \times n \) matrix to the right of the second equal sign.

Understand also that (27) has nothing to do with the \( \hat{e}_j \)'s being eigenvectors. The steps in (26) rely only on the fact that the columns of \( Q \) are ON. Any square matrix, the columns of which are ON, satisfies (27) and is called an orthogonal matrix. Of course, the property (27) is very nice because if we ever need the inverse of \( Q \) it is simply \( Q^T \).

Getting back to diagonalization, note that if \( A \) is symmetric then \( Q^{-1} A Q = D \) is diagonal whether or not the columns of the modal matrix \( Q \) are normalized. However, let us agree (at least within this text) to always normalize them, if \( A \) is symmetric, so as to have access to the property (27) if we need it.

Let us close with one more application.

**EXAMPLE 2. A Free-Vibration Problem.** Consider a mass \( m \) constrained by two mechanical springs, of stiffnesses \( k_1 \) and \( k_2 \), as sketched in Fig. 2. Imagine Fig. 2 as a view looking down on the apparatus, which lies in a horizontal plane on a frictionless table. In the configuration shown, the springs are neither stretched nor compressed, and \( m \) is at rest in static equilibrium. However, if some initial displacement and/or velocity is imparted to \( m \), some motion, no doubt vibrational, will result, and it is that motion that we wish to determine.

The first step in the formulation is to introduce a coordinate system. A reasonable choice is the Cartesian system shown in Fig. 2, with its origin at the equilibrium position of the mass \( m \) (which we regard as a "point mass").

If \( m \) is at some point \( x, y \) other than the origin, then one or both springs will be stretched or compressed and will exert forces \( F_1 \) and \( F_2 \) on \( m \) (Fig. 2). The magnitude

---

*Recall our encounter with orthogonal matrices in the optional Section 10.7.*
The Eigenvalue Problem

The magnitude of $F_1$ is

$$
\|F_1\| = k_1 \text{ times the stretch in spring } \# 1
= k_1 \left\{ \sqrt{(x-(-1))^2 + (y-0)^2} - 1 \right\}
= k_1 \left\{ \sqrt{(x+1)^2 + y^2} - 1 \right\}.
$$  \hspace{1cm} (28)

If we multiply this magnitude by a unit vector directed from $(x, y)$ toward $(-1,0)$, we will have $F_1$. The vector from $(x, y)$ to $(-1,0)$ is $(-1-x)i + (0-y)j$, where $i, j$ are unit base vectors in the $x, y$ directions, respectively. Normalizing that vector gives the desired unit vector

$$
\hat{F}_1 = -\frac{(1+x)i + yj}{\sqrt{(1+x)^2 + y^2}},
$$  \hspace{1cm} (29)

so

$$
F_1 = \|F_1\| \hat{F}_1 = -k_1 \left[ \frac{\sqrt{(x+1)^2 + y^2} - 1}{\sqrt{(x+1)^2 + y^2}} \right] (x+1)i + (y+1)j.
$$  \hspace{1cm} (30)

In like manner, we find that

$$
F_2 = \|F_2\| \hat{F}_2 = -k_2 \left[ \frac{\sqrt{(x+1)^2 + (y+1)^2} - \sqrt{2}}{\sqrt{(x+1)^2 + (y+1)^2}} \right] (x+1)i + (y+1)j.
$$  \hspace{1cm} (31)

According to Newton’s second law,

$$
mx'' = F_x \quad \text{and} \quad my'' = F_y,
$$  \hspace{1cm} (32)

where $F_x$ is the sum of the $x$ components of $F_1$ and $F_2$, and $F_y$ is the sum of the $y$ components of $F_1$ and $F_2$. Thus, the governing equations of motion are

\begin{align*}
mx'' &= -k_1 \left[ \frac{\sqrt{(x+1)^2 + y^2} - 1}{\sqrt{(x+1)^2 + y^2}} \right] (x+1) \\
&\quad -k_2 \left[ \frac{\sqrt{(x+1)^2 + (y+1)^2} - \sqrt{2}}{\sqrt{(x+1)^2 + (y+1)^2}} \right] (x+1), \hspace{1cm} (33a)
\end{align*}

\begin{align*}
my'' &= -k_1 \left[ \frac{\sqrt{(x+1)^2 + y^2} - 1}{\sqrt{(x+1)^2 + y^2}} \right] y \\
&\quad -k_2 \left[ \frac{\sqrt{(x+1)^2 + (y+1)^2} - \sqrt{2}}{\sqrt{(x+1)^2 + (y+1)^2}} \right] (y+1). \hspace{1cm} (33b)
\end{align*}

The latter, coupled, nonlinear differential equations are, clearly, quite intractable analytically. Two possibilities present themselves. First, if we assign numerical values to $m, k_1, k_2, x(0), y(0), x'(0), y'(0)$, then we can generate $x(t)$ and $y(t)$ by one of the numerical methods studied in Chapter 6 (such as fourth-order Runge–Kutta integration) or using computer software (such as the Maple dsolve command).
Second, we can limit our attention to small motions, motions that remain close to the equilibrium position at the origin: $|x| \ll 1$ and $|y| \ll 1$. In that case we can simplify (33) in essentially the same way that we can simplify the nonlinear differential equation

$$x'' + \frac{g}{l} \sin x = 0$$

(34)

governing the motion of a pendulum (Fig. 4): for small motions, near the equilibrium point $x = 0$,

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots$$

(Taylor series)

for $|x| \ll 1$ so (34) can be approximated by the simple linear equation

$$x'' + \frac{g}{l} x = 0.$$  

(35)

We will follow the same steps for (33), but instead of the Taylor series in one variable (about $x = 0$),

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \cdots,$$

(36)

we will need the Taylor series in two variables (about $x = y = 0$):

\[
f(x, y) = f(0, 0) + f_x(0, 0)x + f_y(0, 0)y + \frac{1}{2!} \left[ f_{xx}(0, 0)x^2 + 2f_{xy}(0, 0)xy + f_{yy}(0, 0)y^2 \right] + \cdots
\]

(37)

because the right-hand sides of (33a,b) are functions of $x$ and $y$. First, let $f(x, y)$ in (37) be the right-hand side of (33a). Then (37) gives

\[
f(x, y) = 0 + \left( -k_1 - \frac{k_2}{2} \right) x + \left( -\frac{k_2}{2} \right) y + \cdots
\]

$$\approx -\left( k_1 + \frac{k_2}{2} \right) x - \frac{k_2}{2} y.$$  

(38)

Next, let $f(x, y)$ in (37) be the right-hand side of (33b). Then (37) gives

\[
f(x, y) = 0 - \frac{k_2}{2} x - \frac{k_2}{2} y + \cdots
\]

$$\approx -\frac{k_2}{2} x - \frac{k_2}{2} y.$$  

(39)

With these approximations (linearizations) of the right-hand sides of (33a,b), we have the linearized equations

$$m x'' = -\left( k_1 + \frac{k_2}{2} \right) x - \frac{k_2}{2} y,$$

(40a)

$$m y'' = -\frac{k_2}{2} x - \frac{k_2}{2} y.$$  

(40b)

*Taylor series in more than one variable is discussed in Chapter 13.
or

\[ x'' + Ax = 0, \quad (41) \]

where

\[
x = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix}, \quad A = \begin{bmatrix} \frac{2k_1 + k_2}{2m} & k_2 \\ k_2 & \frac{k_2}{2m} \end{bmatrix}.
\]

Let us solve (41) by diagonalization. Observe that \( A \) is symmetric (even though there is no "physical" symmetry to be seen in Fig. 2). For definiteness, let us set

\[ m = 1, \quad k_1 = 3, \quad k_2 = 4. \]

Then the eigenvalues and normalized eigenvectors of \( A \) are

\[ \lambda_1 = 1, \quad \vec{e}_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ -2 \end{bmatrix}; \quad \lambda_2 = 6, \quad \vec{e}_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix}. \]

With

\[
Q = [\vec{e}_1, \vec{e}_2] = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix},
\]

set \( x = Q\vec{x} \) in (41). Thus,

\[ Q\vec{x}'' + AQ\vec{x} = 0 \]

and hence

\[ \vec{x}'' + Q^{-1}AQ\vec{x} = 0 \]

or

\[ \vec{x}'' + D\vec{x} = 0, \]

where

\[ D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 6 \end{bmatrix}. \]

In scalar form, (48) gives the uncoupled equations

\[ \vec{x}'' + \vec{x} = 0, \]

\[ \vec{y}'' + 6\vec{y} = 0 \]

with general solution [expressed in the \( A \sin(\omega t + \phi) \) form]:

\[
\vec{x} = A_1 \sin(t + \phi_1), \quad \vec{y} = A_2 \sin(\sqrt{6}t + \phi_2),
\]

where the amplitudes \( A_1, A_2 \) and phase angles \( \phi_1, \phi_2 \) are the four constants of integration. To return to the original \( x, y \) variables, write

\[
\begin{bmatrix} x \\ y \end{bmatrix} = Q\vec{x} = \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} \vec{x} \\ \vec{y} \end{bmatrix} = A_1 \begin{bmatrix} 1/\sqrt{5} \sin(t + \phi_1) \\ -2/\sqrt{5} \sin(t + \phi_1) \end{bmatrix} + A_2 \begin{bmatrix} 2/\sqrt{5} \sin(\sqrt{6}t + \phi_2) \\ 1/\sqrt{5} \sin(\sqrt{6}t + \phi_2) \end{bmatrix}
\]

\[ = A_1 \vec{e}_1 \sin(t + \phi_1) + A_2 \vec{e}_2 \sin(\sqrt{6}t + \phi_2) \]}
or, in scalar form,

\[
x(t) = C_1 \sin(t + \varphi_1) + 2C_2 \sin(\sqrt{6}t + \varphi_2),
\]

\[
y(t) = -2C_1 \sin(t + \varphi_1) + C_2 \sin(\sqrt{6}t + \varphi_2),
\]

where \( C_1 = A_1 / \sqrt{5} \) and \( C_2 = A_2 / \sqrt{5} \), for brevity.

COMMENT 1. It is seen from (52) that the general solution is a linear combination of two orthogonal modes, as in Example 2 of Section 11.3. The low mode is a vibration along the \( \hat{e}_1 \) direction, at a frequency that is the square root of \( \lambda_1 \), and the high mode is a vibration along the \( \hat{e}_2 \) direction, at a frequency that is the square root of \( \lambda_2 \), as summarized in Fig. 5. In this example the orthogonality of the modes is geometric since the low- and high-mode motions are 90° apart; in Example 2 of Section 11.3, the orthogonality is more mathematical (\( \hat{e}_1 \cdot \hat{e}_2 \) being zero) than geometric.

COMMENT 2. Why are the directions of the low and high modes, shown in Fig. 5, physically reasonable? Recall that the natural frequency of the classical harmonic mechanical oscillator, governed by the equation \( m\ddot{x} + kx = 0 \), is \( \sqrt{k/m} \), which increases as the stiffness \( k \) increases. It should be clear intuitively that the \( \hat{e}_2 \) axis is the line of maximum stiffness (of the two-spring system) and the \( \hat{e}_1 \) axis is the line of minimum stiffness, at least roughly speaking. Strikingly, the mathematics reveals that these directions are necessarily 90° apart.

COMMENT 3. Why do we show the positive \( \bar{x} \) and \( \bar{y} \) coordinate axes as being in the \( \hat{e}_1 \) and \( \hat{e}_2 \) directions, respectively, in Fig. 5? Because if we set \( \bar{x} = 1 \) and \( \bar{y} = 0 \) in

\[
\begin{bmatrix}
x \\
y
\end{bmatrix} = \begin{bmatrix}
1/\sqrt{5} & 2/\sqrt{5} \\
-2/\sqrt{5} & 1/\sqrt{5}
\end{bmatrix} \begin{bmatrix}
\bar{x} \\
\bar{y}
\end{bmatrix}
\]

we get \( [x, y]^T = [1/\sqrt{5}, -2/\sqrt{5}]^T = \hat{e}_1 \), and if we set \( \bar{x} = 0 \) and \( \bar{y} = 1 \) we get \( [x, y]^T = [2/\sqrt{5}, 1/\sqrt{5}]^T = \hat{e}_2 \). In fact, if you studied the optional Section 10.7, you will appreciate that the effect of the change of variables \( x = QX \), where \( Q \) is an orthogonal matrix with its determinant equal to +1, as here, is a pure rotation of axes. Thus, we have the vivid visual image of the elements of the coupling matrix varying as we rotate the Cartesian coordinate system (somewhat like looking into a kaleidoscope), until the off-diagonal terms become zero and the equations uncouple.

COMMENT 4. In principle, it would have been best to choose the \( \bar{x}, \bar{y} \) coordinate system in the first place, but its orientation was not known. Thus, we chose any \( x, y \) system, to get started, and then used the method of diagonalization to find the optimal \( \bar{x}, \bar{y} \) coordinate system. \( \blacksquare \)

**Closure.** From a mathematical viewpoint, this section is about finding an \( n \times n \) matrix \( Q \), given an \( n \times n \) matrix \( A \), such that \( Q^{-1}AQ = D \) is diagonal. We find that in the generic case \( A \) is diagonalizable: it is diagonalizable if and only if it has \( n \) LI eigenvectors, and it is diagonalizable if it has \( n \) distinct eigenvalues or is symmetric. \( Q \) can be made up of columns which are the eigenvectors of \( A \), and the diagonal elements of \( D \) are the corresponding eigenvalues. In the event that \( A \)
is symmetric, we suggest always normalizing the eigenvectors that are the columns of \( Q \), so that \( Q \) will admit the useful property \( Q^{-1} = Q^T \); that is, so that \( Q \) will be an orthogonal matrix.

From an applications standpoint, we look only at the use of diagonalization in uncoupling systems of coupled differential equations, but additional applications are to be found in the exercises and in the next two sections.

It turns out that even if an \( n \times n \) matrix \( A \) cannot be diagonalized, it can be triangularized. That is, a generalized modal matrix \( P \) can be found for \( A \) so that

\[
P^{-1}AP = J
\]

is triangular. Called the Jordan normal form, or simply the Jordan form, for \( A \), \( J \) is upper triangular, with zeros above its main diagonal — except for 1’s immediately above one or more diagonal elements. This case is discussed briefly in the exercises.

### EXERCISES 11.4

1. Diagonalize each of the given \( A \) matrices. That is, determine matrices \( Q \) and \( D \) such that \( Q^{-1}AQ = D \) is diagonal.

   Also, work out \( Q^{-1} \) and verify that \( Q^{-1}AQ \) is diagonal and that its diagonal elements are the eigenvalues of \( A \). If \( A \) is not diagonalizable, state that and give the reason.

   (a) \[
   \begin{bmatrix}
   2 & -3 \\
   0 & 0 
   \end{bmatrix}
   \]

   (b) \[
   \begin{bmatrix}
   2 & 4 \\
   -1 & -2 
   \end{bmatrix}
   \]

   (c) \[
   \begin{bmatrix}
   1 & 0 \\
   1 & 0 
   \end{bmatrix}
   \]

   (d) \[
   \begin{bmatrix}
   0 & 1 \\
   1 & 1 
   \end{bmatrix}
   \]

   (e) \[
   \begin{bmatrix}
   0 & 1 & 1 \\
   0 & 0 & 0 
   \end{bmatrix}
   \]

   (f) \[
   \begin{bmatrix}
   2 & 0 & 0 \\
   0 & 1 & 1 \\
   0 & 1 & 1 
   \end{bmatrix}
   \]

   (g) \[
   \begin{bmatrix}
   2 & 1 & -1 \\
   1 & 4 & 3 \\
   -1 & 3 & 4 
   \end{bmatrix}
   \]

   (h) \[
   \begin{bmatrix}
   0 & 1 & 0 \\
   1 & 0 & 0 \\
   0 & 0 & 0 
   \end{bmatrix}
   \]

   (i) \[
   \begin{bmatrix}
   4 & 0 & 0 & 0 \\
   0 & 1 & 1 & 0 \\
   0 & 1 & 1 & 0 \\
   0 & 0 & 0 & 2 
   \end{bmatrix}
   \]

   (j) \[
   \begin{bmatrix}
   1 & 1 & 1 & 1 \\
   1 & 1 & 1 & 1 \\
   1 & 1 & 1 & 1 \\
   1 & 1 & 1 & 1 
   \end{bmatrix}
   \]

2. Use the method of diagonalization to obtain the general solution of the given system of differential equations, where primes denote \( d/dt \).

   (a) \[
   \begin{align*}
   x' &= x + y \\
y' &= x + y 
   \end{align*}
   \]

   (b) \[
   \begin{align*}
   x' &= x + 4y \\
y' &= x + y 
   \end{align*}
   \]

   (c) \[
   \begin{align*}
   x'' &= 2x + 4y \\
y'' &= x - y 
   \end{align*}
   \]

   (d) \[
   \begin{align*}
   x' &= 2x + y \\
y' &= x + 2y + 3z = 0 \\
z'' &= y + 2z = 0 
   \end{align*}
   \]

   (e) \[
   \begin{align*}
   x' &= 4x + y + 3z \\
y' &= x - z \\
z' &= 3x - y + 4z 
   \end{align*}
   \]

   (f) \[
   \begin{align*}
   x' &= -y + z \\
y' &= -x - 3y - 4z \\
z'' &= x - 4y - 3z 
   \end{align*}
   \]


4. We see from Fig. 5 that the line of action of the high mode falls between the two springs. Show that that situation holds for all possible combinations of stiffnesses \( k_1 \) and \( k_2 \).

5. Determine (as in Example 2) the natural frequencies and mode shapes for each of the systems shown below. You need carry only three or four significant figures. Each spring is of unit length.

   (a) \[
   \begin{align*}
   m = k_1 = k_2 = 1. \quad k_3 = 10 
   \end{align*}
   \]

   (b) \[
   \begin{align*}
   m = k_1 = 4, \quad k_2 = 3, \quad k_3 = 1 
   \end{align*}
   \]
6. Show why the second equality in (9) is true.

7. (Application to exponentiation) Diagonalization can be helpful in raising a square matrix to a large power. Specifically, show that if $A$ is diagonalizable so that $Q^{-1}AQ = D$, then

$$A^n = QD^nQ^{-1}, \quad (7.1)$$

the advantage being that $D^n$ is simply

$$D^n = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}^m = \begin{bmatrix} \lambda_1^m & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n^m \end{bmatrix}. \quad (7.2)$$

8. Use (7.1), above, to evaluate $A^{1000}$, where

(a) $A = \begin{bmatrix} 0 & 2 & 2 \\ 2 & 0 & 2 \\ 2 & 2 & 0 \end{bmatrix}$  \hspace{1cm} (b) $A = \begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix}$

9. (Application to principal inertias and principal axes) Two vectors of importance in studying the dynamics of a rigid body $B$ are the moment of momentum $H_p$ and the angular velocity $\omega$ of $B$. These are related according to $H_p = I\omega$, where $I$ is the inertia matrix. Written out, we have

$$\begin{bmatrix} (H_p)_x \\ (H_p)_y \\ (H_p)_z \end{bmatrix} = \begin{bmatrix} I_{xx} & -I_{xy} & -I_{xz} \\ -I_{yx} & I_{yy} & -I_{yz} \\ -I_{zx} & -I_{zy} & I_{zz} \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \\ \omega_z \end{bmatrix}, \quad (9.1)$$

where $P$ is the origin of a Cartesian $x, y, z$ coordinate system (see the figure), and

$$I_{xx} = \int_B (y^2 + z^2) \, dm, \quad I_{xy} = I_{yx} = \int_B xy \, dm$$
$$I_{yy} = \int_B (x^2 + z^2) \, dm, \quad I_{xz} = I_{zx} = \int_B xz \, dm$$
$$I_{zz} = \int_B (x^2 + y^2) \, dm, \quad I_{yz} = I_{zy} = \int_B yz \, dm.$$

(9.2)

$I_{xx}, I_{yy}, I_{zz}$ are known as the moments of inertia of $B$ about the $x, y, z$ axes, respectively, and $I_{xy}, I_{xz}, I_{yz}$ are known as the products of inertia of $B$, dm in (9.2) is “d(mass)”.

Now, the relation (9.1) and hence the subsequent dynamic analysis (which will be of no concern here) will be simplest if $I$ is diagonal, i.e., if all of the products of inertia are zero. In general, it is too difficult to see, by inspection, how to orient the coordinate axes to achieve this result. Instead, we go ahead and choose some $x, y, z$ reference frame, compute the nine inertia components, and then rotate to a new Cartesian $\tilde{x}, \tilde{y}, \tilde{z}$ frame so as to diagonalize $I$. That is, if $x = Q\tilde{x}$, where $x = [x, y, z]^T$ and $\tilde{x} = [\tilde{x}, \tilde{y}, \tilde{z}]^T$, then $H_p = QH_p$ and $\omega = Q\tilde{\omega}$ so that $H_p = \tilde{I}\tilde{\omega}$ becomes

$$Q\tilde{H}_p = \tilde{I}Q\tilde{\omega} \quad \text{and} \quad \tilde{H}_p = (Q^{-1}IQ)\tilde{\omega}. \quad (9.3)$$

where $Q^{-1}IQ = \tilde{I}$ is diagonal. That such diagonalization is possible follows from the fact that $I$ is symmetric since

$I_{xy} = I_{yx}, I_{xz} = I_{zx},$ and $I_{yz} = I_{zy}$.

$I_{xx}, I_{yy}, I_{zz}$ are called the principal inertias of $B$ (with respect to coordinates with origin at $P$), and the $\tilde{x}, \tilde{y}, \tilde{z}$ axes are called the principal axes. We now state the problem: compute the principal inertias and determine the principal axes for each of the following bodies; sketch the principal axes. In each case $B$ can be assumed, for simplicity, to be infinitely thin, with mass density $\sigma$ mass units per unit area.
(Jordan form) Recall that an $n \times n$ matrix $A$ is diagonalizable if and only if it has $n$ LI eigenvectors. Thus, not all matrices can be diagonalized. However, our experience in Chapters 8–11 has shown that triangular matrices are "almost as nice" as diagonal ones, and it turns out that even if $A$ cannot be diagonalized, it can be triangularized.

More specifically, if $A$ is diagonalizable and its modal matrix is $Q = [e_1, \ldots, e_n]$, where $e_j$'s are the LI eigenvectors of $A$, then $Q^{-1}AQ = D$ is diagonal, with $d_{jj}$ equal to the $j$th eigenvalue of $A$. But suppose $A$ is not diagonalizable. Though it does not have $n$ LI eigenvectors, it has $n$ LI generalized eigenvectors $e_1, \ldots, e_n$ (defined below) and if its generalized modal matrix is $P = [e_1, \ldots, e_n]$, then $P^{-1}AP = J$, where $J$ is the Jordan form for $A$. $J$ will be upper triangular, with the eigenvalues of $A$ on its diagonal and zeros everywhere else except perhaps immediately above repeated eigenvalues. For details, we refer you to M. Greenberg, Advanced Engineering Mathematics, 1st ed. (Englewood Cliffs, NJ: Prentice Hall, 1988). Here, we will merely go through an example with you and, at the end, ask you to supply various steps.

Consider, as a representative case,

$$
A = \begin{bmatrix}
2 & -1 & 2 & 0 \\
0 & 3 & -1 & 0 \\
0 & 1 & 1 & 0 \\
0 & 1 & -3 & 5
\end{bmatrix}.
$$

(10.1)

Calculation reveals these eigenvalues and eigenvectors:

$$
\lambda_1 = \lambda_2 = \lambda_3 = 2, \quad e = [1, 0, 0, 0]^T \equiv e_1,
\lambda_4 = 5, \quad e = [0, 0, 0, 1]^T \equiv e_4.
$$

(10.2)

In this case the eigenvalue of multiplicity three, $\lambda = 2$, contributes only one LI eigenvector, instead of three, so we end up with only two LI eigenvectors, $e_1$ and $e_4$, instead of the four that are needed for diagonalization. Thus, $A$ is not diagonalizable. But we can find "generalized eigenvectors" $e_2$ and $e_3$, associated with $\lambda = 2$, such that $e_1, e_2, e_3, e_4$ are LI. Noting that $e_1$ satisfies

$$
(A - \lambda_1 I)e_1 = 0,
$$

(10.3)

we introduce $e_2, e_3$ so as to satisfy

$$
(A - \lambda_1 I)e_2 = e_1, \quad (A - \lambda_1 I)e_3 = e_2.
$$

(10.4)

(10.5)

With $\lambda_1$ and $e_1$ given in (10.2), solution of (10.4) and (10.5) by Gauss elimination gives

$$
e_2 = [\alpha, 1, 1, 2/3]^T, \quad (10.6)
$$

$$
e_3 = [\beta, \alpha + 2, \alpha + 1, 2\alpha/3 + 5/9]^T, \quad (10.7)
$$

where $\alpha, \beta$ are arbitrary. Take $\alpha = \beta = 0$, say. Then, with the generalized modal matrix

$$
P = [e_1, e_2, e_3, e_4] = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 2 & 0 \\
0 & 1 & 1 & 0 \\
0 & 2/3 & 5/9 & 1
\end{bmatrix},
$$

(10.8)

we obtain

$$
P^{-1}AP = \begin{bmatrix}
\lambda_1 & 1 & 0 & 0 \\
0 & \lambda_1 & 1 & 0 \\
0 & 0 & \lambda_4 & 0 \\
0 & 0 & 0 & \lambda_4
\end{bmatrix} = \begin{bmatrix}
2 & 1 & 0 & 0 \\
0 & 2 & 1 & 0 \\
0 & 0 & 2 & 0 \\
0 & 0 & 0 & 5
\end{bmatrix} = J.
$$

(10.9)

(a) Derive the eigenvalues and eigenvectors given in (10.2).
(b) Use Gauss elimination to derive (10.6) from (10.4), and (10.7) from (10.5).
11.5 Application to First-Order Systems with Constant Coefficients (Optional)

In Section 11.4 we studied diagonalization, and showed how to use that method to uncouple and solve systems of differential equations. In this section we continue that discussion, but this time our emphasis is on the theory of differential equations rather than on diagonalization.

Consider the initial-value problem

\[ x'_1 = a_{11}x_1 + \cdots + a_{1n}x_n + f_1(t); \quad x_1(0) = c_1 \]

\[ \vdots \]

\[ x'_n = a_{n1}x_1 + \cdots + a_{nn}x_n + f_n(t); \quad x_n(0) = c_n \]

where the \( a_{ij} \)'s are constants. Or, in matrix form,

\[ x' = Ax + f(t); \quad x(0) = c. \]  

Unlike Section 11.4, here we allow for forcing functions \( f_1(t), \ldots, f_n(t) \).

To solve the first-order system (2) by diagonalization, we suppose that \( A \) has \( n \) LI eigenvectors, a modal matrix \( Q \), and eigenvalues \( \lambda_1, \ldots, \lambda_n \) that are not necessarily distinct. Setting

\[ x(t) = Q\tilde{x}(t), \]

where the \( \tilde{x} \)'s are the generalized eigenvectors corresponding to \( \lambda \).

Then multiply each term in (10.10) by \( A - \lambda \mathbf{I} \). Then multiply each term in the resulting equation by \( A - \lambda \mathbf{I} \). Explain why the resulting set of four equations implies that \( \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0 \).

HINT: Suppose that

\[ \alpha_1 e_1 + \alpha_2 e_2 + \alpha_3 e_3 + \alpha_4 e_4 = 0. \]  

(10.10)

Then multiply each term in (10.10) by \( A - \lambda \mathbf{I} \). Then multiply each term in the resulting equation by \( A - \lambda \mathbf{I} \). Again multiply each term in the resulting equation by \( A - \lambda \mathbf{I} \), where \( k < K \). Within that eigenspace there can be found \( k \) LI eigenvectors of \( A \), say \( e_1, \ldots, e_k \). Vectors \( e_{k+1} \) through \( e_K \) satisfying

\[ (A - \lambda \mathbf{I})e_{k+1} = e_k, \]

(10.11)

\[ (A - \lambda \mathbf{I})e_{k+2} = e_{k+1}, \]

(10.12)

\[ \vdots \]

\[ (A - \lambda \mathbf{I})e_K = e_{K-1} \]

are the generalized eigenvectors corresponding to \( \lambda \).

Consider the system of differential equations

\[ x' = Ax, \]  

(10.12)

where the prime denotes \( d/dt \) and \( A \) is given by (10.1). Under the change of variables \( x = P\tilde{x} \), reduce (10.12) to the Jordan form

\[ \tilde{x}' = P^{-1}AP\tilde{x} = J\tilde{x}, \]  

(10.13)

where \( J \) is given in (10.9). Solve the triangular system (10.13) for \( \tilde{x}(t) \), and thus obtain the general solution \( x(t) = P\tilde{x}(t) \) of (10.12).
(2) becomes
\[ Q\ddot{x}' = AQ\dot{x} + f(t); \quad Q\dot{x}(0) = c \] (4)
or, equivalently,
\[ \dot{x}' = Q^{-1}AQ\dot{x} + Q^{-1}f(t); \quad \dot{x}(0) = Q^{-1}c, \] (5)
where \( Q^{-1} \) necessarily exists because the columns of \( Q \) are LI. We know that
\[ Q^{-1}AQ = D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}, \] (6)
and if we denote \( Q^{-1}f(t) \) as \( \tilde{f}(t) \) and \( Q^{-1}c \) as \( \tilde{c} \), for brevity, then
\[ \ddot{x}' = D\dot{x} + \tilde{f}(t); \quad \dot{x}(0) = \tilde{c} \] (7)
or, returning to scalar form,
\[ \ddot{x}'_1 = \lambda_1\dot{x}_1 + \tilde{f}_1(t); \quad \dot{x}_1(0) = \tilde{c}_1 \\
\vdots \\
\ddot{x}'_n = \lambda_n\dot{x}_n + \tilde{f}_n(t); \quad \dot{x}_n(0) = \tilde{c}_n. \] (8)
Each of these equations is first-order linear, like equation (2) in Section 2.2, and its solution is given by (24) in Section 2.2.2 [with \( x \to t, y \to \tilde{x}_j, p(x) \to -\lambda_j \), and so on]:
\[ \tilde{x}_1(t) = \tilde{c}_1 e^{\lambda_1 t} + \int_0^t e^{\lambda_1 (t-\tau)} \tilde{f}_1(\tau) d\tau; \] (9)
\[ \vdots \]
\[ \tilde{x}_n(t) = \tilde{c}_n e^{\lambda_n t} + \int_0^t e^{\lambda_n (t-\tau)} \tilde{f}_n(\tau) d\tau. \]
If we define
\[ e^{Dt} \equiv \begin{bmatrix} e^{\lambda_1 t} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2 t} & \cdots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & e^{\lambda_n t} \end{bmatrix}, \] (10)
then we can express (9) as
\[ \ddot{x}(t) = e^{Dt}\tilde{c} + \int_0^t e^{D(t-\tau)}\tilde{f}(\tau) d\tau. \] (11)
Finally, since \( \tilde{x} = Q^{-1} x, \tilde{c} = Q^{-1} c, \) and \( \tilde{f} = Q^{-1} f, \)

\[
x(t) = Q e^{Dt} Q^{-1} c + \int_0^t Q e^{D(t-\tau)} Q^{-1} f(\tau) d\tau
\]

(12)

is the unique solution of (1). Note the order: \( e^{Dt} \tilde{c}, \) not \( \tilde{c} e^{Dt}, \) in (11).

**EXAMPLE 1.** Use (12) to solve the system

\[
x' = x + 4y - 4t^2 - 3; \quad x(0) = 2
\]
\[
y' = x + y - t^2 + 2t - 3; \quad y(0) = 3.
\]

Then

\[
A = \begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix}, \quad f(t) = \begin{bmatrix} -4t^2 - 3 \\ -t^2 + 2t - 3 \end{bmatrix}, \quad c = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.
\]

(14)

The eigenvalues and eigenvectors of \( A \) are found to be

\[
\lambda_1 = 3, \quad e_1 = \alpha \begin{bmatrix} 2 \\ 1 \end{bmatrix}; \quad \lambda_2 = -1, \quad e_1 = \beta \begin{bmatrix} -2 \\ 1 \end{bmatrix}.
\]

(15)

Since these \( e_j \)'s are LI, \( A \) is diagonalizable so (12) applies. With

\[
D = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}, \quad Q = \begin{bmatrix} 2 & -2 \\ 1 & 1 \end{bmatrix}, \quad Q^{-1} = \begin{bmatrix} 1/4 & 1/2 \\ -1/4 & 1/2 \end{bmatrix},
\]

(16)

(12) gives

\[
x(t) = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} e^{3t} & 0 \\ 0 & e^{-t} \end{bmatrix} \begin{bmatrix} 2/4 \\ 1 \\ 1/2 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix}
\]

\[
+ \int_0^t \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} e^{3(t-\tau)} & 0 \\ 0 & e^{-(t-\tau)} \end{bmatrix} \begin{bmatrix} 1/4 \\ 1/2 \\ 1/2 \end{bmatrix} \begin{bmatrix} -4\tau^2 - 3 \\ -\tau^2 + 2\tau - 3 \end{bmatrix} d\tau
\]

(17)

so

\[
x(t) = 3 + 5/2 e^{3t} - 7/2 e^{-t}, \quad y(t) = t^2 + 5/4 e^{3t} + 7/4 e^{-t}
\]

(18)

is the desired solution.

**EXAMPLE 2.** Can (12) be used for the system

\[
x'' + 3x' + 2x - y = e^t
\]
\[
y'' - y' + 5x' - x - 6y = t
\]

(19a)

(19b)
Yes, if we can reduce (19) to a first-order system. How to do that is described in Section 3.9 (see Example 5, therein, and the paragraph preceding that example). Specifically, set \( x' = u \) and \( y' = v \). Then (19) can be re-expressed as

\[
\begin{align*}
x' &= u, & \text{(by definition)} \\
u' &= -3u - 2x + y + e^t, & \text{[from (19a)]} \\
y' &= v, & \text{(by definition)} \\
v' &= v - 5u + x + 6y + t & \text{[from (19b)]}
\end{align*}
\]

or, in matrix form, as

\[
\begin{bmatrix}
x' \\
u' \\
y' \\
v'
\end{bmatrix} =
\begin{bmatrix}
0 & 1 & 0 & 0 \\
-2 & -3 & 1 & 0 \\
0 & 0 & 0 & 1 \\
1 & -5 & 6 & 1
\end{bmatrix}
\begin{bmatrix}
x \\
u \\
y \\
v
\end{bmatrix} +
\begin{bmatrix}
0 \\
e^t \\
0 \\
t
\end{bmatrix}.
\]

The latter is a first-order system, to which (12) could be applied. 

In deriving (12) an interesting quantity arose, the exponential matrix function \( e^B \), where \( B = \{b_{ij}\} \) is an \( n \times n \) diagonal matrix:

\[
e^B = \begin{bmatrix}
e^{b_{11}} & 0 & \ldots & 0 \\
0 & e^{b_{22}} & \ldots & \vdots \\
\vdots & \ddots & \ddots & \vdots \\
0 & \ldots & e^{b_{nn}}
\end{bmatrix} \quad \text{(B diagonal)}
\]

More generally, let \( B \) be any \( n \times n \) matrix, not necessarily diagonal. Following the familiar Taylor series formula,

\[
e^x = 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \ldots,
\]

it seems reasonable to define

\[
e^B \equiv I + \frac{1}{1!}B + \frac{1}{2!}B^2 + \ldots
\]

Fine, but observe that the right-hand side of (23) is an infinite series of matrices, which we have not yet defined. Mimicking the usual definition of convergence for series of scalars, we define an infinite series of matrices \( \sum_{j=1}^{\infty} A_j \) as the limit of the sequence of partial sums \( S_N \), where \( S_N = \sum_{j=1}^{N} A_j \). That is,

\[
\sum_{j=1}^{\infty} A_j = \lim_{N \to \infty} S_N.
\]
The infinite series is said to converge if the limit on the right exists and to diverge if that limit does not exist. Finally, observe that $\lim_{N \to \infty} S_N$ is the limit of a sequence of matrices, which we have not yet defined so we are not done.

Let $C_1, C_2, \ldots$ be a sequence of $m \times n$ matrices, with $(c_{ij})_n$ as the $i, j$ element of $C_n$. We say that the sequence converges to a matrix $C = \{c_{ij}\}$ if

$$\lim_{n \to \infty} (c_{ij})_n = c_{ij}$$

for each $i, j$, and we denote such convergence by writing either $\lim_{n \to \infty} C_n = C$ or $C_n \to C$ as $n \to \infty$. If the sequence does not converge, then it is said to diverge.

**EXAMPLE 3.**

$$\begin{bmatrix} 2 + \frac{3}{n} & e^{-\frac{n}{n+2}} \\ 1 & \frac{4}{n+2} \end{bmatrix} \to \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix},$$

(25)

but

$$\begin{bmatrix} 7 - \frac{3n}{3} & 2 + e^{-\frac{4n}{n}} \\ 3 & \frac{1}{n} \end{bmatrix} \text{ and } \begin{bmatrix} 2 & 5 \\ \sin n & \frac{1}{n+2} \end{bmatrix}$$

(26)

diverge because $\lim_{n \to \infty} (7 - \frac{3n}{3})$ and $\lim_{n \to \infty} \sin n$ do not exist. 

With (23), let us return to (12) and consider the combination $Q e^{Dt} Q^{-1}$:

$$Q e^{Dt} Q^{-1} = Q \left( I + Dt + \frac{1}{2!} D^2 t^2 + \cdots \right) Q^{-1}$$

$$= I + QDQ^{-1} t + \frac{1}{2!} QD^2 Q^{-1} t^2 + \cdots$$

$$= I + QDQ^{-1} t + \frac{1}{2!} QDQ^{-1} QDQ^{-1} t^2 + \cdots,$$

(27)

and recalling that $QDQ^{-1} = A$, this series gives

$$Q e^{Dt} Q^{-1} = I + At + \frac{1}{2!} A^2 t^2 + \cdots = e^{At}.$$ 

(28)

Similarly, $Q e^{D(t-r)} Q^{-1} = e^{A(t-r)}$ so (12) can be expressed in the equivalent form

$$x(t) = e^{At} c + \int_0^t e^{A(t-\tau)} f(\tau) d\tau,$$

(29)

How are we to evaluate the exponentials in (29)? We could use the series formula (23), but it is more efficient to reverse (28) and to use $e^{At} = Q e^{Dt} Q^{-1}$ and $e^{A(t-r)} = Q e^{D(t-r)} Q^{-1}$ because $e^{Dt}$ is given simply by (10), and similarly
for $e^{D(t-\tau)}$. That is, computationally, we continue to rely on (12) rather than (29). Nonetheless, (29) is of considerable interest because it shows how the solution

$$x(t) = e^{At}c + \int_0^t e^{A(t-\tau)}f(\tau)d\tau$$

(30)

of the single initial-value problem

$$x'(t) = Ax + f(t); \quad x(0) = c$$

(31)

can be generalized to the solution (29) of (2) using the exponential matrix function.\footnote{Of course, it is more natural to write the first term on the right-hand side of (30) as $ce^{At}$, but we have written it as $e^{At}c$ to emphasize the correspondence between (29) and (30).}

**Closure.** There were two objectives in this section. One was to derive the matrix solution (12) of the initial-value problem (2), and the other was to introduce the exponential matrix function $e^B$ for any $n \times n$ matrix $B$. In fact, one can introduce other functions of square matrices as well: sines, cosines, fractional powers, and so on. For discussion of such functions we refer you to Wylie and Barrett.\footnote{C. R. Wylie and L. C. Barrett. *Advanced Engineering Mathematics*, 5th ed. (New York: McGraw–Hill, 1982).}

**Computer software.** Using Maple, the exponential matrix function can be evaluated by means of the exponential command within the linalg package. For instance, to evaluate $e^{A^t}$, where

$$A = \begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix},$$

enter

```maple
with(linalg):
```

and return, then

```maple
B := matrix(2,2,[t,4*t,t,t]):
```

and return. Then,

```maple
exponential(B);
```

and return gives the output

$$\begin{bmatrix} \frac{1}{2}e^{(-t)} + \frac{1}{2}e^{(3t)} & e^{(3t)} - e^{(-t)} \\ \frac{1}{4}e^{(3t)} - \frac{1}{4}e^{(-t)} & \frac{1}{2}e^{(-t)} + \frac{1}{2}e^{(3t)} \end{bmatrix}$$

Actually, when you enter the $B$ matrix you may prefer to end with a semicolon rather than a colon because then the $B$ matrix will be printed, and you can inspect it to see if its elements were entered correctly.
11.6 Quadratic Forms (Optional)

A function of the form
\[ f(x_1, x_2) = a_{11}x_1^2 + a_{22}x_2^2 + 2a_{12}x_1x_2 \]  
(1)

is called a quadratic form in the variables \( x_1 \) and \( x_2 \). Similarly,
\[ f(x_1, x_2, x_3) = a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + 2a_{12}x_1x_2 + 2a_{13}x_1x_3 + 2a_{23}x_2x_3 \]  
(2)

### EXERCISES 11.5

1. Solve, using (12). If (12) does not apply, explain why it does not. NOTE: You may need to first re-express the problem in the standard form (1).
   
   (a) \( x' = x + y; \quad x(0) = 0 \)
   \( y' = x + y + e^{3t}; \quad y(0) = 0 \)
   
   (b) \( x' = 2x + 3y + 1; \quad x(0) = 0 \)
   \( y' = x - y; \quad y(0) = 0 \)
   
   (c) \( x' = 2x - y - 3t + 1; \quad x(0) = 0 \)
   \( y' = x - y; \quad y(0) = 0 \)
   
   (d) \( x' = x + 2y - t - 1; \quad x(0) = 0 \)
   \( y' = 4x + 8y - 4t - 8; \quad y(0) = 3 \)
   
   (e) \( x' = \sin t - y; \quad x(0) = 0 \)
   \( y' = -9x + 4; \quad y(0) = 1 \)
   
   (f) \( x' = x - 8y; \quad x(0) = 0 \)
   \( y' = -x - y - 3t - 3; \quad y(0) = 0 \)
   
   (g) \( x' = 3x + y + 4e^t; \quad x(0) = 1 \)
   \( y' = 3x + y - 6; \quad y(0) = -1 \)
   
   (h) \( x' = x - y - 3t; \quad x(0) = 0 \)
   \( x' + y' - 5x - 2y = 5; \quad y(0) = 3 \)
   
   (i) \( x'' + 3x' + 2x = e^t; \quad x(0) = x'(0) = 0 \)
   
   (j) \( x'' + 2x' + x = 0; \quad x(0) = 2, x'(0) = 9 \)
   
   (k) \( x_1' = 4x_1 + x_2 + 3x_3; \quad x_1(0) = 1 \)
   \( x_2' = x_1 - x_2 + 6; \quad x_2(0) = 0 \)
   \( x_3' = 3x_1 - x_2 + 4x_3; \quad x_3(0) = 0 \)
   
   (l) \( x' = x + y + z - 1; \quad x(0) = 0 \)
   \( y' = x + y + z; \quad y(0) = 0 \)
   \( z' = x + y + z; \quad z(0) = 0 \)

2. (a)–(l) Same as Exercise 1, but drop the initial conditions and find the general solution instead.

3. Evaluate \( e^A \) for the given \( A \) matrix. HINT: If \( A \) is diagonalizable you can use the formula
\[ e^A = Qe^{D}Q^{-1} \]
(3.1)
[i.e., equation (28) with \( t = 1 \)]. Whether or not \( A \) is diagonalizable, you can use the series definition
\[ e^A = I + \frac{1}{1!}A + \frac{1}{2!}A^2 + \cdots \]
(3.2)

Generally, (3.2) is unwieldy because one needs to sum an infinite series of matrices. However, if \( A \) happens to be nilpotent then the series reduces to a finite number of terms.

(b) \[
\begin{bmatrix}
1 & 2 \\
0 & 3
\end{bmatrix}
\]

(c) \[
\begin{bmatrix}
0 & 4 \\
9 & 0
\end{bmatrix}
\]

(d) \[
\begin{bmatrix}
2 & 0 \\
7 & 3
\end{bmatrix}
\]

(f) \[
\begin{bmatrix}
0 & 0 & 1 & 2 & 1 \\
0 & 0 & 1 & 2 & 3
\end{bmatrix}
\]

(h) \[
\begin{bmatrix}
1 & 1 & 3 & 1 & 2 & 2 \\
1 & 2 & 2 & 1 & 3 & 2
\end{bmatrix}
\]

(i) \[
\begin{bmatrix}
0 & 1 & 2 & 3 & 0 & 0 & 0 \\
0 & 0 & 4 & 5 & 0 & 0 & 0 \\
0 & 0 & 0 & 6 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 2 & 1
\end{bmatrix}
\]

4. (a)–(l) Evaluate \( e^A \) using computer software, where \( A \) is given in the corresponding part of Exercise 3.
is a quadratic form in \(x_1, x_2, x_3\), and so on for any number of variables \(x_1, \ldots, x_n\). The \(a_{ij}\)'s are constants (where \(a_{ij}\) is the coefficient of the product \(x_i x_j\)), and the 2's are included for our subsequent convenience.

It turns out that a quadratic form \(f(x_1, \ldots, x_n)\) can be expressed concisely, in matrix notation, as

\[
f(x_1, \ldots, x_n) = x^T A x,
\]

where \(x = [x_1, \ldots, x_n]^T\), and \(A = \{a_{ij}\}\) is an \(n \times n\) matrix. For instance, suppose that \(n = 2\) and

\[
f(x_1, x_2) = 2x_1^2 - x_2^2 + 6x_1x_2.
\]

Writing out the right-hand side of (3) gives

\[
x^T A x = [x_1, x_2] \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = [x_1, x_2] \begin{bmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{bmatrix}
\]

\[
= a_{11}x_1^2 + a_{22}x_2^2 + (a_{12} + a_{21})x_1x_2.
\]

Comparing (4) and (5), we see that \(a_{11} = 2, a_{22} = -1, \) and \(a_{12} + a_{21} = 6\). Since the sum \(a_{12} + a_{21}\) is prescribed, but not \(a_{12}\) and \(a_{21}\) individually, we are free to ask \(a_{12}\) and \(a_{21}\) to be equal, so that \(a_{12} = a_{21} = 3\). Then the right-hand side of (5) becomes \(2a_{11}x_1^2 + 2a_{22}x_2^2 + 2a_{12}x_1x_2\), as written in (1). The benefit in asking \(a_{12}\) to equal \(a_{21}\) is that then the \(A\) matrix in (3) is symmetric, and symmetric matrices have advantages over nonsymmetric matrices, as we have seen in Sections 11.3 and 11.4. Similarly for \(n = 3, 4, \ldots\).

Thus, if we ask the \(A\) matrix in (3) to be symmetric, for convenience, then (3) gives (1) for \(n = 2\), (2) for \(n = 3\), and so on.

**EXAMPLE 1.** Let

\[
f(x_1, \ldots, x_4) = x_1^2 + 5x_2^2 + 6x_1x_3 + 10x_3x_4.
\]

Then \(a_{22} = 1, a_{33} = 5, 2a_{13} = 6\) so \(a_{13} = 3, 2a_{23} = -1\) so \(a_{23} = -\frac{1}{2}, 2a_{34} = 10\) so \(a_{34} = 5\), and all the other \(a_{ij}\)'s are zero. Thus

\[
A = \begin{bmatrix} 0 & 0 & 3 & 0 \\ 0 & 1 & -\frac{1}{2} & 0 \\ 3 & -\frac{1}{2} & 5 & 5 \\ 0 & 0 & 5 & 0 \end{bmatrix}
\]

A quadratic form is said to be **canonical** if all mixed terms (such as \(x_1x_2, x_1x_3\), and \(x_2x_3\)) are absent, that is, if \(a_{ij} = 0\) for \(i \neq j\). Thus,

\[
f(x_1, \ldots, x_n) = a_{11}x_1^2 + a_{22}x_2^2 + \cdots + a_{nn}x_n^2
\]
is canonical, and its associated matrix

\[
A = \begin{bmatrix}
  a_{11} & 0 & \cdots & 0 \\
  0 & a_{22} & \cdots & 0 \\
  \vdots & \ddots & \ddots & \vdots \\
  0 & \cdots & 0 & a_{nn}
\end{bmatrix}
\]

is diagonal. For instance, (4) is not canonical, but \( f(x_1, x_2) = 5x_1^2 - 7x_2^2 \) and \( f(x_1, x_2, x_3) = 6x_1^2 + x_3^2 \) are.

There is interest in being able to reduce a given quadratic form to canonical form (i.e., to its simplest form) through a linear change of variables

\[
x = Q\tilde{x}
\]

from \( x_1, \ldots, x_n \) to \( \tilde{x}_1, \ldots, \tilde{x}_n \). Putting (7) into (3) gives

\[
f = (Q\tilde{x})^T A (Q\tilde{x}) = \tilde{x}^T (Q^T A Q) \tilde{x}.
\]

Thus, given \( A \), we wish to determine a \( Q \) matrix such that \( Q^T A Q \) is diagonal, for then (8) will be canonical in the new variables \( \tilde{x}_1, \ldots, \tilde{x}_n \). Theorem 11.4.1 tells us that \( Q^{-1}AQ \) will be a diagonal matrix, with its diagonal elements being the eigenvalues of \( A \), if \( A \) has \( n \) LI eigenvectors and if these eigenvectors are used as the columns of \( Q \). But since \( A \) is symmetric it has \( n \) orthogonal (and hence LI) eigenvectors, and if these are normalized and used as the columns of \( Q \) then \( Q^T = Q^{-1} \), and

\[
Q^T A Q = Q^{-1} A Q = \begin{bmatrix}
  \lambda_1 & 0 & \cdots & 0 \\
  0 & \lambda_2 & \cdots & 0 \\
  \vdots & \ddots & \ddots & \vdots \\
  0 & \cdots & 0 & \lambda_n
\end{bmatrix} = D,
\]

where the \( \lambda_j \)'s are the eigenvalues of \( A \), so (8) is the desired canonical form of \( f \),

\[
f = \lambda_1 \tilde{x}_1^2 + \cdots + \lambda_n \tilde{x}_n^2.
\]

**EXAMPLE 2.** Reduce

\[
f(x_1, x_2) = 3x_1^2 + 3x_2^2 + 2x_1x_2
\]

to canonical form. First, identify \( A \):

\[
A = \begin{bmatrix}
  3 & 1 \\
  1 & 3
\end{bmatrix}
\]

*We do not claim that \( Q \) must be a modal matrix of \( A \) for \( Q^T A Q \) to be diagonal.*
Next, determine the eigenvalues and normalized eigenvectors of $A$. These are found to be

$$
\begin{align*}
\lambda_1 &= 4, \quad \hat{e}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \\
\lambda_2 &= 2, \quad \hat{e}_2 = \frac{1}{\sqrt{-1}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.
\end{align*}
$$

(Whether we call $\lambda_1 = 4$ and $\lambda_2 = 2$, or vice versa, will not matter.) According to (10), the desired canonical form of $f$ is then

$$
f = \lambda_1 \bar{x}_1^2 + \lambda_2 \bar{x}_2^2 = 4\bar{x}_1^2 + 2\bar{x}_2^2.
$$

It looks like we are done, so why do we need to know $\hat{e}_1, \hat{e}_2$? To know the connection between $x_1, x_2$ and $\bar{x}_1, \bar{x}_2$. Specifically,

$$
x = Q\bar{x} = [\hat{e}_1, \hat{e}_2]\bar{x},
$$
or

$$
\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}
$$

so that

$$
\begin{align*}
x_1 &= \frac{1}{\sqrt{2}} \bar{x}_1 + \frac{1}{\sqrt{2}} \bar{x}_2 \\
x_2 &= \frac{1}{\sqrt{2}} \bar{x}_1 - \frac{1}{\sqrt{2}} \bar{x}_2.
\end{align*}
$$

Or, if we wish to have it the other way around, we can use $\bar{x} = Q^{-1}x = Q^T x$. 

Let us summarize:

### Theorem 11.6.1 Canonical Form

A quadratic form $f(x_1, \ldots, x_n) = x^T A x$ can be reduced to the canonical form $\lambda_1 \bar{x}_1^2 + \cdots + \lambda_n \bar{x}_n^2$ by introducing the change of variables $x = Q\bar{x}$, where the $\lambda_j$'s are the eigenvalues of $A$ and the columns of $Q$ are the corresponding normalized eigenvectors of $A$. The reverse transformation is given by $\bar{x} = Q^T x$.

A quadratic form $x^T A x$ is classified as **positive definite** (i.e., "definitely positive") if $x^T A x > 0$ for all $x \neq 0$, and as **negative definite** if $x^T A x < 0$ for all $x \neq 0$. Likewise, $A$ is classified as positive (negative) definite if the quadratic form $x^T A x$ is positive (negative) definite.

### Theorem 11.6.2 Definiteness

Let $A$ be symmetric. Then $A$ and its quadratic form $x^T A x$ are positive (negative) definite if every eigenvalue of $A$ is positive (negative).
Proof: It is simplest to work with the canonical form $x^T A x = \lambda_1 \overline{x}_1^2 + \cdots + \lambda_n \overline{x}_n^2$. Since the latter is a sum of squares, it is evident that if $\lambda_1, \ldots, \lambda_n$ are all positive (negative), then $x^T A x$ is positive (negative) for all $x \neq 0$. It remains to show that $\bar{x} \neq 0$ if and only if $x \neq 0$. Since $x = Q \bar{x}$ and $\bar{x} = Q^{-1} x = Q^T x$ (i.e., $Q$ is nonsingular), $x = 0$ implies that $x = Q \bar{x} = 0$, and $x = 0$ implies that $\bar{x} = Q^T 0 = 0$. \qed

For instance, $f$ and $A$ in Example 2 are positive definite because $\lambda_1 = 4 > 0$ and $\lambda_2 = 2 > 0$.

**EXAMPLE 3.** If
\[
f(x_1, x_2, x_3) = 2x_1^2 + 4x_2^2 + 4x_3^2 + 2x_1 x_2 - 2x_1 x_3 + 6x_2 x_3,
\]
then
\[
A = \begin{bmatrix} 2 & 1 & -1 \\ 1 & 4 & 3 \\ -1 & 3 & 4 \end{bmatrix} \quad \text{and} \quad \lambda_1 = 7, \ \lambda_2 = 3, \ \lambda_3 = 0.
\]

It might appear that $f$ is positive definite but it is not, because $\lambda_3$ is not positive. Being zero is not good enough, for remember that being positive definite means that $x^T A x > 0$ for all $x \neq 0$ or, equivalently, $\bar{x}^T (Q^T A Q) \bar{x} > 0$ for all $\bar{x} \neq 0$. Yet,
\[
f = 7\bar{x}_1^2 + 3\bar{x}_2^2 + 6\bar{x}_3^2
\]
is zero if
\[
\bar{x} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 8 \end{bmatrix},
\]
for instance. \qed

Let us conclude this section with a physical application.

**EXAMPLE 4.** *Buckling Load.* Three rigid rods, each of length $L$, are pinned at their ends, with the end $A$ constrained to move in a frictionless vertical slot, as shown in Fig. 1. The three lateral springs, each of stiffness $k$, are unstretched and uncompressed when the system is undellected (i.e., when $x = y = 0$), and the middle spring is attached at the middle of the middle rod. We find that as the load $P$ is slowly increased the deflection remains zero ($x = y = z = 0$) so that $x = y = z = 0$ constitutes an equilibrium state of the system. Eventually, however, when $P$ reaches a critical value, say $P_{cr}$, the system "buckles," that is, collapses. The problem posed here is to determine $P_{cr}$ in terms of the given quantities. It will be convenient to work with the potential energy $V$ of the system, and to use the physical principle that the system will arrange itself so as to minimize its potential $V$.

Recall, from elementary physics, that the potential energy stored in a spring of stiffness $k$, deflected by an amount $x$, is $\frac{1}{2} k x^2$. Further, note that we may associate a potential

![Figure 1. Buckling.](image)
with the load \( P \) by regarding it as being due to a weight \( P \) sitting on the slider. Then the gravitational potential of the weight is \(-Pz\), if we define the potential to be zero when \( z = 0 \). Thus,
\[
V = \frac{1}{2}kx^2 + \frac{1}{2}ky^2 + \frac{1}{2}k\left(\frac{x+y}{2}\right)^2 - Pz.
\]
(20)

But \( x, y, z \) are not independent variables. For instance, we can express \( z \) in terms of \( x \) and \( y \) through the geometry. Specifically, if we introduce the angles \( \alpha, \beta, \gamma \), shown in Fig. 1, then
\[
z + L \cos \alpha + L \cos \beta + L \cos \gamma = 3L.
\]
(21)

For the onset of the buckling \( x, y, z \) are all infinitesimal (arbitrarily small), so
\[
\cos \alpha = \sqrt{1 - \sin^2 \alpha} = \sqrt{1 - \left(\frac{x}{L}\right)^2}
\]
\[
= 1 - \frac{1}{2} \left(\frac{x}{L}\right)^2 - \frac{1}{8} \left(\frac{x}{L}\right)^4 - \cdots \quad \text{(Taylor series)}
\]
\[
\sim 1 - \frac{1}{2} \left(\frac{x}{L}\right)^2 \quad \text{as } x \to 0.
\]
(22)

Similarly for \( \cos \beta \) and \( \cos \gamma \) so (21) becomes
\[
z + L \left[1 - \frac{1}{2} \left(\frac{x}{L}\right)^2\right] + L \left[1 - \frac{1}{2} \left(\frac{y}{L}\right)^2\right] + L \left[1 - \frac{1}{2} \left(\frac{y}{L}\right)^2\right] \sim 3L
\]
(23)

or, upon simplification,
\[
z \sim \frac{x^2 - xy + y^2}{L}
\]
(24)

so that
\[
V(x, y) \sim \left(\frac{5k}{8} - \frac{P}{L}\right) x^2 + \left(\frac{5k}{8} - \frac{P}{L}\right) y^2 + \left(\frac{k}{4} + \frac{P}{L}\right) xy.
\]
(25)

The right-hand side of (25) is a quadratic form
\[
f(x, y) = \left(\frac{5k}{8} - \frac{P}{L}\right) x^2 + \left(\frac{5k}{8} - \frac{P}{L}\right) y^2 + \left(\frac{k}{4} + \frac{P}{L}\right) xy,
\]
(26)

with the associated matrix
\[
A = \begin{bmatrix}
\frac{5k}{8} & \frac{k}{4} + \frac{P}{L} \\
\frac{5k}{8} & \frac{k}{4} + \frac{P}{L} \\
\end{bmatrix}
\]
(27)

The crucial point is whether or not \( f \) is positive definite for if \( f(x, y) \) is positive definite, then \( V(x, y) \) has a minimum at \( x = y = 0 \) and the undeflected equilibrium configuration \( x = y = 0 \) is stable. Otherwise, the equilibrium configuration will be unstable and the system will buckle.
To assess the positive definiteness of $f$, we need merely evaluate the eigenvalues of $A$. These are found to be

$$\lambda_1 = \frac{3k}{2} \left( \frac{1}{3} - \frac{P}{kL} \right), \quad \lambda_2 = \frac{k}{2} \left( \frac{3}{2} - \frac{P}{kL} \right)$$

(28)

from which we see that $f$ is positive definite (i.e., both $\lambda_j$'s are positive) if and only if $P/(kL) < \frac{1}{3}$. Thus, the critical load, the "buckling load," is $P_{cr} = kL/3$.

**COMMENT 1.** If $P/(kL) < \frac{1}{3}$, then both $\lambda_j$'s are positive in the canonical version $\lambda_1 \bar{x}^2 + \lambda_2 \bar{y}^2$ of $V$ so the graph of $V$ has a "valley" at the origin ($\bar{x} = \bar{y} = 0$ or, equivalently $x = y = 0$). If $P/(kL) > \frac{1}{3}$, then both $\lambda_j$'s are negative so the graph of $V$ has a "hill" at the origin. And if $\frac{1}{3} < P/(kL) < \frac{3}{2}$, then $\lambda_1 < 0$ and $\lambda_2 > 0$ so the graph of $V$ has a "saddle" at the origin (i.e., like an equestrian saddle it has a valley in one direction and a hill in the other). In the case of the valley $V$ has a *minimum* at the origin so the equilibrium solution $x = y = 0$ is stable and buckling does not occur. In the other two cases (hill and saddle) $V$ does not have a minimum at the origin, so the equilibrium solution $x = y = 0$ is unstable and buckling occurs. The borderline between these two cases gives the critical buckling load.

**COMMENT 2.** Physically, the idea is that as the system begins to deflect, the springs gain potential energy while the weight $P$ loses potential energy. For instability (buckling) we need the loss to exceed the gain. Since the loss is proportional to $P$, one anticipates the existence of a critical value $P_{cr}$ such that if $P > P_{cr}$, the loss just balances the gain, and if $P < P_{cr}$, then the loss exceeds the gain. 

**Closure.** The chief result of this section is given in Theorem 11.6.1, that a quadratic form can be reduced to canonical form by a normalized modal matrix. In addition, Theorem 11.6.2 tells us that a quadratic form $x^T A x$ is positive definite if every eigenvalue of $A$ is positive, and negative definite if every eigenvalue is negative. Positive and negative definiteness is central when we discuss the maxima and minima of functions of several variables in Chapter 13.

**EXERCISES 11.6**

1. For each of the following quadratic forms in $n$ variables determine the associated symmetric $A$ matrix.

   (a) $2x_1^2 + 4x_2^2 + x_1x_2$  \hspace{1cm} (b) $x_1^2 + 2x_1x_3 + 2x_2x_3$  \hspace{1cm} (c) $x_1^2 + x_2^2 + x_1x_2$  \hspace{1cm} (d) $x_1^2 - 4x_2^2 + 3x_1x_2$  \hspace{1cm} (e) $x_1^2 + x_2^2 + 2x_1x_3$  \hspace{1cm} (f) $4x_1x_4 + 4x_2x_3$  \hspace{1cm} (g) $3x_1x_2$  \hspace{1cm} (h) $4x_1x_2 - 2x_2^2$  \hspace{1cm} (i) $-2x_1x_3$  \hspace{1cm} (j) $x_1^2 + x_1x_2 + x_1x_3 + x_2x_3$  \hspace{1cm} (k) $6x_2x_3$  \hspace{1cm} (l) $2x_1x_2 + 2x_1x_3 + 2x_2x_3 - 4x_1^2$  \hspace{1cm} (n) $3$

2. (a)–(l) Reduce each quadratic form in Exercise 1 to the canonical form (10) by means of a normalized modal matrix transformation $Q$. Further, classify the quadratic form and the associated symmetric $A$ matrix as positive definite or negative definite, where applicable.

3. We state, without proof, that a necessary and sufficient con-
dation for a quadratic form \( f = x^T A x \), and its associated symmetric matrix \( A \), to be positive definite is that

\[
\begin{vmatrix}
  a_{11} & a_{12} & a_{13} \\
  a_{21} & a_{22} & a_{23} \\
  a_{31} & a_{32} & a_{33}
\end{vmatrix} > 0,
\]

Apply this condition to each of the following \( A \) matrices, and compare your conclusion with that drawn from direct examination of the eigenvalues.

(a) \[
\begin{bmatrix}
  4 & 1 \\
  1 & 4
\end{bmatrix}
\]

(b) \[
\begin{bmatrix}
  1 & 1 \\
  1 & 1
\end{bmatrix}
\]

(c) \[
\begin{bmatrix}
  2 & -1 \\
  -1 & 2
\end{bmatrix}
\]

(d) \[
\begin{bmatrix}
  0 & 1 \\
  1 & 0
\end{bmatrix}
\]

(e) \[
\begin{bmatrix}
  6 & 0 & 0 \\
  0 & 2 & 0 \\
  0 & 0 & 3
\end{bmatrix}
\]

(f) \[
\begin{bmatrix}
  0 & 0 & 1 \\
  0 & 2 & 0 \\
  1 & 0 & 0
\end{bmatrix}
\]

(g) \[
\begin{bmatrix}
  -2 & 1 & 2 \\
  1 & -2 & 2 \\
  2 & 2 & 1
\end{bmatrix}
\]

(h) \[
\begin{bmatrix}
  0 & 0 & 1 \\
  0 & 0 & 1 \\
  0 & 0 & 0
\end{bmatrix}
\]

4. Reduce the following quadratic equations to the canonical form \( ax^2 + by^2 = c \), and hence decide whether they correspond to ellipses or hyperbolas. Sketch their graphs, showing both the \( x \), \( y \) and \( \bar{x}, \bar{y} \) axes, as well as the intercepts on the \( \bar{x}, \bar{y} \) axes. HINT: Recall that \((\bar{x}/A)^2 + (\bar{y}/B)^2 = 1\) is the equation of an ellipse with intercepts at \( \pm A \) on the \( \bar{x} \) axis and \( \pm B \) on the \( \bar{y} \) axis (or a circle of radius \( A \) if \( A = B \)). Further, \((\bar{x}/A)^2 - (\bar{y}/B)^2 = 1\) is the equation of a hyperbola with intercepts at \( \pm A \) on the \( \bar{x} \) axis, and \((\bar{y}/B)^2 - (\bar{x}/A)^2 = 1\) is the equation of a hyperbola with intercepts at \( \pm B \) on the \( \bar{y} \) axis.

(a) \( 3x^2 + 2y^2 - 2xy = 6 \)

(b) \( xy = 1 \)

(c) \( 4y^2 + 3xy = 1 \)

(d) \( x^2 + y^2 - 10xy = 4 \)

(e) \( x^2 + y^2 + 2xy = 4 \)  \( \text{This case will be found to correspond to a limiting case. Explain what we mean by that.)} \)

5. (Completing Squares) The successful \( Q \) matrix, in (7), is not necessarily a modal matrix of \( A \). Given \( f = x_1^2 + x_2^2 + x_1x_2 \), for example, suppose that we proceed, instead, by "completing squares:"

\[
f = x_1^2 + x_2^2 + x_1x_2
\]

\[
= (x_1^2 + x_1x_2 + \frac{1}{4}x_2^2) + x_2^2 - \frac{1}{4}x_2^2
\]

so the transformation

\[
x_1 = x_1 + \frac{1}{2}x_2, \quad x_2 = x_2
\]

reduces \( f \) to the canonical form \( f = \frac{x_1^2}{4} + \frac{3}{4}x_2^2 \).

(a) Show that the matrix \( Q \) corresponding to (5.2) is not a modal matrix.

(b) Reduce \( f = x_1^2 + x_2^2 + 4x_1^2 + 2x_2x_3 - x_1x_3 \) to canonical form by completing squares.

(c) Repeat part (b), for \( f = x_1^2 + 4x_1x_2 + x_2x_3 \).

(d) Repeat part (b), for \( f = 4x_1^2 + 2x_1x_2 + x_2x_3 \).

(e) Repeat part (b), for \( f = x_1^2 + 4x_1x_2 \).

(f) Does the method of completing squares work for \( f = 2x_1x_2 \)? Explain.

6. Rework Example 4 for the case where the middle spring is removed. Show that the buckling load is then \( R_{cr} = \frac{1}{3}kL \).

Note that this result is the same as in Example 4, where the middle spring is included. Explain, in physical terms, why this is so.

---

**Chapter 11 Review**

The matrix eigenvalue problem is the search for nontrivial solutions of \( Ax = \lambda x \) or, equivalently,

\[
(A - \lambda I)x = 0;
\]

that is, solutions other than the trivial solution \( x = 0 \). The eigenvalues are found
by setting
\[ \det(A - \lambda I) = 0, \]  
(2)
which condition guarantees the existence of nontrivial solutions of (1). Known as the characteristic equation of \( A \), (2) is an \( n \)-degree polynomial equation, which always has at least one and at most \( n \) distinct roots. For each eigenvalue \( \lambda_j \) thus found, the solution of \( (A - \lambda_j I)x = 0 \) by Gauss elimination gives the corresponding eigenvectors \( e_j \).

Rather than being rare, the case of symmetric matrices is common in applications (as noted, for instance, in Example 2 of Section 11.3, Example 1 in Section 11.4, and in all of Section 11.6). If an \( n \times n \) matrix \( A \) is symmetric, then all of its eigenvalues are real, eigenvectors corresponding to distinct eigenvalues are orthogonal, and its eigenvectors provide an orthogonal basis for \( n \)-space.

A pattern emerges insofar as choice of basis. Namely, when a basis is needed, to expand vectors, the most convenient basis to use is probably the basis provided by the eigenvectors of the \( A \) matrix to be found within the given problem. For instance, we do that to study the stability of the equilibrium solution of the Markov process (Example 4, Section 11.2), to solve the nonhomogeneous equation \( Ax = Ax + c \) (Section 11.3.2), and to prove the convergence of the power method (Exercise 12, Section 11.3). The reason that eigenvector bases are convenient is that if we multiply a vector equation by \( A \), then we need to evaluate the vectors \( Ae_j \). If the \( e_j \)'s are eigenvectors of \( A \), then \( Ae_j \) is simply the single term \( \lambda_j e_j \); if not, it is a linear combination of the \( n \) base vectors, \( e_1, \ldots, e_n \).

In Section 11.4 we study the diagonalization of an \( n \times n \) matrix \( A \). There, Theorem 11.4.1 is the most significant result because it gives a necessary and sufficient condition for \( A \) to be diagonalizable (namely, that it have \( n \) LI eigenvectors), and it tells us how to choose \( Q \) so that
\[ Q^{-1}AQ = D \]  
(3)
is diagonal. Namely, if we use the \((n \text{ LI})\) eigenvectors of \( A \) as the columns of \( Q \), then \( D = \{d_{ij}\} \) is diagonal, with \( d_{jj} = \lambda_j \). Further, Theorem 11.4.3 gives a sufficient condition for diagonalizability, that \( A \) have \( n \) distinct eigenvalues. Since the generic case is for the characteristic equation to have distinct roots, the generic case is for a given \( n \times n \) matrix to be diagonalizable. If \( A \) is symmetric, then it is diagonalizable whether or not it has \( n \) distinct eigenvalues because every \( n \times n \) symmetric matrix has \( n \) orthogonal (and hence LI) eigenvectors. For symmetric matrices we urge you to use the normalized eigenvectors of \( A \) to form its modal matrix \( Q \) because then \( Q \) admits the useful property that
\[ Q^{-1} = Q^T, \]  
(4)
that is, the inverse of \( Q \) is simply its transpose.

Even if \( A \) is not diagonalizable, it can be reduced to Jordan normal form. That is, a generalized modal matrix \( P \) can be found so that
\[ P^{-1}AP = J \]  
(5)
is triangular; $J$ is upper triangular with zeros above its main diagonal except for 1's immediately above one or more diagonal elements. That case is left for the exercises.

Whereas in Section 11.4 we concentrate on the concept of diagonalization, and show how to use diagonalization to uncouple systems of differential equations, in Section 11.5 we shift our focus to differential equation theory itself, using matrix theory and diagonalization as tools, and solve the general system of $n$ first-order nonhomogeneous linear differential equations with constant coefficients.

In the final section, 11.6, we show that a quadratic form can be expressed in matrix terminology as $x^{T}Ax$, where $A$ is symmetric, and can be reduced to canonical form $\lambda_{1}x_{1}^{2} + \cdots + \lambda_{n}x_{n}^{2}$ by the change of variables $x = Q\tilde{x}$, where $Q$ is a normalized modal matrix of $A$. That is, $Q = [e_{1}, \ldots, e_{n}]$, where $e_{j}$ is a normalized eigenvector corresponding to the eigenvalue $\lambda_{j}$.

Applications of the eigenvalue problem are extensive. Examples studied in this chapter are drawn from the areas of oscillation theory, systems of coupled differential equations, and population dynamics, with other applications covered in the exercises. In Chapter 13 we use the theory of quadratic forms to help us classify the extrema of functions of several variables.
Chapter 12

Extension to Complex Case (Optional)

PREREQUISITE: Familiarity with the algebra of complex numbers, covered in Section 21.2.

12.1 Introduction

In Chapters 8–11 all scalars are understood to be real: the coefficients in systems of linear equations, the components of vectors, the elements of matrices, and so on. In some applications, however, perhaps more so in physics and chemistry than in engineering, complex numbers enter. For example, you may have met the Euler angles $\theta, \phi, \psi$ used to specify the orientation of a rigid body such as a gyroscope. Alternatively, it is sometimes advantageous to employ so-called Cayley–Klein parameters, and in doing so one meets the Pauli spin matrices

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

(1)

of which $\sigma_y$ is seen to have complex elements.

The purpose of this chapter is to indicate the changes that need to be made in extending our vector and matrix systems so as to include complex numbers. Although this could have been done from the start, it was felt that the gain in simplicity, in the preceding chapters, offsets the need for this special chapter, which we make quite brief.

12.2 Complex $n$-Space

All scalars in the vector space $\mathbb{R}^n$ defined in Section 9.4 (namely, the scalars that multiplied vectors and the scalar components of the vectors themselves) are real.
Chapter 12. Extension to Complex Case

If we allow these scalars to be complex, then in place of \( \mathbb{R}^n \) we have the **complex \( n \)-space** denoted here as \( \mathbb{C}^n \):

\[
\mathbb{C}^n = \{(a_1, \ldots, a_n) \mid a_1, \ldots, a_n \text{ complex numbers} \}.
\]  (1)

The definitions \( u + v \equiv (u_1 + v_1, \ldots, u_n + v_n) \), \( \alpha u \equiv (\alpha u_1, \ldots, \alpha u_n) \), \( 0 \equiv (0, \ldots, 0) \), and \( -u \equiv (-u_1, \ldots, -u_n) \) are the same as for \( \mathbb{R}^n \), except that now the scalars are complex numbers. From these definitions the same properties follow as in (10) in Section 9.4.

However, the Euclidean dot product \( u \cdot v = u_1v_1 + \cdots + u_nv_n \), which we adopted for real vector space, is unacceptable because the resulting norm \( \|u\| = \sqrt{u \cdot u} = \sqrt{u_1^2 + \cdots + u_n^2} \) fails to display the key properties expected of a norm, in particular, the nonnegativeness condition

\[
\|u\| > 0 \quad \text{for all } u \neq 0,
\]
\[
= 0 \quad \text{for } u = 0.
\]  (2)

For example, if \( u = [2, 2i, 0, 0] \), then \( \|u\| = \sqrt{(2)^2 + (2i)^2 + (0)^2 + (0)^2} = 0 \) even though \( u \neq 0 \); and if \( u = [0, 2i, 0, 0] \), then \( \|u\| = \sqrt{0 - 4 + 0 + 0} = 2i \) is not even real so it cannot satisfy the condition \( \|u\| > 0 \).

To avoid this problem with the norm, we adopt the modified dot product

\[
u \cdot v \equiv u_1\overline{v}_1 + u_2\overline{v}_2 + \cdots + u_n\overline{v}_n = \sum_{j=1}^{n} u_j\overline{v}_j
\]  (3)

where the overhead bar denotes complex conjugate for then

\[
\|u\| \equiv \sqrt{u \cdot \overline{u}} = \sqrt{\sum_{j=1}^{n} u_j\overline{u}_j} = \sqrt{\sum_{j=1}^{n} |u_j|^2}
\]  (4)

does satisfy the nonnegativeness condition (2). [Recall that if \( z = a + ib \), then \( z\overline{z} = (a + ib)(a - ib) = a^2 + b^2 = |z|^2 \).]

**EXAMPLE 1.** If \( u = [2, 3 - 5i, 0, 4i] \), then

\[
\|u\| = \sqrt{(2)(2) + (3 - 5i)(3 + 5i) + (0)(0) + (4i)(-4i)}
\]
\[
= \sqrt{4 + 34 + 16} = \sqrt{54},
\]

*Recall from Section 22.2 (which is the prerequisite for this chapter) that inequalities such as \( z > 0 \) and \( z < 0 \) are not meaningful if \( z \) is complex; see the paragraph below equation (12) in that section.*
which is real and positive. □

**Properties of the dot product.** Observe that

\[
\mathbf{v} \cdot \mathbf{u} = v_1u_1 + \cdots + v_nu_n = v_1\overline{u}_1 + \cdots + v_n\overline{u}_n = \mathbf{v}^* \mathbf{u}
\]

Thus, in place of the commutativity condition \( \mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u} \), satisfied in the real case, we have the so-called conjugate commutativity \( \mathbf{u} \cdot \mathbf{v} = \mathbf{v}^* \mathbf{u} \) in the complex case. In fact, the properties

- **Conjugate Commutative:** \( \mathbf{u} \cdot \mathbf{v} = \mathbf{v}^* \mathbf{u} \), \( (5a) \)
- **Nonnegative:** \( \mathbf{u} \cdot \mathbf{u} > 0 \) for all \( \mathbf{u} \neq \mathbf{0} \),
  \( \mathbf{u} \cdot \mathbf{u} = 0 \) for \( \mathbf{u} = \mathbf{0} \), \( (5b) \)
- **Linear:** \( (\alpha \mathbf{u} + \beta \mathbf{v}) \cdot \mathbf{w} = \alpha (\mathbf{u} \cdot \mathbf{w}) + \beta (\mathbf{v} \cdot \mathbf{w}) \), \( (5c) \)

of the complex dot product are the same as in the real case [(12) in Section 9.5.2] except for the complex conjugate bar in \( (5a) \).

**EXAMPLE 2.** If \( \mathbf{u} = [1 + 2i, -4] \) and \( \mathbf{v} = [i, 3 - i] \), then \( \mathbf{u} \cdot \mathbf{v} = (1 + 2i)(-i) + (-4)(3 + i) = -10 + 5i \), which does equal \( \mathbf{u}^* \mathbf{v} \), in accordance with \( (5a) \).

The Schwarz inequality,

\[
|\mathbf{u} \cdot \mathbf{v}| \leq ||\mathbf{u}|| \cdot ||\mathbf{v}||
\]

is found to hold (Exercise 7), as in the real case, except that here \( |\mathbf{u} \cdot \mathbf{v}| \) is the modulus of a complex number rather than the absolute value of a real number.

**Properties of the norm.** The norm \( (4) \) admits the properties

- **Scaling:** \( ||\alpha \mathbf{u}|| = |\alpha| \cdot ||\mathbf{u}|| \), \( (7a) \)
- **Nonnegative:** \( ||\mathbf{u}|| > 0 \) for all \( \mathbf{u} \neq \mathbf{0} \),
  \( ||\mathbf{u}|| = 0 \) for \( \mathbf{u} = \mathbf{0} \), \( (7b) \)
- **Triangle Inequality:** \( ||\mathbf{u} + \mathbf{v}|| \leq ||\mathbf{u}|| + ||\mathbf{v}|| \). \( (7c) \)
These properties are identical to those for the real case [(17) in Section 9.5.3] but here $|\alpha|$ is the modulus of a complex number rather than the absolute value of a real number. Also, the proofs are slightly different due to the presence of complex numbers. To illustrate, consider (7c):

$$
||u + v||^2 = (u + v) \cdot (u + v)
= u \cdot u + u \cdot v + v \cdot u + v \cdot v
= ||u||^2 + 2u \cdot v + ||v||^2
= ||u||^2 + 2 \Re(u \cdot v) + ||v||^2
$$

so that

$$
||u + v|| \leq ||u|| + ||v||.
$$

Proofs of (7a) and (7b) are left for the exercises.

Thus far, then, the only substantial change (beyond the fact that the scalars are now complex) is the change in the dot product from $u \cdot v \equiv u_1v_1 + \cdots + u_nv_n$ to $u \cdot v \equiv u_1\overline{v}_1 + \cdots + u_n\overline{v}_n$. Furthermore, whereas we defined the angle $\theta$ between $u$ and $v$ as $\theta = \cos^{-1} \left( \frac{u \cdot v}{||u|| ||v||} \right)$ in the real case, this definition is awkward in the complex case since $u \cdot v$ (and hence the argument of the arccosine) is complex. Instead of trying to patch things up, we simply choose not to define $\theta$ for the complex case, although we do retain a notion of perpendicularity; that is, we still say that $u$ and $v$ are orthogonal if $u \cdot v = 0$.

Finally,

**THEOREM 12.2.1** Dimension of $\mathbb{C}^n$

The dimension of $\mathbb{C}^n$ is $n$.

You may have expected the dimension to be $2n$, on the grounds that each of the $n$ vector components has both real and imaginary parts. However, observe that the ON set

$$
\hat{e}_1 = [1, 0, \ldots, 0],
\hat{e}_2 = [0, 1, 0, \ldots, 0],
\vdots
\hat{e}_n = [0, \ldots, 0, 1]
$$

is a basis for $\mathbb{C}^n$, just as it is for $\mathbb{R}^n$ since the set is LI and spans the space. That it spans $\mathbb{C}^n$ follows from the fact that every vector $u = (u_1, \ldots, u_n)$ in $\mathbb{C}^n$ can be
expanded as \( u = u_1 \hat{e}_1 + \cdots + u_n \hat{e}_n \). And since the basis (8) contains \( n \) vectors, it follows from the definition of dimension in Section 9.9.2 that \( \mathbb{C}^n \) is \( n \)-dimensional.

### EXERCISES 12.2

1. Normalize each of the following vectors

   \( \begin{align*}
   \text{(a)} & \quad [1, i] \\
   \text{(b)} & \quad [1 + i, 1 - i] \\
   \text{(c)} & \quad [1, 3, -2, 0] \\
   \text{(d)} & \quad [2, 1 - 3i, 0, 5] \\
   \text{(e)} & \quad [2 + 3i, 1 - i, 4i] \\
   \text{(f)} & \quad [i, 0, 0, -i] \\
   \text{(g)} & \quad [x, y, z, ic] \\
   \end{align*} \)

   \[ \text{and use it to expand each vector in terms of the bases given in Exercise 3.} \]

2. Show whether the following vector sets are bases for \( \mathbb{C}^3 \);

   \( \begin{align*}
   \text{(a)} & \quad [i, 2, 1 + i], [0, 1, 2 + i], [4, 1, -i] \\
   \text{(b)} & \quad [1, 0, 2], [3, 2, -2], [0, 0, 4] \\
   \text{(c)} & \quad [4, 1 - 2i, 0], [3 + i, i, -2i], [-2 - 2i, 1 - 4i, 4i] \\
   \text{(d)} & \quad [1, 0, 0], [0, 1, 0], [0, 0, i] \\
   \text{(e)} & \quad [i, 1, 2], [3, 1 - i, -i], [3 + 2i, 3 - i, 4 - i], [3 - i, -i, -2 - i] \\
   \end{align*} \)

3. Show that the set \( e_1 = [i, 1, 0], e_2 = [2, 2i, 1], e_3 = [1, i, -i] \) is an orthogonal basis for \( \mathbb{C}^3 \).

4. Show that (23) in Section 9.9 (with \( k = n \)) is valid for \( \mathbb{C}^n \).

5. Show that \( (\alpha x) \cdot y = \alpha (x \cdot y) \), \( \text{ whereas } x \cdot (\alpha y) = \overline{\alpha} (x \cdot y) \). \( \text{ (5.1) } \)

6. The property (5a) was proved in the text. Prove (5b) and (5c).

7. Prove the Schwarz inequality (6).

8. Prove the properties (7a) and (7b).

9. Vectors in \( \mathbb{R}^1 \), \( \mathbb{R}^2 \), and \( \mathbb{R}^3 \) can be displayed, graphically, as arrow vectors. Is the same true for \( \mathbb{C}^1 \), \( \mathbb{C}^2 \), \( \mathbb{C}^3 \)? Explain.

### 12.3 Complex Matrices

All of Chapter 10 (on matrices, determinants, and linear equations) holds even if the scalars are allowed to be complex. To illustrate, consider a representative example.

#### EXAMPLE 1.

Given

\[
A = \begin{bmatrix}
2 & 1 + i \\
0 & i
\end{bmatrix},
\]

compute \( A^{-1} \). We proceed as usual, although the numbers are now complex:

\[
\det A = (2)(i) - (0)(1 + i) = 2i \quad (\neq 0, \text{ so } A^{-1} \text{ exists})
\]

<table>
<thead>
<tr>
<th>Minors: ( M_{11} = i )</th>
<th>Cofactors: ( A_{11} = i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M_{12} = 0 )</td>
<td>( A_{12} = 0 )</td>
</tr>
<tr>
<td>( M_{21} = 1 + i )</td>
<td>( A_{21} = -1 - i )</td>
</tr>
<tr>
<td>( M_{22} = 2 )</td>
<td>( A_{22} = 2 )</td>
</tr>
</tbody>
</table>
Chapter 12. Extension to Complex Case

so

$$A^{-1} = \frac{1}{2i} \begin{bmatrix} i & -1 - i \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} \frac{-1+i}{2} \\ 0 & -i \end{bmatrix},$$

as is verified by showing that $A^{-1}A = I$.  

However, in Chapter 11, on the eigenvalue problem, dot products begin to appear so the impact of the change in the dot product (from $x \cdot y = x_1y_1 + \cdots + x_ny_n$ to $x \cdot y = \Re(x_1y_1 + \cdots + x_ny_n)$) begins to be felt. Expressed in matrix form, the dot product is now

$$x \cdot y = x^T y$$

We begin with two definitions. Given an $m \times n$ matrix $A = \{a_{ij}\}$, we define the complex conjugate of $A$ as

$$\bar{A} = \{\bar{a}_{ij}\}$$

and the Hermitian conjugate of $A$ as

$$A^* = \{\bar{a}_{ji}\}, \quad \text{i.e.,} \quad A^* = \bar{A}^T.$$  

If $\bar{A} = A$ then $A$ is real, and if $A^* = A$ then $A$ is Hermitian. If $A$ is not square then it cannot be Hermitian.

**EXAMPLE 2.** If

$$A = \begin{bmatrix} 2 + i & 0 & 3 - 5i \\ 7 & 1 & 4i \\ 2 & i & 3 \end{bmatrix},$$

then

$$\bar{A} = \begin{bmatrix} 2 - i & 0 & 3 + 5i \\ 7 & 1 & -4i \\ 2 & -i & 3 \end{bmatrix} \quad \text{and} \quad A^* = \begin{bmatrix} 2 - i & 7 & 2 \\ 0 & 1 & -i \\ 3 + 5i & -4i & 3 \end{bmatrix}. $$

Since $A^* \neq A$, $A$ is not Hermitian. 

**EXAMPLE 3.** If

$$A = \begin{bmatrix} 3 & 1 + 4i \\ 1 - 4i & 0 \end{bmatrix},$$

†Charles Hermite (1822–1901), a professor at the Sorbonne and at the Ecole Polytechnique, contributed to the theory of elliptic functions and is also well known for his introduction of the Hermite polynomials.
then
\[ \text{then} \]
\[ \overline{A} = \begin{bmatrix} 3 & 1 - 4i \\ 1 + 4i & 0 \end{bmatrix} \quad \text{and} \quad A^* = \begin{bmatrix} 3 & 1 + 4i \\ 1 - 4i & 0 \end{bmatrix}. \]
Since \( A^* = A \), \( A \) is Hermitian. \( \blacklozenge \)

Some properties of the complex conjugate and Hermitian conjugate matrices are as follows:
\[ \overline{A} = A, \quad (\overline{A + B}) = \overline{A} + \overline{B}, \quad \overline{AB} = \overline{A} \overline{B}, \quad (A^*)^* = A, \quad (A + B)^* = A^* + B^*, \quad (AB)^* = B^* A^*. \]

These properties are readily verified. In addition, the key property of the Hermitian conjugate matrix \( A^* \) is that
\[ (Ax) \cdot y = x \cdot (A^* y) \]
holds for all vectors \( x \) and \( y \); specifically, if \( A \) is \( m \times n \), \( x \) is any \( n \times 1 \) vector and \( y \) is any \( m \times 1 \) vector. To prove (7), observe that
\[ (Ax) \cdot y = (Ax)^T y = x^T A^T y = x^T \overline{A} y = x^T A^* y = x \cdot (A^* y). \]

**Hermitian matrices.** Recall from Chapters 10 and 11 that matrices arising in applications are often symmetric (\( A^T = A \)) and that such matrices exhibit several useful properties concerning the eigenvalue problem (Theorems 11.3.1–11.3.4). Likewise, when complex matrices arise in applications they are often Hermitian (\( A^* = \overline{A}^T = A \)), and such matrices exhibit analogous useful properties, given by Theorems 12.3.1–12.3.4 below.

**EXAMPLE 4.** *Lorentz Transformation.* In the special theory of relativity one considers the vector \( [x, y, z, ict] \), where \( x, y, z \) are Cartesian coordinates, \( c \) is the speed of light, and \( t \) is the time. If the corresponding vector, referred to an \( x', y', z' \) system which is translating in the \( z \) direction with constant speed \( v \), is denoted as \( [x', y', z', ict'] \), it turns out that these vectors are related according to
\[ \begin{bmatrix} x' \\ y' \\ z' \\ ict' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & i\beta \\ 0 & 0 & -i\beta & \sqrt{1 - \beta^2} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ ict \end{bmatrix}, \]

where \( \beta = v/c \).
where $\beta = v/c (<1)$. This is the important Lorentz transformation, and it is seen that the
transformation matrix is Hermitian. \[ \]

**THEOREM 12.3.1  Real Eigenvalues**

If $A$ is Hermitian ($A^T = A$), then all of its eigenvalues are real.

**Proof:** Let us both pre-dot and post-dot both sides of $Ax = \lambda x$ with $x$:

- Pre-dot: $x \cdot (Ax) = x \cdot (\lambda x)$
  
- Post-dot: $(Ax) \cdot x = (\lambda x) \cdot x$

But the left-hand sides are equal, by virtue of property (7) together with the assumption that $A$ is Hermitian. Thus, subtracting (10b) from (10a) gives

$$ (\bar{\lambda} - \lambda)(x \cdot x) = 0. \tag{11} $$

Now, $x \cdot x = ||x||^2 \neq 0$ since $x$ is an eigenvector, so it follows from (11) that $\bar{\lambda} - \lambda = 0$, or $\bar{\lambda} = \lambda$. Thus, $\lambda$ is real. \[ \]

**THEOREM 12.3.2  Dimension of Eigenspace**

If an eigenvalue $\lambda$ of an Hermitian matrix $A$ is of multiplicity $k$, then the eigenspace corresponding to $\lambda$ is of dimension $k$.

**THEOREM 12.3.3  Orthogonality of Eigenvectors**

If $A$ is Hermitian, then eigenvectors corresponding to distinct eigenvalues are orthogonal.

**Proof:** Let $e_j$ and $e_k$ be eigenvectors corresponding to distinct eigenvalues $\lambda_j$ and $\lambda_k$, respectively. That is,

$$ A e_j = \lambda_j e_j \quad \text{and} \quad A e_k = \lambda_k e_k. \tag{12a,b} $$

Pre-dotting both sides of (12a) with $e_k$, and post-dotting both sides of (12b) with $e_j$, and using the fact that $\bar{\lambda} = \lambda$, we have

$$ e_k \cdot (A e_j) = e_k \cdot (\lambda_j e_j) = \bar{\lambda}_j (e_k \cdot e_j) = \lambda_j (e_k \cdot e_j) \tag{13} $$

$$ (A e_k) \cdot e_j = (\lambda_k e_k) \cdot e_j = \lambda_k (e_k \cdot e_j). $$
But (7) tells us that \((Ae_k) \cdot e_j = e_k \cdot (A^*e_j)\) and, since \(A^* = A\) by assumption, the equation on the right side of (13) becomes \(e_k \cdot (Ae_j) = \lambda_k (e_k \cdot e_j)\). Subtracting that equation from the one on the left side of (13) gives

\[
0 = (\lambda_j - \lambda_k)(e_k \cdot e_j).
\]  

(14)

Finally, \(\lambda_j - \lambda_k \neq 0\) since \(\lambda_j\) and \(\lambda_k\) are distinct, by assumption, so it follows from (14) that \(e_k \cdot e_j = 0\), as was to be shown. ■

**THEOREM 12.3.4 Orthogonal Basis**

If an \(n \times n\) matrix \(A\) is Hermitian, then its eigenvectors provide an orthogonal basis for \(n\)-space.

Proof of Theorem 12.3.4 follows the same lines as the proof of Theorem 11.3.4.

**EXAMPLE 5.** Find the eigenvalues and eigenvectors of

\[
A = \begin{bmatrix}
3 & 2i \\
-2i & 0
\end{bmatrix}.
\]

Setting \(\det(A - \lambda I) = 0\) gives the characteristic equation

\[(3 - \lambda)(-\lambda) - (2i)(-2i) = \lambda^2 - 3\lambda - 4 = (\lambda + 1)(\lambda - 4) = 0.
\]

Hence,

\[
\lambda = -1\quad \text{and} \quad \lambda = 4.
\]

Observe that \(A\) is Hermitian and, sure enough, in accordance with Theorem 12.3.1, the \(\lambda\)'s are real.

To find the eigenvectors, solve \((A - \lambda I)x = 0\).

\[
\lambda = \lambda_1 = -1: \begin{bmatrix}
4 & 2i \\
-2i & 1
\end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{or} \quad 4x_1 + 2ix_2 = 0 \quad \text{and} \quad -2ix_1 + x_2 = 0.
\]

Solving, \(x_2 = 2ix_1\) so the eigenspace corresponding to \(\lambda_1\) is

\[
e_1 = \alpha \begin{bmatrix} 1 \\ 2i \end{bmatrix}.
\]

Similarly, we find that the eigenspace corresponding to \(\lambda_2\) is

\[
e_1 = \beta \begin{bmatrix} 2i \\ 1 \end{bmatrix}.
\]

Since \(A\) is Hermitian and \(\lambda_1\) and \(\lambda_2\) are distinct, the eigenvectors should be orthogonal (Theorem 12.3.3). Let us see: with \(\alpha = \beta = 1\), say, \(e_1 \cdot e_2 = (1)(2i) + (2i)(1) = 2i + 2i = 0\) so they are, indeed, orthogonal. ■
Hermitian forms, diagonalization, and unitary matrices. The analog of the quadratic form \( f(x_1, \ldots, x_n) = x^T A x \), where \( A \) is symmetric and \( x = (x_1, \ldots, x_n)^T \), is the Hermitian form

\[
f = \bar{x}^T A x,
\]

where \( A \) is Hermitian. For example, if \( n = 2 \), then (15) becomes

\[
f = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{12}^* & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = a_{11}x_1 \bar{x}_1 + a_{12}x_2 \bar{x}_2 + a_{12}^*x_1 \bar{x}_2 + a_{22}x_2 \bar{x}_2.
\]

(16)

The right-hand side of (16) is seen to be real since \( a_{11} \) and \( a_{22} \) are diagonal elements of an Hermitian matrix and hence are real, \( x_1 \bar{x}_1 = |x_1|^2 \) and \( x_2 \bar{x}_2 = |x_2|^2 \) are real, and the fourth term is the conjugate of the third term. In fact, \( f \) is real for all values \( n \geq 1 \):

\[
\bar{f} = x^T A x = (Ax)^T x = x^T A^T x = x^T A x = f.
\]

(17)

(See Exercise 9.)

To reduce \( f \) to canonical form, set \( x = U \bar{x} \) in (15). Then

\[
f = (U \bar{x})^T A U \bar{x} = \bar{x}^T (U^* A U) \bar{x}
\]

(18)

so \( U \) is to be chosen so that

\[
U^* A U = D
\]

(19)

is diagonal. Recalling our discussion of the real case (Section 11.4), it is not surprising that a suitable \( U \) matrix is a normalized modal matrix of \( A \) (Exercise 10). With that choice, the diagonal elements of \( D \) are the eigenvalues of \( A \), and \( f \) reduces to the canonical form

\[
f = \lambda_1 \bar{x}_1^2 + \lambda_2 \bar{x}_2^2 + \cdots + \lambda_n \bar{x}_n^2
\]

(20)

In the real case the normalized modal matrix was denoted as \( Q \), and it turned out that \( Q^T = Q^{-1} \). Analogously, we find (Exercise 11), in the present case, that

\[
U^T = U^{-1} \quad \text{or} \quad U^* = U^{-1},
\]

(21)

and such a matrix is said to be unitary.

\[\text{THEOREM 12.3.5} \quad \text{Eigenvalues of Unitary Matrix}\]

If \( \lambda \) is an eigenvalue of a unitary matrix, then \( |\lambda| = 1 \).
Proof: Let \( U \) be unitary, with an eigenvalue \( \lambda \) and corresponding eigenvector \( e \):
\[
Ue = \lambda e \quad (e \neq 0).
\]  
(22)

Seeking to employ (21), take the conjugate transpose of both sides:
\[
(Ue)^* = (\lambda e)^* \quad \text{or} \quad \overline{e}^T U^{-1} \overline{e} = \overline{\lambda} \overline{e}^T.
\]  
(23)

And post-multiplying the left with \( Ue \) and the right with \( \lambda e \) [which are equal by (22) and nonzero], (23) gives
\[
\overline{e}^T U^{-1} Ue = \lambda \overline{e}^T \lambda e \quad \text{or} \quad \overline{e}^T e = |\lambda|^2 \overline{e}^T e.
\]  
(24)

But \( \overline{e}^T e = ||e||^2 \neq 0 \) so (24) implies \( |\lambda|^2 = 1 \), or \( |\lambda| = 1 \), as claimed. \( \blacksquare \)

**EXAMPLE 6.** Reduce
\[
f = 3x_1 \overline{x}_1 + 2i(x_1 x_2 - 2ix_1 \overline{x}_2)
\]  
(25)
to a canonical form. Comparing (25) with (16), we see that \( a_{11} = 3, a_{22} = 0, \) and \( a_{12} = 2i, \) so that
\[
A = \begin{bmatrix}
3 & 2i \\
-2i & 0
\end{bmatrix}.
\]

This is the same \( A \) as in Example 5 so
\[
\lambda_1 = -1, \quad \overline{e}_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2i \end{bmatrix}; \quad \lambda_2 = 4, \quad \overline{e}_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2i \\ 1 \end{bmatrix}.
\]

Thus, the desired canonical form is
\[
f = \lambda_1 |x_1|^2 + \lambda_2 |\overline{x}_2|^2 = -|x_1|^2 + 4|\overline{x}_2|^2,
\]
where \( x_1, x_2 \) and \( \overline{x}_1, \overline{x}_2 \) are related according to
\[
x = UX = [\overline{e}_1, \overline{e}_2] \overline{x}
\]
or
\[
\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 2i \\ 2i & 1 \end{bmatrix} \begin{bmatrix} \overline{x}_1 \\ \overline{x}_2 \end{bmatrix}.
\]

Or, if we prefer it the other way around,
\[
\overline{x} = U^{-1} \overline{x} = U^* \overline{x} \quad \text{or} \quad \begin{bmatrix} \overline{x}_1 \\ \overline{x}_2 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -2i \\ -2i & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.
\]  

**Computer software.** Using Maple, for instance, no new commands are needed; just type \( I \)'s for \( i \)'s. For instance, to find the eigenvalues and eigenvectors of
\[
A = \begin{bmatrix}
3 & 2i \\
-2i & 0
\end{bmatrix},
\]
type
and return. Then enter
\[
A := \text{matrix}(2, 2, [3, 2*I, -2*I, 0])
\]
and the command
eigenvects(A);
The result,
\[
[4, 1, \{[2I, 1]\}], \{-1, 1, \{[-1/2I, 1]\}\}
\]
is the same as found, by hand, in Example 5. The 1's following the eigenvalues 4
and -1 indicate the multiplicity of those eigenvalues.

EXERCISES 12.3

1. Invert each of the following matrices. If the matrix is not
invertible (i.e., if it is singular), state that.
(a) \[
\begin{bmatrix}
2 & i \\
1 & 4
\end{bmatrix}
\]
(b) \[
\begin{bmatrix}
1 + i & 3i \\
1 & 2 - i
\end{bmatrix}
\]
(c) \[
\begin{bmatrix}
3 & 1 + i \\
1 - i & 2
\end{bmatrix}
\]
(d) \[
\begin{bmatrix}
1 & 3 \\
0 & 1 + i
\end{bmatrix}
\]
(e) \[
\begin{bmatrix}
0 & i \\
-i & 0
\end{bmatrix}
\]
(f) \[
\begin{bmatrix}
0 & 2 - 2i \\
0 & 2 - 2i
\end{bmatrix}
\]
2. (a)–(f) Same as Exercise 1, but using computer software.
3. (a)–(f) Determine the eigenvalues and eigenvectors of each
of the matrices in Exercise 1, and show that the results are in
accord with the relevant theorems in this section.
4. (a)–(f) Same as Exercise 3, but using computer software.
5. Determine the eigenvalues and eigenvectors in each case.
(a) \[
\begin{bmatrix}
1 & 2 \\
1 & 0
\end{bmatrix}
\]
(b) \[
\begin{bmatrix}
2 & 1 \\
1 & -1
\end{bmatrix}
\]
(c) \[
\begin{bmatrix}
1 & 2 \\
2 & 4
\end{bmatrix}
\]
(d) \[
\begin{bmatrix}
4 & -5 \\
1 & 2
\end{bmatrix}
\]
6. Give necessary and sufficient conditions on a, b, c, d such
that \(A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\) will have complex eigenvalues, i.e., such
that \(\text{Im} \lambda \neq 0\).
7. Reduce the following to the canonical form (20), and give
the matrix transformations from \(x\) to \(\bar{x}\), and from \(\bar{x}\) to \(x\).
8. Given each of the following \(A\) matrices, evaluate \(A^{1000}\).
(a) \[
\begin{bmatrix}
2 & 1 - i \\
1 + i & 3
\end{bmatrix}
\]
(b) \[
\begin{bmatrix}
2 & -2i \\
2i & 5
\end{bmatrix}
\]
(c) \[
\begin{bmatrix}
0 & -i \\
i & 2
\end{bmatrix}
\]
(d) \[
\begin{bmatrix}
3 & 2 - i \\
2 + i & -1
\end{bmatrix}
\]
9. Justify the second through fifth equalities in (17), citing
relevant equation numbers or theorems.
10. Verify that if \(U\) is a normalized modal matrix of \(A\), then
\(f\) does reduce to the canonical form (20), as claimed.
11. Verify that a normalized modal matrix \(U\) satisfies (21).
12. Is the Lorentz transformation matrix, in (9), unitary? Or-
thogonal?
13. If \(A = \cdot (A_2) = x \cdot (Bx)\) for all \(x\), where \(A\) and \(B\) are Her-
mitian, does it follow that \(A = B\)? Explain.
14. (Necessary condition for existence of solutions of \(A_2 = x\))
Prove that a necessary condition for the existence of solution(s)
to \(A_2 = c\) (where \(A\) is \(m \times n\), \(x\) is \(n \times 1\), and \(c\) is \(m \times 1\))
is that \(c\) be orthogonal to every solution \(z\) of the associated
homogeneous equation \(A_2z = 0\). HINT: Dot both sides of
\(A_2x = c\) with \(z\), and then use (7).
15. Use the result stated in Exercise 14 to determine necessary conditions, if any, on the components of \( c \), for the given system to be consistent. Compare your result with those obtained by direct application of Gauss elimination to the given system.

(a) \[
\begin{bmatrix}
2 & 1 \\
2 & 1
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2
\end{bmatrix} = \begin{bmatrix} c_1 \\
c_2
\end{bmatrix}
\]

(b) \[
\begin{bmatrix}
2 & 1 \\
4 & 1
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2
\end{bmatrix} = \begin{bmatrix} c_1 \\
c_2
\end{bmatrix}
\]

(c) \[
\begin{bmatrix}
2 & 1 \\
i & 1
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2
\end{bmatrix} = \begin{bmatrix} c_1 \\
c_2
\end{bmatrix}
\]

(d) \[
\begin{bmatrix}
1 & 3 & 2 & 1 \\
2 & -1 & 1 & 0 \\
3 & 2 & 3 & 1
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4
\end{bmatrix} = \begin{bmatrix} c_1 \\
c_2 \\
c_3
\end{bmatrix}
\]

(e) \[
\begin{bmatrix}
1 & 3 & 2 & i \\
2 & -1 & 1 & 0 \\
3 & 2 & 3 & i
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4
\end{bmatrix} = \begin{bmatrix} c_1 \\
c_2 \\
c_3
\end{bmatrix}
\]

(f) \[
\begin{bmatrix}
2 & 1 & 1 \\
i & 1 & 2i - 3 \\
1 & 2 & -4 \\
-1 & 1 & -5
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4
\end{bmatrix} = \begin{bmatrix} c_1 \\
c_2 \\
c_3 \\
c_4
\end{bmatrix}
\]

16. (Decomposition) Given any \( m \times n \) matrix \( A \), Hermitian or not, show that \( A \) can be split as \( A = B + iC \), where \( B \) and \( C \) are each Hermitian. Show that \( B = (A^* + A)/2 \) and \( C = i(A^* - A)/2 \).

17. (a) If \( A \) is Hermitian, is \( iA \) Hermitian? Explain.
(b) If \( A \) is Hermitian, is \( A^2 \) Hermitian? Explain.

---

**Chapter 12 Review**

This chapter is so compact that it hardly warrants review. But, let us stress three points.

First, the key difference between \( \mathbb{R}^n \) and \( \mathbb{C}^n \) is in the *dot product*, which is \( \mathbf{x} \cdot \mathbf{y} = \mathbf{x}^\top \mathbf{y} \) for \( \mathbb{R}^n \) and

\[
\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^\top \mathbf{y}
\]

for \( \mathbb{C}^n \), the complex conjugate being introduced so that the dot product \( \mathbf{x} \cdot \mathbf{y} \) and the norm \( \|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}} \) satisfy the properties listed in equations (12) and (17), respectively, in Section 9.5. (If you studied the optional Section 9.6 you will recall that these properties were elevated to requirements, or axioms, for any normed inner product space.)

Second, just as real matrices are found, in Section 11.3, to admit special useful properties (e.g., the eigenvalues are real, eigenvectors corresponding to distinct eigenvalues are orthogonal, and the eigenvectors provide an orthogonal basis for the \( n \)-space) if they are symmetric \( (A^T = A) \), we find in this chapter that complex matrices admit the same properties if they are *Hermitian* \( (A^* = A, \text{where } A^* \text{ is } A^T) \).

Finally, we call your attention to the result

\[
(A\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (A^* \mathbf{y}),
\]

given as equation (7) in Section 12.3. For real matrices the latter becomes

\[
(A\mathbf{x}) \cdot \mathbf{y} = \mathbf{x} \cdot (A\mathbf{y})
\]
if $A$ is symmetric, and for complex matrices (3) holds if $A$ is Hermitian; (3) is the key relation used in proving that eigenvectors corresponding to distinct eigenvalues are orthogonal, which result, in turn, is needed in proving that the eigenvectors provide an orthogonal basis for the $n$-space. We will meet a function-space version of (3) known as Lagrange's identity, in Chapter 17, when we study the Sturm–Liouville theory. Expansions in terms of bases consisting of orthogonal eigenvectors (or “eigenfunctions” in the function space case) will be of great importance to us, and therefore the underlying relation (3) is of great importance as well.