CHAPTER 1

Vectors in R^n and C^n, Spatial Vectors

1.1 Introduction

There are two ways to motivate the notion of a vector: one is by means of lists of numbers and subscripts, and the other is by means of certain objects in physics. We discuss these two ways below.

Here we assume the reader is familiar with the elementary properties of the field of real numbers, denoted by **R**. On the other hand, we will review properties of the field of complex numbers, denoted by **C**. In the context of vectors, the elements of our number fields are called *scalars*.

Although we will restrict ourselves in this chapter to vectors whose elements come from **R** and then from **C**, many of our operations also apply to vectors whose entries come from some arbitrary field K.

Lists of Numbers

Suppose the weights (in pounds) of eight students are listed as follows:

$$156, \quad 125, \quad 145, \quad 134, \quad 178, \quad 145, \quad 162, \quad 193$$

One can denote all the values in the list using only one symbol, say w, but with different subscripts; that is,

$$w_1, \quad w_2, \quad w_3, \quad w_4, \quad w_5, \quad w_6, \quad w_7, \quad w_8$$

Observe that each subscript denotes the position of the value in the list. For example,

$$w_1 = 156, \text{ the first number, } w_2 = 125, \text{ the second number, } \ldots$$

Such a list of values,

$$w = (w_1, w_2, w_3, \ldots, w_8)$$

is called a *linear array* or *vector*.

Vectors in Physics

Many physical quantities, such as temperature and speed, possess only "magnitude." These quantities can be represented by real numbers and are called *scalars*. On the other hand, there are also quantities, such as force and velocity, that possess both "magnitude" and "direction." These quantities, which can be represented by arrows having appropriate lengths and directions and emanating from some given reference point O, are called *vectors*.

Now we assume the reader is familiar with the space $\mathbf{R}^3$ where all the points in space are represented by ordered triples of real numbers. Suppose the origin of the axes in $\mathbf{R}^3$ is chosen as the reference point O for the vectors discussed above. Then every vector is uniquely determined by the coordinates of its endpoint, and vice versa.

There are two important operations, vector addition and scalar multiplication, associated with vectors in physics. The definition of these operations and the relationship between these operations and the endpoints of the vectors are as follows.

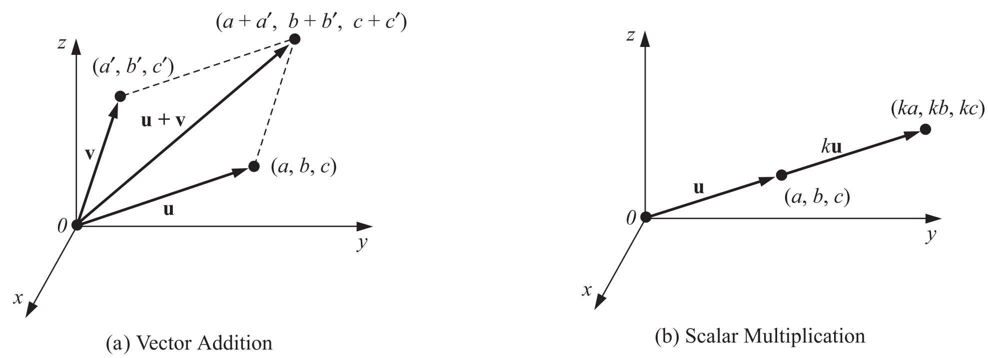


Figure 1-1

(i) ***Vector Addition:*** The resultant $\mathbf{u} + \mathbf{v}$ of two vectors $\mathbf{u}$ and $\mathbf{v}$ is obtained by the *parallelogram law*; that is, $\mathbf{u} + \mathbf{v}$ is the diagonal of the parallelogram formed by $\mathbf{u}$ and $\mathbf{v}$. Furthermore, if (a, b, c) and (a', b', c') are the endpoints of the vectors $\mathbf{u}$ and $\mathbf{v}$, then $(a + a',\ b + b',\ c + c')$ is the endpoint of the vector $\mathbf{u} + \mathbf{v}$. These properties are pictured in Fig. 1-1(a).

(ii) ***Scalar Multiplication:*** The product $k\mathbf{u}$ of a vector $\mathbf{u}$ by a real number k is obtained by multiplying the magnitude of $\mathbf{u}$ by k and retaining the same direction if $k > 0$ or the opposite direction if $k < 0$. Also, if (a, b, c) is the endpoint of the vector $\mathbf{u}$, then (ka, kb, kc) is the endpoint of the vector $k\mathbf{u}$. These properties are pictured in Fig. 1-1(b).

Mathematically, we identify the vector $\mathbf{u}$ with its (a, b, c) and write $\mathbf{u} = (a, b, c)$. Moreover, we call the ordered triple (a, b, c) of real numbers a point or vector depending upon its interpretation. We generalize this notion and call an n-tuple $(a_1, a_2, \ldots, a_n)$ of real numbers a vector. However, special notation may be used for the vectors in $\mathbf{R}^3$ called *spatial* vectors (Section 1.6).

1.2 Vectors in $\mathbf{R}^n$

The set of all n-tuples of real numbers, denoted by $\mathbf{R}^n$, is called *n-space*. A particular n-tuple in $\mathbf{R}^n$, say

$$u = (a_1, a_2, \ldots, a_n)$$

is called a *point* or *vector*. The numbers a_i are called the *coordinates*, *components*, *entries*, or *elements* of u. Moreover, when discussing the space $\mathbf{R}^n$, we use the term *scalar* for the elements of $\mathbf{R}$.

Two vectors, u and v, are *equal*, written $u = v$, if they have the same number of components and if the corresponding components are equal. Although the vectors $(1, 2, 3)$ and $(2, 3, 1)$ contain the same three numbers, these vectors are not equal because corresponding entries are not equal.

The vector $(0, 0, \ldots, 0)$ whose entries are all 0 is called the *zero vector* and is usually denoted by 0.

EXAMPLE 1.1

(a) The following are vectors:

$$(2, -5), \quad (7, 9), \quad (0, 0, 0), \quad (3, 4, 5)$$

The first two vectors belong to $\mathbf{R}^2$, whereas the last two belong to $\mathbf{R}^3$. The third is the zero vector in $\mathbf{R}^3$.

(b) Find x, y, z such that $(x - y,\ x + y,\ z - 1) = (4, 2, 3)$.

By definition of equality of vectors, corresponding entries must be equal. Thus,

$$x - y = 4, \qquad x + y = 2, \qquad z - 1 = 3$$

Solving the above system of equations yields $x = 3$, $y = -1$, $z = 4$.

Column Vectors

Sometimes a vector in n-space $\mathbf{R}^n$ is written vertically rather than horizontally. Such a vector is called a *column vector*, and, in this context, the horizontally written vectors in Example 1.1 are called *row vectors*. For example, the following are column vectors with $2, 2, 3$, and 3 components, respectively:

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \begin{bmatrix} 3 \\ -4 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 5 \\ -6 \end{bmatrix}, \quad \begin{bmatrix} 1.5 \\ \frac{2}{3} \\ -15 \end{bmatrix}$$

We also note that any operation defined for row vectors is defined analogously for column vectors.

1.3 Vector Addition and Scalar Multiplication

Consider two vectors u and v in $\mathbf{R}^n$, say

$$u = (a_1, a_2, \ldots, a_n) \qquad \text{and} \qquad v = (b_1, b_2, \ldots, b_n)$$

Their *sum*, written $u + v$, is the vector obtained by adding corresponding components from u and v. That is,

$$u + v = (a_1 + b_1,\ a_2 + b_2,\ \ldots,\ a_n + b_n)$$

The *scalar product* or, simply, *product*, of the vector u by a real number k, written ku, is the vector obtained by multiplying each component of u by k. That is,

$$ku = k(a_1, a_2, \ldots, a_n) = (ka_1, ka_2, \ldots, ka_n)$$

Observe that $u + v$ and ku are also vectors in $\mathbf{R}^n$. The sum of vectors with different numbers of components is not defined.

Negatives and subtraction are defined in $\mathbf{R}^n$ as follows:

$$-u = (-1)u \qquad \text{and} \qquad u - v = u + (-v)$$

The vector $-u$ is called the *negative* of u, and $u - v$ is called the *difference* of u and v.

Now suppose we are given vectors $u_1, u_2, \ldots, u_m$ in $\mathbf{R}^n$ and scalars $k_1, k_2, \ldots, k_m$ in $\mathbf{R}$. We can multiply the vectors by the corresponding scalars and then add the resultant scalar products to form the vector

$$v = k_1u_1 + k_2u_2 + k_3u_3 + \cdots + k_mu_m$$

Such a vector v is called a *linear combination* of the vectors $u_1, u_2, \ldots, u_m$.

EXAMPLE 1.2

(a) Let $u = (2, 4, -5)$ and $v = (1, -6, 9)$. Then

$$\begin{aligned} u + v &= (2 + 1,\ 4 + (-5),\ -5 + 9) = (3, -1, 4) \\ 7u &= (7(2), 7(4), 7(-5)) = (14, 28, -35) \\ -v &= (-1)(1, -6, 9) = (-1, 6, -9) \\ 3u - 5v &= (6, 12, -15) + (-5, 30, -45) = (1, 42, -60) \end{aligned}$$

(b) The zero vector $0 = (0, 0, \ldots, 0)$ in $\mathbf{R}^n$ is similar to the scalar 0 in that, for any vector $u = (a_1, a_2, \ldots, a_n)$.

$$u + 0 = (a_1 + 0,\ a_2 + 0,\ \ldots,\ a_n + 0) = (a_1, a_2, \ldots, a_n) = u$$

(c) Let $u = \begin{bmatrix} 2 \\ 3 \\ -4 \end{bmatrix}$ and $v = \begin{bmatrix} 3 \\ -1 \\ -2 \end{bmatrix}$. Then $2u - 3v = \begin{bmatrix} 4 \\ 6 \\ -8 \end{bmatrix} + \begin{bmatrix} -9 \\ 3 \\ 6 \end{bmatrix} = \begin{bmatrix} -5 \\ 9 \\ -2 \end{bmatrix}$.

Basic properties of vectors under the operations of vector addition and scalar multiplication are described in the following theorem.

THEOREM 1.1: For any vectors u, v, w in $\mathbf{R}^n$ and any scalars k, k' in $\mathbf{R}$,

(i)	$(u+v)+w = u+(v+w)$,	(v)	$k(u+v) = ku+kv$,
(ii)	$u+0 = u$,	(vi)	$(k+k')u = ku+k'u$,
(iii)	$u+(-u) = 0$,	(vii)	(kk')u=k(k'u),
(iv)	$u+v = v+u$,	(viii)	$1u = u$.

We postpone the proof of Theorem 1.1 until Chapter 2, where it appears in the context of matrices (Problem 2.3).

Suppose u and v are vectors in $\mathbf{R}^n$ for which $u = kv$ for some nonzero scalar k in $\mathbf{R}$. Then u is called a *multiple* of v. Also, u is said to be in the *same* or *opposite direction* as v according to whether $k > 0$ or $k < 0$.

1.4 Dot (Inner) Product

Consider arbitrary vectors u and v in $\mathbf{R}^n$; say,

$$u = (a_1, a_2, \ldots, a_n) \qquad \text{and} \qquad v = (b_1, b_2, \ldots, b_n)$$

The *dot product* or *inner product* or *scalar product* of u and v is denoted and defined by

$$u \cdot v = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$$

That is, $u \cdot v$ is obtained by multiplying corresponding components and adding the resulting products. The vectors u and v are said to be *orthogonal* (or *perpendicular*) if their dot product is zero—that is, if $u \cdot v = 0$.

EXAMPLE 1.3

(a) Let $u = (1, -2, 3)$, $v = (4, 5, -1)$, $w = (2, 7, 4)$. Then,

$$u \cdot v = 1(4) - 2(5) + 3(-1) = 4 - 10 - 3 = -9$$
$$u \cdot w = 2 - 14 + 12 = 0, \qquad v \cdot w = 8 + 35 - 4 = 39$$

Thus, u and w are orthogonal.

(b) Let $u = \begin{bmatrix} 2 \\ 3 \\ -4 \end{bmatrix}$ and $v = \begin{bmatrix} 3 \\ -1 \\ -2 \end{bmatrix}$. Then $u \cdot v = 6 - 3 + 8 = 11$.

(c) Suppose $u = (1, 2, 3, 4)$ and $v = (6, k, -8, 2)$. Find k so that u and v are orthogonal.

First obtain $u \cdot v = 6 + 2k - 24 + 8 = -10 + 2k$. Then set $u \cdot v = 0$ and solve for k:

$$-10 + 2k = 0 \qquad \text{or} \qquad 2k = 10 \qquad \text{or} \qquad k = 5$$

Basic properties of the dot product in $\mathbf{R}^n$ (proved in Problem 1.13) follow.

THEOREM 1.2: For any vectors u, v, w in $\mathbf{R}^n$ and any scalar k in $\mathbf{R}$:

(i)	$(u+v) \cdot w = u \cdot w + v \cdot w$,	(iii)	$u \cdot v = v \cdot u$,
(ii)	$(ku) \cdot v = k(u \cdot v)$,	(iv)	$u \cdot u \geq 0$, and $u \cdot u = 0$ iff $u = 0$.

Note that (ii) says that we can "take k out" from the first position in an inner product. By (iii) and (ii),

$$u \cdot (kv) = (kv) \cdot u = k(v \cdot u) = k(u \cdot v)$$

That is, we can also "take k out" from the second position in an inner product.

The space $\mathbf{R}^n$ with the above operations of vector addition, scalar multiplication, and dot product is usually called *Euclidean n-space*.

Norm (Length) of a Vector

The *norm* or *length* of a vector u in $\mathbf{R}^n$, denoted by $\|u\|$, is defined to be the nonnegative square root of $u \cdot u$. In particular, if $u = (a_1, a_2, \ldots, a_n)$, then

$$\|u\| = \sqrt{u \cdot u} = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}$$

That is, $\|u\|$ is the square root of the sum of the squares of the components of u. Thus, $\|u\| \geq 0$, and $\|u\| = 0$ if and only if $u = 0$.

A vector u is called a *unit* vector if $\|u\| = 1$ or, equivalently, if $u \cdot u = 1$. For any nonzero vector v in $\mathbf{R}^n$, the vector

$$\hat{v} = \frac{1}{\|v\|} v = \frac{v}{\|v\|}$$

is the unique unit vector in the same direction as v. The process of finding $\hat{v}$ from v is called *normalizing* v.

EXAMPLE 1.4

(a) Suppose $u = (1, -2, -4, 5, 3)$. To find $\|u\|$, we can first find $\|u\|^2 = u \cdot u$ by squaring each component of u and adding, as follows:

$$\|u\|^2 = 1^2 + (-2)^2 + (-4)^2 + 5^2 + 3^2 = 1 + 4 + 16 + 25 + 9 = 55$$

Then $\|u\| = \sqrt{55}$.

(b) Let $v = (1, -3, 4, 2)$ and $w = (\frac{1}{2}, -\frac{1}{6}, \frac{5}{6}, \frac{1}{6})$. Then

$$\|v\| = \sqrt{1 + 9 + 16 + 4} = \sqrt{30} \qquad \text{and} \qquad \|w\| = \sqrt{\frac{9}{36} + \frac{1}{36} + \frac{25}{36} + \frac{1}{36}} = \sqrt{\frac{36}{36}} = \sqrt{1} = 1$$

Thus w is a unit vector, but v is not a unit vector. However, we can normalize v as follows:

$$\hat{v} = \frac{v}{\|v\|} = \left(\frac{1}{\sqrt{30}}, \frac{-3}{\sqrt{30}}, \frac{4}{\sqrt{30}}, \frac{2}{\sqrt{30}}\right)$$

This is the unique unit vector in the same direction as v.

The following formula (proved in Problem 1.14) is known as the Schwarz inequality or Cauchy–Schwarz inequality. It is used in many branches of mathematics.

THEOREM 1.3 (Schwarz): For any vectors u, v in $\mathbf{R}^n$, $|u \cdot v| \leq \|u\|\|v\|$.

Using the above inequality, we also prove (Problem 1.15) the following result known as the "triangle inequality" or Minkowski's inequality.

THEOREM 1.4 (Minkowski): For any vectors u, v in $\mathbf{R}^n$, $\|u + v\| \leq \|u\| + \|v\|$.

Distance, Angles, Projections

The *distance* between vectors $u = (a_1, a_2, \ldots, a_n)$ and $v = (b_1, b_2, \ldots, b_n)$ in $\mathbf{R}^n$ is denoted and defined by

$$d(u, v) = \|u - v\| = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2}$$

One can show that this definition agrees with the usual notion of distance in the Euclidean plane $\mathbf{R}^2$ or space $\mathbf{R}^3$.

The *angle* θ between nonzero vectors u, v in $\mathbf{R}^n$ is defined by

$$\cos\theta = \frac{u \cdot v}{\|u\|\|v\|}$$

This definition is well defined, because, by the Schwarz inequality (Theorem 1.3),

$$-1 \le \frac{u \cdot v}{\|u\|\|v\|} \le 1$$

Note that if $u \cdot v = 0$, then $\theta = 90^\circ$ (or $\theta = \pi/2$). This then agrees with our previous definition of orthogonality.

The *projection* of a vector u onto a nonzero vector v is the vector denoted and defined by

$$\text{proj}(u, v) = \frac{u \cdot v}{\|v\|^2} v = \frac{u \cdot v}{v \cdot v} v$$

We show below that this agrees with the usual notion of vector projection in physics.

EXAMPLE 1.5

(a) Suppose $u = (1, -2, 3)$ and $v = (2, 4, 5)$. Then

$$d(u, v) = \sqrt{(1-2)^2 + (-2-4)^2 + (3-5)^2} = \sqrt{1 + 36 + 4} = \sqrt{41}$$

To find $\cos\theta$, where θ is the angle between u and v, we first find

$$u \cdot v = 2 - 8 + 15 = 9, \qquad \|u\|^2 = 1 + 4 + 9 = 14, \qquad \|v\|^2 = 4 + 16 + 25 = 45$$

Then

$$\cos\theta = \frac{u \cdot v}{\|u\|\|v\|} = \frac{9}{\sqrt{14}\sqrt{45}}$$

Also,

$$\text{proj}(u, v) = \frac{u \cdot v}{\|v\|^2} v = \frac{9}{45}(2, 4, 5) = \frac{1}{5}(2, 4, 5) = \left(\frac{2}{5}, \frac{4}{5}, 1\right)$$

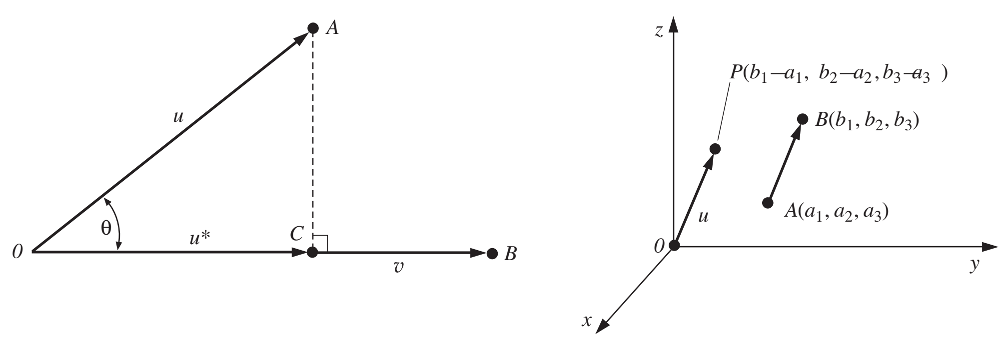
(b) Consider the vectors u and v in Fig. 1-2(a) (with respective endpoints A and B). The (perpendicular) projection of u onto v is the vector u^* with magnitude

$$\|u^*\| = \|u\| \cos\theta = \|u\| \frac{u \cdot v}{\|u\|\|v\|} = \frac{u \cdot v}{\|v\|}$$

To obtain u^*, we multiply its magnitude by the unit vector in the direction of v, obtaining

$$u^* = \|u^*\| \frac{v}{\|v\|} = \frac{u \cdot v}{\|v\|} \frac{v}{\|v\|} = \frac{u \cdot v}{\|v\|^2} v$$

This is the same as the above definition of proj(u, v).



Projection u^* of u onto v

(a)

$u = B - A$

(b)

Figure 1-2

1.5 Located Vectors, Hyperplanes, Lines, Curves in R^n

This section distinguishes between an n-tuple $P(a_i) \equiv P(a_1, a_2, \ldots, a_n)$ viewed as a point in $\mathbf{R}^n$ and an n-tuple $u = [c_1, c_2, \ldots, c_n]$ viewed as a vector (arrow) from the origin O to the point $C(c_1, c_2, \ldots, c_n)$.

Located Vectors

Any pair of points $A(a_i)$ and $B(b_i)$ in $\mathbf{R}^n$ defines the *located vector* or *directed line segment* from A to B, written $\overrightarrow{AB}$. We identify $\overrightarrow{AB}$ with the vector

$$u = B - A = [b_1 - a_1,\ b_2 - a_2,\ \ldots,\ b_n - a_n]$$

because $\overrightarrow{AB}$ and u have the same magnitude and direction. This is pictured in Fig. 1-2(b) for the points $A(a_1, a_2, a_3)$ and $B(b_1, b_2, b_3)$ in $\mathbf{R}^3$ and the vector $u = B - A$ which has the endpoint $P(b_1 - a_1,\ b_2 - a_2,\ b_3 - a_3)$.

Hyperplanes

A *hyperplane* H in $\mathbf{R}^n$ is the set of points $(x_1, x_2, \ldots, x_n)$ that satisfy a linear equation

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

where the vector $u = [a_1, a_2, \ldots, a_n]$ of coefficients is not zero. Thus a hyperplane H in $\mathbf{R}^2$ is a line, and a hyperplane H in $\mathbf{R}^3$ is a plane. We show below, as pictured in Fig. 1-3(a) for $\mathbf{R}^3$, that u is orthogonal to any directed line segment $\overrightarrow{PQ}$, where $P(p_i)$ and $Q(q_i)$ are points in H. [For this reason, we say that u is *normal* to H and that H is *normal* to u.]

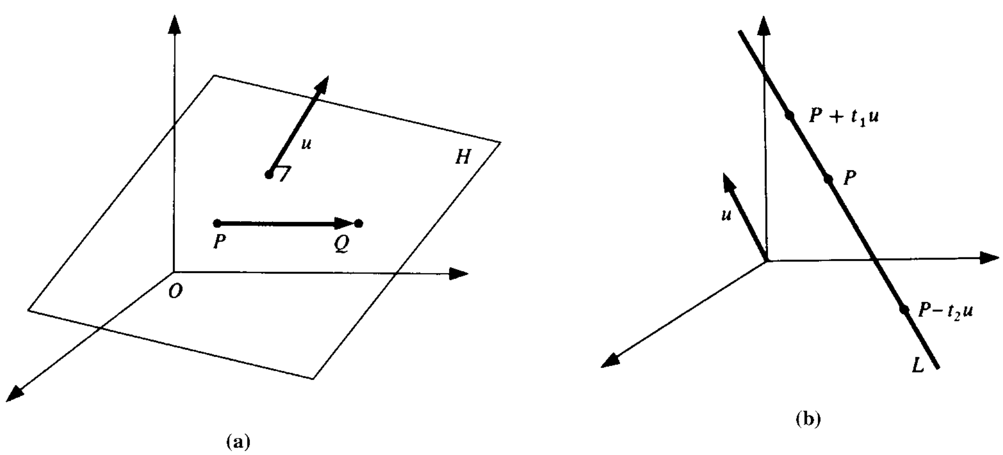


Figure 1-3

Because $P(p_i)$ and $Q(q_i)$ belong to H, they satisfy the above hyperplane equation—that is,

$$a_1p_1 + a_2p_2 + \cdots + a_np_n = b \quad \text{and} \quad a_1q_1 + a_2q_2 + \cdots + a_nq_n = b$$

Let $\quad v = \overrightarrow{PQ} = Q - P = [q_1 - p_1, q_2 - p_2, \ldots, q_n - p_n]$

Then

$$\begin{aligned} u \cdot v &= a_1(q_1 - p_1) + a_2(q_2 - p_2) + \cdots + a_n(q_n - p_n) \\ &= (a_1q_1 + a_2q_2 + \cdots + a_nq_n) - (a_1p_1 + a_2p_2 + \cdots + a_np_n) = b - b = 0 \end{aligned}$$

Thus $v = \overrightarrow{PQ}$ is orthogonal to u, as claimed.

Lines in R^n

The *line* L in $\mathbf{R}^n$ passing through the point $P(b_1, b_2, \ldots, b_n)$ and in the direction of a nonzero vector $u = [a_1, a_2, \ldots, a_n]$ consists of the points $X(x_1, x_2, \ldots, x_n)$ that satisfy

$$X = P + tu \qquad \text{or} \qquad \begin{cases} x_1 = a_1 t + b_1 \\ x_2 = a_2 t + b_2 \\ \cdots\cdots\cdots\cdots \\ x_n = a_n t + b_n \end{cases} \qquad \text{or} \quad L(t) = (a_i t + b_i)$$

where the *parameter* t takes on all real values. Such a line L in $\mathbf{R}^3$ is pictured in Fig. 1-3(b).

EXAMPLE 1.6

(a) Let H be the plane in $\mathbf{R}^3$ corresponding to the linear equation $2x - 5y + 7z = 4$. Observe that $P(1, 1, 1)$ and $Q(5, 4, 2)$ are solutions of the equation. Thus P and Q and the directed line segment

$$v = \overrightarrow{PQ} = Q - P = [5 - 1,\ 4 - 1,\ 2 - 1] = [4, 3, 1]$$

lie on the plane H. The vector $u = [2, -5, 7]$ is normal to H, and, as expected,

$$u \cdot v = [2, -5, 7] \cdot [4, 3, 1] = 8 - 15 + 7 = 0$$

That is, u is orthogonal to v.

(b) Find an equation of the hyperplane H in $\mathbf{R}^4$ that passes through the point $P(1, 3, -4, 2)$ and is normal to the vector $u = [4, -2, 5, 6]$.

The coefficients of the unknowns of an equation of H are the components of the normal vector u; hence, the equation of H must be of the form

$$4x_1 - 2x_2 + 5x_3 + 6x_4 = k$$

Substituting P into this equation, we obtain

$$4(1) - 2(3) + 5(-4) + 6(2) = k \qquad \text{or} \qquad 4 - 6 - 20 + 12 = k \qquad \text{or} \qquad k = -10$$

Thus, $4x_1 - 2x_2 + 5x_3 + 6x_4 = -10$ is the equation of H.

(c) Find the parametric representation of the line L in $\mathbf{R}^4$ passing through the point $P(1, 2, 3, -4)$ and in the direction of $u = [5, 6, -7, 8]$. Also, find the point Q on L when $t = 1$.

Substitution in the above equation for L yields the following parametric representation:

$$x_1 = 5t + 1, \qquad x_2 = 6t + 2, \qquad x_3 = -7t + 3, \qquad x_4 = 8t - 4$$

or, equivalently,

$$L(t) = (5t + 1, 6t + 2, -7t + 3, 8t - 4)$$

Note that $t = 0$ yields the point P on L. Substitution of $t = 1$ yields the point $Q(6, 8, -4, 4)$ on L.

Curves in R^n

Let D be an interval (finite or infinite) on the real line $\mathbf{R}$. A continuous function $F: D \to \mathbf{R}^n$ is a *curve* in $\mathbf{R}^n$. Thus, to each point $t \in D$ there is assigned the following point in $\mathbf{R}^n$:

$$F(t) = [F_1(t), F_2(t), \ldots, F_n(t)]$$

Moreover, the derivative (if it exists) of $F(t)$ yields the vector

$$V(t) = \frac{dF(t)}{dt} = \left[\frac{dF_1(t)}{dt}, \frac{dF_2(t)}{dt}, \ldots, \frac{dF_n(t)}{dt}\right]$$

which is tangent to the curve. Normalizing $V(t)$ yields

$$\mathbf{T}(t) = \frac{V(t)}{\|V(t)\|}$$

Thus, $\mathbf{T}(t)$ is the unit tangent vector to the curve. (Unit vectors with geometrical significance are often presented in bold type.)

EXAMPLE 1.7 Consider the curve $F(t) = [\sin t, \cos t, t]$ in $\mathbf{R}^3$. Taking the derivative of $F(t)$ [or each component of $F(t)$] yields

$$V(t) = [\cos t, -\sin t, 1]$$

which is a vector tangent to the curve. We normalize $V(t)$. First we obtain

$$\|V(t)\|^2 = \cos^2 t + \sin^2 t + 1 = 1 + 1 = 2$$

Then the unit tangent vection $\mathbf{T}(t)$ to the curve follows:

$$\mathbf{T}(t) = \frac{V(t)}{\|V(t)\|} = \left[\frac{\cos t}{\sqrt{2}}, \frac{-\sin t}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]$$

1.6 Vectors in R^3 (Spatial Vectors), ijk Notation

Vectors in $\mathbf{R}^3$, called *spatial vectors*, appear in many applications, especially in physics. In fact, a special notation is frequently used for such vectors as follows:

$\mathbf{i} = [1, 0, 0]$ denotes the unit vector in the x direction.
$\mathbf{j} = [0, 1, 0]$ denotes the unit vector in the y direction.
$\mathbf{k} = [0, 0, 1]$ denotes the unit vector in the z direction.

Then any vector $u = [a, b, c]$ in $\mathbf{R}^3$ can be expressed uniquely in the form

$$u = [a, b, c] = a\mathbf{i} + b\mathbf{j} + c\mathbf{j}$$

Because the vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are unit vectors and are mutually orthogonal, we obtain the following dot products:

$$\mathbf{i} \cdot \mathbf{i} = 1, \quad \mathbf{j} \cdot \mathbf{j} = 1, \quad \mathbf{k} \cdot \mathbf{k} = 1 \qquad \text{and} \qquad \mathbf{i} \cdot \mathbf{j} = 0, \quad \mathbf{i} \cdot \mathbf{k} = 0, \quad \mathbf{j} \cdot \mathbf{k} = 0$$

Furthermore, the vector operations discussed above may be expressed in the **ijk** notation as follows. Suppose

$$u = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k} \qquad \text{and} \qquad v = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$$

Then

$$u + v = (a_1 + b_1)\mathbf{i} + (a_2 + b_2)\mathbf{j} + (a_3 + b_3)\mathbf{k} \qquad \text{and} \qquad cu = ca_1\mathbf{i} + ca_2\mathbf{j} + ca_3\mathbf{k}$$

where c is a scalar. Also,

$$u \cdot v = a_1b_1 + a_2b_2 + a_3b_3 \qquad \text{and} \qquad \|u\| = \sqrt{u \cdot u} = a_1^2 + a_2^2 + a_3^2$$

EXAMPLE 1.8 Suppose $u = 3\mathbf{i} + 5\mathbf{j} - 2\mathbf{k}$ and $v = 4\mathbf{i} - 8\mathbf{j} + 7\mathbf{k}$.

(a) To find $u + v$, add corresponding components, obtaining $u + v = 7\mathbf{i} - 3\mathbf{j} + 5\mathbf{k}$

(b) To find $3u - 2v$, first multiply by the scalars and then add:

$$3u - 2v = (9\mathbf{i} + 13\mathbf{j} - 6\mathbf{k}) + (-8\mathbf{i} + 16\mathbf{j} - 14\mathbf{k}) = \mathbf{i} + 29\mathbf{j} - 20\mathbf{k}$$

(c) To find $u \cdot v$, multiply corresponding components and then add:

$$u \cdot v = 12 - 40 - 14 = -42$$

(d) To find $\|u\|$, take the square root of the sum of the squares of the components:

$$\|u\| = \sqrt{9 + 25 + 4} = \sqrt{38}$$

Cross Product

There is a special operation for vectors u and v in $\mathbf{R}^3$ that is not defined in $\mathbf{R}^n$ for $n \neq 3$. This operation is called the *cross product* and is denoted by $u \times v$. One way to easily remember the formula for $u \times v$ is to use the determinant (of order two) and its negative, which are denoted and defined as follows:

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

Here a and d are called the *diagonal* elements and b and c are the *nondiagonal* elements. Thus, the determinant is the product ad of the diagonal elements minus the product bc of the nondiagonal elements, but vice versa for the negative of the determinant.

Now suppose $u = a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ and $v = b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$. Then

$$\begin{aligned} u \times v &= (a_2b_3 - a_3b_2)\mathbf{i} + (a_3b_1 - a_1b_3)\mathbf{j} + (a_1b_2 - a_2b_1)\mathbf{k} \\ &= \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}\mathbf{i} - \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}\mathbf{j} + \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}\mathbf{i} \end{aligned}$$

That is, the three components of $u \times v$ are obtained from the array

$$\begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix}$$

(which contain the components of u above the component of v) as follows:

(1) Cover the first column and take the determinant.
(2) Cover the second column and take the negative of the determinant.
(3) Cover the third column and take the determinant.

Note that $u \times v$ is a vector; hence, $u \times v$ is also called the *vector product* or *outer product* of u and v.

EXAMPLE 1.9 Find $u \times v$ where: (a) $u = 4\mathbf{i} + 3\mathbf{j} + 6\mathbf{k}$, $v = 2\mathbf{i} + 5\mathbf{j} - 3\mathbf{k}$, (b) $u = [2, -1, 5]$, $v = [3, 7, 6]$.

(a) Use $\begin{bmatrix} 4 & 3 & 6 \\ 2 & 5 & -3 \end{bmatrix}$ to get $u \times v = (-9 - 30)\mathbf{i} + (12 + 12)\mathbf{j} + (20 - 6)\mathbf{k} = -39\mathbf{i} + 24\mathbf{j} + 14\mathbf{k}$

(b) Use $\begin{bmatrix} 2 & -1 & 5 \\ 3 & 7 & 6 \end{bmatrix}$ to get $u \times v = [-6 - 35, 15 - 12, 14 + 3] = [-41, 3, 17]$

Remark: The cross products of the vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are as follows:

$$\mathbf{i} \times \mathbf{j} = \mathbf{k}, \qquad \mathbf{j} \times \mathbf{k} = \mathbf{i}, \qquad \mathbf{k} \times \mathbf{i} = \mathbf{j}$$
$$\mathbf{j} \times \mathbf{i} = -\mathbf{k}, \qquad \mathbf{k} \times \mathbf{j} = -\mathbf{i}, \qquad \mathbf{i} \times \mathbf{k} = -\mathbf{j}$$

Thus, if we view the triple $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ as a cyclic permutation, where $\mathbf{i}$ follows $\mathbf{k}$ and hence $\mathbf{k}$ precedes $\mathbf{i}$, then the product of two of them in the given direction is the third one, but the product of two of them in the opposite direction is the negative of the third one.

Two important properties of the cross product are contained in the following theorem.

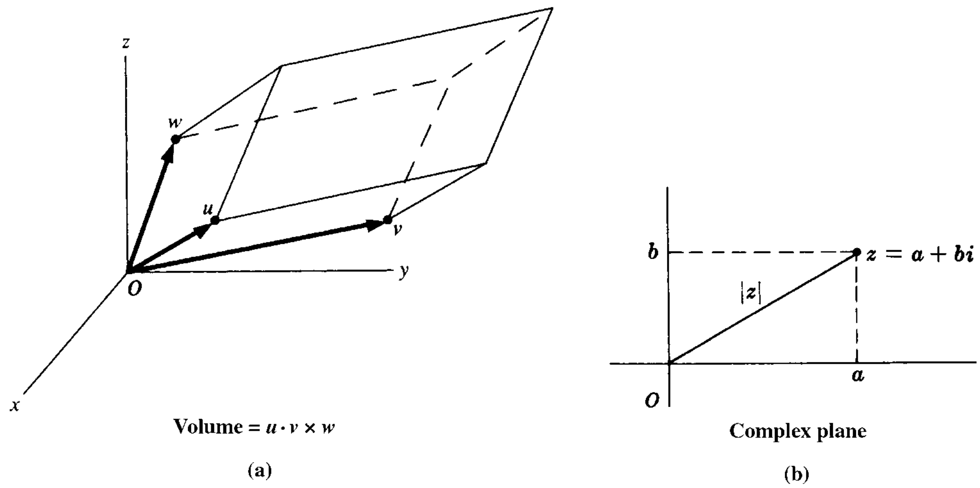


Figure 1-4

THEOREM 1.5: Let u, v, w be vectors in $\mathbf{R}^3$.

(a) The vector $u \times v$ is orthogonal to both u and v.

(b) The absolute value of the "triple product"

$$u \cdot v \times w$$

represents the volume of the parallelopiped formed by the vectors u, v, w. [See Fig. 1-4(a).]

We note that the vectors $u, v, u \times v$ form a right-handed system, and that the following formula gives the magnitude of $u \times v$:

$$\|u \times v\| = \|u\|\|v\| \sin \theta$$

where θ is the angle between u and v.

1.7 Complex Numbers

The set of complex numbers is denoted by $\mathbf{C}$. Formally, a complex number is an ordered pair (a, b) of real numbers where equality, addition, and multiplication are defined as follows:

$$(a, b) = (c, d) \quad \text{if and only if } a = c \text{ and } b = d$$
$$(a, b) + (c, d) = (a + c,\ b + d)$$
$$(a, b) \cdot (c, d) = (ac - bd,\ ad + bc)$$

We identify the real number a with the complex number $(a, 0)$; that is,

$$a \leftrightarrow (a, 0)$$

This is possible because the operations of addition and multiplication of real numbers are preserved under the correspondence; that is,

$$(a, 0) + (b, 0) = (a + b,\ 0) \qquad \text{and} \qquad (a, 0) \cdot (b, 0) = (ab, 0)$$

Thus we view $\mathbf{R}$ as a subset of $\mathbf{C}$, and replace $(a, 0)$ by a whenever convenient and possible.

We note that the set $\mathbf{C}$ of complex numbers with the above operations of addition and multiplication is a *field* of numbers, like the set $\mathbf{R}$ of real numbers and the set $\mathbf{Q}$ of *rational numbers*.

The complex number $(0, 1)$ is denoted by i. It has the important property that

$$i^2 = ii = (0,1)(0,1) = (-1, 0) = -1 \qquad \text{or} \qquad i = \sqrt{-1}$$

Accordingly, any complex number $z = (a, b)$ can be written in the form

$$z = (a, b) = (a, 0) + (0, b) = (a, 0) + (b, 0) \cdot (0, 1) = a + bi$$

The above notation $z = a + bi$, where $a \equiv \text{Re}\, z$ and $b \equiv \text{Im}\, z$ are called, respectively, the *real* and *imaginary parts* of z, is more convenient than (a, b). In fact, the sum and product of complex numbers $z = a + bi$ and $w = c + di$ can be derived by simply using the commutative and distributive laws and $i^2 = -1$:

$$\begin{aligned} z + w &= (a + bi) + (c + di) = a + c + bi + di = (a + b) + (c + d)i \\ zw &= (a + bi)(c + di) = ac + bci + adi + bdi^2 = (ac - bd) + (bc + ad)i \end{aligned}$$

We also define the *negative* of z and subtraction in $\mathbf{C}$ by

$$-z = -1z \qquad \text{and} \qquad w - z = w + (-z)$$

Warning: The letter i representing $\sqrt{-1}$ has no relationship whatsoever to the vector $\mathbf{i} = [1, 0, 0]$ in Section 1.6.

Complex Conjugate, Absolute Value

Consider a complex number $z = a + bi$. The *conjugate* of z is denoted and defined by

$$\bar{z} = \overline{a + bi} = a - bi$$

Then $z\bar{z} = (a + bi)(a - bi) = a^2 - b^2i^2 = a^2 + b^2$. Note that z is real if and only if $\bar{z} = z$.

The *absolute value* of z, denoted by $|z|$, is defined to be the nonnegative square root of $z\bar{z}$. Namely,

$$|z| = \sqrt{z\bar{z}} = \sqrt{a^2 + b^2}$$

Note that $|z|$ is equal to the norm of the vector (a, b) in $\mathbf{R}^2$.

Suppose $z \neq 0$. Then the inverse z^{-1} of z and division in $\mathbf{C}$ of w by z are given, respectively, by

$$z^{-1} = \frac{\bar{z}}{z\bar{z}} = \frac{a}{a^2 + b^2} - \frac{b}{a^2 + b^2}i \qquad \text{and} \qquad \frac{w}{z} - \frac{w\bar{z}}{z\bar{z}} = wz^{-1}$$

EXAMPLE 1.10 Suppose $z = 2 + 3i$ and $w = 5 - 2i$. Then

$$\begin{aligned} z + w &= (2 + 3i) + (5 - 2i) = 2 + 5 + 3i - 2i = 7 + i \\ zw &= (2 + 3i)(5 - 2i) = 10 + 15i - 4i - 6i^2 = 16 + 11i \\ \bar{z} &= \overline{2 + 3i} = 2 - 3i \qquad \text{and} \qquad \bar{w} = \overline{5 - 2i} = 5 + 2i \\ \frac{w}{z} &= \frac{5 - 2i}{2 + 3i} = \frac{(5 - 2i)(2 - 3i)}{(2 + 3i)(2 - 3i)} = \frac{4 - 19i}{13} = \frac{4}{13} - \frac{19}{13}i \\ |z| &= \sqrt{4 + 9} = \sqrt{13} \qquad \text{and} \qquad |w| = \sqrt{25 + 4} = \sqrt{29} \end{aligned}$$

Complex Plane

Recall that the real numbers $\mathbf{R}$ can be represented by points on a line. Analogously, the complex numbers $\mathbf{C}$ can be represented by points in the plane. Specifically, we let the point (a, b) in the plane represent the complex number $a + bi$ as shown in Fig. 1-4(b). In such a case, $|z|$ is the distance from the origin O to the point z. The plane with this representation is called the *complex plane*, just like the line representing $\mathbf{R}$ is called the *real line*.

1.8 Vectors in $\mathbf{C}^n$

The set of all n-tuples of complex numbers, denoted by $\mathbf{C}^n$, is called *complex* n-*space*. Just as in the real case, the elements of $\mathbf{C}^n$ are called *points* or *vectors*, the elements of $\mathbf{C}$ are called *scalars*, and vector addition in $\mathbf{C}^n$ and scalar multiplication on $\mathbf{C}^n$ are given by

$$[z_1, z_2, \ldots, z_n] + [w_1, w_2, \ldots, w_n] = [z_1 + w_1,\ z_2 + w_2,\ \ldots,\ z_n + w_n]$$
$$z[z_1, z_2, \ldots, z_n] = [zz_1, zz_2, \ldots, zz_n]$$

where the z_i, w_i, and z belong to $\mathbf{C}$.

EXAMPLE 1.11 Consider vectors $u = [2+3i,\ 4-i,\ 3]$ and $v = [3-2i,\ 5i,\ 4-6i]$ in $\mathbf{C}^3$. Then

$$u + v = [2+3i,\ 4-i,\ 3] + [3-2i,\ 5i,\ 4-6i] = [5+i,\ 4+4i,\ 7-6i]$$
$$(5-2i)u = [(5-2i)(2+3i),\ (5-2i)(4-i),\ (5-2i)(3)] = [16+11i,\ 18-13i,\ 15-6i]$$

Dot (Inner) Product in $\mathbf{C}^n$

Consider vectors $u = [z_1, z_2, \ldots, z_n]$ and $v = [w_1, w_2, \ldots, w_n]$ in $\mathbf{C}^n$. The *dot* or *inner product* of u and v is denoted and defined by

$$u \cdot v = z_1\bar{w}_1 + z_2\bar{w}_2 + \cdots + z_n\bar{w}_n$$

This definition reduces to the real case because $\bar{w}_i = w_i$ when w_i is real. The norm of u is defined by

$$\|u\| = \sqrt{u \cdot u} = \sqrt{z_1\bar{z}_1 + z_2\bar{z}_2 + \cdots + z_n\bar{z}_n} = \sqrt{|z_1|^2 + |z_2|^2 + \cdots + |v_n|^2}$$

We emphasize that $u \cdot u$ and so $\|u\|$ are real and positive when $u \neq 0$ and 0 when $u = 0$.

EXAMPLE 1.12 Consider vectors $u = [2+3i,\ 4-i,\ 3+5i]$ and $v = [3-4i,\ 5i,\ 4-2i]$ in $\mathbf{C}_3$. Then

$$\begin{aligned} u \cdot v &= (2+3i)(\overline{3-4i}) + (4-i)(\overline{5i}) + (3+5i)(\overline{4-2i}) \\ &= (2+3i)(3+4i) + (4-i)(-5i) + (3+5i)(4+2i) \\ &= (-6+13i) + (-5-20i) + (2+26i) = -9+19i \\ u \cdot u &= |2+3i|^2 + |4-i|^2 + |3+5i|^2 = 4+9+16+1+9+25 = 64 \\ \|u\| &= \sqrt{64} = 8 \end{aligned}$$

The space $\mathbf{C}^n$ with the above operations of vector addition, scalar multiplication, and dot product, is called *complex Euclidean* n-*space*. Theorem 1.2 for $\mathbf{R}^n$ also holds for $\mathbf{C}^n$ if we replace $u \cdot v = v \cdot u$ by

$$u \cdot v = \overline{u \cdot v}$$

On the other hand, the Schwarz inequality (Theorem 1.3) and Minkowski's inequality (Theorem 1.4) are true for $\mathbf{C}^n$ with no changes.

SOLVED PROBLEMS

Vectors in $\mathbf{R}^n$

1.1. Determine which of the following vectors are equal:

$$u_1 = (1, 2, 3), \qquad u_2 = (2, 3, 1), \qquad u_3 = (1, 3, 2), \qquad u_4 = (2, 3, 1)$$

Vectors are equal only when corresponding entries are equal; hence, only $u_2 = u_4$.

CHAPTER 2

Algebra of Matrices

2.1 Introduction

This chapter investigates matrices and algebraic operations defined on them. These matrices may be viewed as rectangular arrays of elements where each entry depends on two subscripts (as compared with vectors, where each entry depended on only one subscript). Systems of linear equations and their solutions (Chapter 3) may be efficiently investigated using the language of matrices. Furthermore, certain abstract objects introduced in later chapters, such as ''change of basis,'' ''linear transformations,'' and ''quadratic forms,'' can be represented by these matrices (rectangular arrays). On the other hand, the abstract treatment of linear algebra presented later on will give us new insight into the structure of these matrices.

The entries in our matrices will come from some arbitrary, but fixed, field K. The elements of K are called *numbers* or *scalars*. Nothing essential is lost if the reader assumes that K is the real field $\mathbf{R}$.

2.2 Matrices

A *matrix A over a field K* or, simply, a *matrix A* (when K is implicit) is a rectangular array of scalars usually presented in the following form:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

The *rows* of such a matrix A are the m horizontal lists of scalars:

$$(a_{11}, a_{12}, \dots, a_{1n}), \quad (a_{21}, a_{22}, \dots, a_{2n}), \quad \dots, \quad (a_{m1}, a_{m2}, \dots, a_{mn})$$

and the *columns* of A are the n vertical lists of scalars:

$$\begin{bmatrix} a_{11} \\ a_{21} \\ \dots \\ a_{m1} \end{bmatrix}, \quad \begin{bmatrix} a_{12} \\ a_{22} \\ \dots \\ a_{m2} \end{bmatrix}, \quad \dots, \quad \begin{bmatrix} a_{1n} \\ a_{2n} \\ \dots \\ a_{mn} \end{bmatrix}$$

Note that the element a_{ij}, called the *ij-entry* or *ij-element*, appears in row i and column j. We frequently denote such a matrix by simply writing $A = [a_{ij}]$.

A matrix with m rows and n columns is called an *m by n* matrix, written $m \times n$. The pair of numbers m and n is called the *size* of the matrix. Two matrices A and B are *equal*, written $A = B$, if they have the same size and if corresponding elements are equal. Thus, the equality of two $m \times n$ matrices is equivalent to a system of mn equalities, one for each corresponding pair of elements.

A matrix with only one row is called a *row matrix* or *row vector*, and a matrix with only one column is called a *column matrix* or *column vector*. A matrix whose entries are all zero is called a *zero matrix* and will usually be denoted by 0.

Matrices whose entries are all real numbers are called *real matrices* and are said to be *matrices over* **R**. Analogously, matrices whose entries are all complex numbers are called *complex matrices* and are said to be *matrices over* **C**. This text will be mainly concerned with such real and complex matrices.

EXAMPLE 2.1

(a) The rectangular array $A = \begin{bmatrix} 1 & -4 & 5 \\ 0 & 3 & -2 \end{bmatrix}$ is a 2×3 matrix. Its rows are $(1, -4, 5)$ and $(0, 3, -2)$, and its columns are

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

(b) The 2×4 zero matrix is the matrix $0 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

(c) Find x, y, z, t such that

$$\begin{bmatrix} x+y & 2z+t \\ x-y & z-t \end{bmatrix} = \begin{bmatrix} 3 & 7 \\ 1 & 5 \end{bmatrix}$$

By definition of equality of matrices, the four corresponding entries must be equal. Thus,

$$x + y = 3, \qquad x - y = 1, \qquad 2z + t = 7, \qquad z - t = 5$$

Solving the above system of equations yields $x = 2$, $y = 1$, $z = 4$, $t = -1$.

2.3 Matrix Addition and Scalar Multiplication

Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be two matrices with the same size, say $m \times n$ matrices. The *sum* of A and B, written $A + B$, is the matrix obtained by adding corresponding elements from A and B. That is,

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \dots & a_{mn} + b_{mn} \end{bmatrix}$$

The *product* of the matrix A by a scalar k, written $k \cdot A$ or simply kA, is the matrix obtained by multiplying each element of A by k. That is,

$$kA = \begin{bmatrix} ka_{11} & ka_{12} & \dots & ka_{1n} \\ ka_{21} & ka_{22} & \dots & ka_{2n} \\ \dots & \dots & \dots & \dots \\ ka_{m1} & ka_{m2} & \dots & ka_{mn} \end{bmatrix}$$

Observe that $A + B$ and kA are also $m \times n$ matrices. We also define

$$-A = (-1)A \qquad \text{and} \qquad A - B = A + (-B)$$

The matrix $-A$ is called the *negative* of the matrix A, and the matrix $A - B$ is called the *difference* of A and B. The sum of matrices with different sizes is not defined.

EXAMPLE 2.2 Let $A = \begin{bmatrix} 1 & -2 & 3 \\ 0 & 4 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} 4 & 6 & 8 \\ 1 & -3 & -7 \end{bmatrix}$. Then

$$A + B = \begin{bmatrix} 1+4 & -2+6 & 3+8 \\ 0+1 & 4+(-3) & 5+(-7) \end{bmatrix} = \begin{bmatrix} 5 & 4 & 11 \\ 1 & 1 & -2 \end{bmatrix}$$

$$3A = \begin{bmatrix} 3(1) & 3(-2) & 3(3) \\ 3(0) & 3(4) & 3(5) \end{bmatrix} = \begin{bmatrix} 3 & -6 & 9 \\ 0 & 12 & 15 \end{bmatrix}$$

$$2A - 3B = \begin{bmatrix} 2 & -4 & 6 \\ 0 & 8 & 10 \end{bmatrix} + \begin{bmatrix} -12 & -18 & -24 \\ -3 & 9 & 21 \end{bmatrix} = \begin{bmatrix} -10 & -22 & -18 \\ -3 & 17 & 31 \end{bmatrix}$$

The matrix $2A - 3B$ is called a *linear combination* of A and B.

Basic properties of matrices under the operations of matrix addition and scalar multiplication follow.

THEOREM 2.1: Consider any matrices A, B, C (with the same size) and any scalars k and k'. Then

(i) $(A + B) + C = A + (B + C)$, (v) $k(A + B) = kA + kB$,

(ii) $A + 0 = 0 + A = A$, (vi) $(k + k')A = kA + k'A$,

(iii) $A + (-A) = (-A) + A = 0$, (vii) $(kk')A = k(k'A)$,

(iv) $A + B = B + A$, (viii) $1 \cdot A = A$.

Note first that the 0 in (ii) and (iii) refers to the zero matrix. Also, by (i) and (iv), any sum of matrices

$$A_1 + A_2 + \cdots + A_n$$

requires no parentheses, and the sum does not depend on the order of the matrices. Furthermore, using (vi) and (viii), we also have

$$A + A = 2A, \qquad A + A + A = 3A, \qquad \ldots$$

and so on.

The proof of Theorem 2.1 reduces to showing that the *ij*-entries on both sides of each matrix equation are equal. (See Problem 2.3.)

Observe the similarity between Theorem 2.1 for matrices and Theorem 1.1 for vectors. In fact, the above operations for matrices may be viewed as generalizations of the corresponding operations for vectors.

2.4 Summation Symbol

Before we define matrix multiplication, it will be instructive to first introduce the *summation symbol* Σ (the Greek capital letter sigma).

Suppose $f(k)$ is an algebraic expression involving the letter k. Then the expression

$$\sum_{k=1}^{n} f(k) \qquad \text{or equivalently} \qquad \textstyle\sum_{k=1}^{n} f(k)$$

has the following meaning. First we set $k = 1$ in $f(k)$, obtaining

$$f(1)$$

Then we set $k = 2$ in $f(k)$, obtaining $f(2)$, and add this to $f(1)$, obtaining

$$f(1) + f(2)$$

Then we set $k = 3$ in $f(k)$, obtaining $f(3)$, and add this to the previous sum, obtaining

$$f(1) + f(2) + f(3)$$

We continue this process until we obtain the sum

$$f(1) + f(2) + \cdots + f(n)$$

Observe that at each step we increase the value of k by 1 until we reach n. The letter k is called the *index*, and 1 and n are called, respectively, the *lower* and *upper* limits. Other letters frequently used as indices are i and j.

We also generalize our definition by allowing the sum to range from any integer n_1 to any integer n_2. That is, we define

$$\sum_{k=n_1}^{n_2} f(k) = f(n_1) + f(n_1 + 1) + f(n_1 + 2) + \cdots + f(n_2)$$

EXAMPLE 2.3

(a) $\sum_{k=1}^{5} x_k = x_1 + x_2 + x_3 + x_4 + x_5$ and $\sum_{i=1}^{n} a_i b_i = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$

(b) $\sum_{j=2}^{5} j^2 = 2^2 + 3^2 + 4^2 + 5^2 = 54$ and $\sum_{i=0}^{n} a_i x^i = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$

(c) $\sum_{k=1}^{p} a_{ik} b_{kj} = a_{i1} b_{1j} + a_{i2} b_{2j} + a_{i3} b_{3j} + \cdots + a_{ip} b_{pj}$

2.5 Matrix Multiplication

The product of matrices A and B, written AB, is somewhat complicated. For this reason, we first begin with a special case.

The product AB of a row matrix $A = [a_i]$ and a column matrix $B = [b_i]$ with the same number of elements is defined to be the scalar (or 1×1 matrix) obtained by multiplying corresponding entries and adding; that is,

$$AB = [a_1, a_2, \ldots, a_n] \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_n \end{bmatrix} = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = \sum_{k=1}^{n} a_k b_k$$

We emphasize that AB is a scalar (or a 1×1 matrix). The product AB is not defined when A and B have different numbers of elements.

EXAMPLE 2.4

(a) $[7, -4, 5] \begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix} = 7(3) + (-4)(2) + 5(-1) = 21 - 8 - 5 = 8$

(b) $[6, -1, 8, 3] \begin{bmatrix} 4 \\ -9 \\ -2 \\ 5 \end{bmatrix} = 24 + 9 - 16 + 15 = 32$

We are now ready to define matrix multiplication in general.

DEFINITION: Suppose $A = [a_{ik}]$ and $B = [b_{kj}]$ are matrices such that the number of columns of A is equal to the number of rows of B; say, A is an $m \times p$ matrix and B is a $p \times n$ matrix. Then the product AB is the $m \times n$ matrix whose ij-entry is obtained by multiplying the ith row of A by the jth column of B. That is,

$$\begin{bmatrix} a_{11} & \dots & a_{1p} \\ \cdot & \dots & \cdot \\ a_{i1} & \dots & a_{ip} \\ \cdot & \dots & \cdot \\ a_{m1} & \dots & a_{mp} \end{bmatrix} \begin{bmatrix} b_{11} & \dots & b_{1j} & \dots & b_{1n} \\ \cdot & \dots & \cdot & \dots & \cdot \\ \cdot & \dots & \cdot & \dots & \cdot \\ \cdot & \dots & \cdot & \dots & \cdot \\ b_{p1} & \dots & b_{pj} & \dots & b_{pn} \end{bmatrix} = \begin{bmatrix} c_{11} & \dots & c_{1n} \\ \cdot & \dots & \cdot \\ \cdot & c_{ij} & \cdot \\ \cdot & \dots & \cdot \\ c_{m1} & \dots & c_{mn} \end{bmatrix}$$

where $\quad c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ip}b_{pj} = \sum_{k=1}^{p} a_{ik}b_{kj}$

The product AB is not defined if A is an $m \times p$ matrix and B is a $q \times n$ matrix, where $p \neq q$.

EXAMPLE 2.5

(a) Find AB where $A = \begin{bmatrix} 1 & 3 \\ 2 & -1 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 0 & -4 \\ 5 & -2 & 6 \end{bmatrix}$.

Because A is 2×2 and B is 2×3, the product AB is defined and AB is a 2×3 matrix. To obtain the first row of the product matrix AB, multiply the first row [1, 3] of A by each column of B,

$$\begin{bmatrix} 2 \\ 5 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ -2 \end{bmatrix}, \quad \begin{bmatrix} -4 \\ 6 \end{bmatrix}$$

respectively. That is,

$$AB = \begin{bmatrix} 2+15 & 0-6 & -4+18 \\ & & \end{bmatrix} = \begin{bmatrix} 17 & -6 & 14 \\ & & \end{bmatrix}$$

To obtain the second row of AB, multiply the second row $[2, -1]$ of A by each column of B. Thus,

$$AB = \begin{bmatrix} 17 & -6 & 14 \\ 4-5 & 0+2 & -8-6 \end{bmatrix} = \begin{bmatrix} 17 & -6 & 14 \\ -1 & 2 & -14 \end{bmatrix}$$

(b) Suppose $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 5 & 6 \\ 0 & -2 \end{bmatrix}$. Then

$$AB = \begin{bmatrix} 5+0 & 6-4 \\ 15+0 & 18-8 \end{bmatrix} = \begin{bmatrix} 5 & 2 \\ 15 & 10 \end{bmatrix} \quad \text{and} \quad BA = \begin{bmatrix} 5+18 & 10+24 \\ 0-6 & 0-8 \end{bmatrix} = \begin{bmatrix} 23 & 34 \\ -6 & -8 \end{bmatrix}$$

The above example shows that matrix multiplication is not commutative—that is, in general, $AB \neq BA$. However, matrix multiplication does satisfy the following properties.

THEOREM 2.2: Let A, B, C be matrices. Then, whenever the products and sums are defined,

(i) $(AB)C = A(BC)$ (associative law),

(ii) $A(B + C) = AB + AC$ (left distributive law),

(iii) $(B + C)A = BA + CA$ (right distributive law),

(iv) $k(AB) = (kA)B = A(kB)$, where k is a scalar.

We note that $0A = 0$ and $B0 = 0$, where 0 is the zero matrix.

2.6 Transpose of a Matrix

The *transpose* of a matrix A, written A^T, is the matrix obtained by writing the columns of A, in order, as rows. For example,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \qquad \text{and} \qquad [1, -3, -5]^T = \begin{bmatrix} 1 \\ -3 \\ -5 \end{bmatrix}$$

In other words, if $A = [a_{ij}]$ is an $m \times n$ matrix, then $A^T = [b_{ij}]$ is the $n \times m$ matrix where $b_{ij} = a_{ji}$.

Observe that the tranpose of a row vector is a column vector. Similarly, the transpose of a column vector is a row vector.

The next theorem lists basic properties of the transpose operation.

THEOREM 2.3: Let A and B be matrices and let k be a scalar. Then, whenever the sum and product are defined,

(i) $(A+B)^T = A^T + B^T$, (iii) $(kA)^T = kA^T$,

(ii) $(A^T)^T = A$, (iv) $(AB)^T = B^T A^T$.

We emphasize that, by (iv), the transpose of a product is the product of the transposes, but in the reverse order.

2.7 Square Matrices

A *square matrix* is a matrix with the same number of rows as columns. An $n \times n$ square matrix is said to be of *order* n and is sometimes called an n-*square matrix*.

Recall that not every two matrices can be added or multiplied. However, if we only consider square matrices of some given order n, then this inconvenience disappears. Specifically, the operations of addition, multiplication, scalar multiplication, and transpose can be performed on any $n \times n$ matrices, and the result is again an $n \times n$ matrix.

EXAMPLE 2.6 The following are square matrices of order 3:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -4 & -4 & -4 \\ 5 & 6 & 7 \end{bmatrix} \qquad \text{and} \qquad B = \begin{bmatrix} 2 & -5 & 1 \\ 0 & 3 & -2 \\ 1 & 2 & -4 \end{bmatrix}$$

The following are also matrices of order 3:

$$A + B = \begin{bmatrix} 3 & -3 & 4 \\ -4 & -1 & -6 \\ 6 & 8 & 3 \end{bmatrix}, \qquad 2A = \begin{bmatrix} 2 & 4 & 6 \\ -8 & -8 & -8 \\ 10 & 12 & 14 \end{bmatrix}, \qquad A^T = \begin{bmatrix} 1 & -4 & 5 \\ 2 & -4 & 6 \\ 3 & -4 & 7 \end{bmatrix}$$

$$AB = \begin{bmatrix} 5 & 7 & -15 \\ -12 & 0 & 20 \\ 17 & 7 & -35 \end{bmatrix}, \qquad BA = \begin{bmatrix} 27 & 30 & 33 \\ -22 & -24 & -26 \\ -27 & -30 & -33 \end{bmatrix}$$

Diagonal and Trace

Let $A = [a_{ij}]$ be an n-square matrix. The *diagonal* or *main diagonal* of A consists of the elements with the same subscripts—that is,

$$a_{11}, \quad a_{22}, \quad a_{33}, \quad \ldots, \quad a_{nn}$$

The *trace* of A, written $\text{tr}(A)$, is the sum of the diagonal elements. Namely,

$$\text{tr}(A) = a_{11} + a_{22} + a_{33} + \cdots + a_{nn}$$

The following theorem applies.

THEOREM 2.4: Suppose $A = [a_{ij}]$ and $B = [b_{ij}]$ are n-square matrices and k is a scalar. Then

(i) $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$, (iii) $\text{tr}(A^T) = \text{tr}(A)$,

(ii) $\text{tr}(kA) = k\,\text{tr}(A)$, (iv) $\text{tr}(AB) = \text{tr}(BA)$.

EXAMPLE 2.7 Let A and B be the matrices A and B in Example 2.6. Then

$$\text{diagonal of } A = \{1, -4, 7\} \quad \text{and} \quad \text{tr}(A) = 1 - 4 + 7 = 4$$
$$\text{diagonal of } B = \{2, 3, -4\} \quad \text{and} \quad \text{tr}(B) = 2 + 3 - 4 = 1$$

Moreover,

$$\text{tr}(A + B) = 3 - 1 + 3 = 5, \qquad \text{tr}(2A) = 2 - 8 + 14 = 8, \quad \text{tr}(A^T) = 1 - 4 + 7 = 4$$
$$\text{tr}(AB) = 5 + 0 - 35 = -30, \qquad \text{tr}(BA) = 27 - 24 - 33 = -30$$

As expected from Theorem 2.4,

$$\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B), \qquad \text{tr}(A^T) = \text{tr}(A), \qquad \text{tr}(2A) = 2\,\text{tr}(A)$$

Furthermore, although $AB \neq BA$, the traces are equal.

Identity Matrix, Scalar Matrices

The n-square *identity* or *unit* matrix, denoted by I_n, or simply I, is the n-square matrix with 1's on the diagonal and 0's elsewhere. The identity matrix I is similar to the scalar 1 in that, for any n-square matrix A,

$$AI = IA = A$$

More generally, if B is an $m \times n$ matrix, then $BI_n = I_m B = B$.

For any scalar k, the matrix kI that contains k's on the diagonal and 0's elsewhere is called the *scalar matrix* corresponding to the scalar k. Observe that

$$(kI)A = k(IA) = kA$$

That is, multiplying a matrix A by the scalar matrix kI is equivalent to multiplying A by the scalar k.

EXAMPLE 2.8 The following are the identity matrices of orders 3 and 4 and the corresponding scalar matrices for $k = 5$:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}, \quad \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix}, \quad \begin{bmatrix} 5 & & & \\ & 5 & & \\ & & 5 & \\ & & & 5 \end{bmatrix}$$

Remark 1: It is common practice to omit blocks or patterns of 0's when there is no ambiguity, as in the above second and fourth matrices.

Remark 2: The *Kronecker delta function* δ_{ij} is defined by

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

Thus, the identity matrix may be defined by $I = [\delta_{ij}]$.

2.8 Powers of Matrices, Polynomials in Matrices

Let A be an n-square matrix over a field K. *Powers* of A are defined as follows:

$$A^2 = AA, \qquad A^3 = A^2A, \qquad \dots, \qquad A^{n+1} = A^nA, \qquad \dots, \qquad \text{and} \qquad A^0 = I$$

Polynomials in the matrix A are also defined. Specifically, for any polynomial

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

where the a_i are scalars in K, $f(A)$ is defined to be the following matrix:

$$f(A) = a_0I + a_1A + a_2A^2 + \cdots + a_nA^n$$

[Note that $f(A)$ is obtained from $f(x)$ by substituting the matrix A for the variable x and substituting the scalar matrix a_0I for the scalar a_0.] If $f(A)$ is the zero matrix, then A is called a *zero* or *root* of $f(x)$.

EXAMPLE 2.9 Suppose $A = \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix}$. Then

$$A^2 = \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix} = \begin{bmatrix} 7 & -6 \\ -9 & 22 \end{bmatrix} \quad \text{and} \quad A^3 = A^2A = \begin{bmatrix} 7 & -6 \\ -9 & 22 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix} = \begin{bmatrix} -11 & 38 \\ 57 & -106 \end{bmatrix}$$

Suppose $f(x) = 2x^2 - 3x + 5$ and $g(x) = x^2 + 3x - 10$. Then

$$f(A) = 2\begin{bmatrix} 7 & -6 \\ -9 & 22 \end{bmatrix} - 3\begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix} + 5\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 16 & -18 \\ -27 & 61 \end{bmatrix}$$

$$g(A) = \begin{bmatrix} 7 & -6 \\ -9 & 22 \end{bmatrix} + 3\begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix} - 10\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Thus, A is a zero of the polynomial $g(x)$.

2.9 Invertible (Nonsingular) Matrices

A square matrix A is said to be *invertible* or *nonsingular* if there exists a matrix B such that

$$AB = BA = I$$

where I is the identity matrix. Such a matrix B is unique. That is, if $AB_1 = B_1A = I$ and $AB_2 = B_2A = I$, then

$$B_1 = B_1I = B_1(AB_2) = (B_1A)B_2 = IB_2 = B_2$$

We call such a matrix B the *inverse* of A and denote it by A^{-1}. Observe that the above relation is symmetric; that is, if B is the inverse of A, then A is the inverse of B.

EXAMPLE 2.10 Suppose that $A = \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix}$. Then

$$AB = \begin{bmatrix} 6-5 & -10+10 \\ 3-3 & -5+6 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \text{and} \qquad BA = \begin{bmatrix} 6-5 & 15-15 \\ -2+2 & -5+6 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Thus, A and B are inverses.

It is known (Theorem 3.16) that $AB = I$ if and only if $BA = I$. Thus, it is necessary to test only one product to determine whether or not two given matrices are inverses. (See Problem 2.17.)

Now suppose A and B are invertible. Then AB is invertible and $(AB)^{-1} = B^{-1}A^{-1}$. More generally, if $A_1, A_2, \dots, A_k$ are invertible, then their product is invertible and

$$(A_1A_2\dots A_k)^{-1} = A_k^{-1}\dots A_2^{-1}A_1^{-1}$$

the product of the inverses in the reverse order.

Inverse of a 2 × 2 Matrix

Let A be an arbitrary 2×2 matrix, say $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. We want to derive a formula for A^{-1}, the inverse of A. Specifically, we seek $2^2 = 4$ scalars, say x_1, y_1, x_2, y_2, such that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} ax_1 + by_1 & ax_2 + by_2 \\ cx_1 + dy_1 & cx_2 + dy_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Setting the four entries equal to the corresponding entries in the identity matrix yields four equations, which can be partitioned into two 2×2 systems as follows:

$$ax_1 + by_1 = 1, \qquad ax_2 + by_2 = 0$$
$$cx_1 + dy_1 = 0, \qquad cx_2 + dy_2 = 1$$

Suppose we let $|A| = ab - bc$ (called the *determinant* of A). Assuming $|A| \neq 0$, we can solve uniquely for the above unknowns x_1, y_1, x_2, y_2, obtaining

$$x_1 = \frac{d}{|A|}, \qquad y_1 = \frac{-c}{|A|}, \qquad x_2 = \frac{-b}{|A|}, \qquad y_2 = \frac{a}{|A|}$$

Accordingly,

$$A^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \begin{bmatrix} d/|A| & -b/|A| \\ -c/|A| & a/|A| \end{bmatrix} = \frac{1}{|A|}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

In other words, when $|A| \neq 0$, the inverse of a 2×2 matrix A may be obtained from A as follows:

(1) Interchange the two elements on the diagonal.

(2) Take the negatives of the other two elements.

(3) Multiply the resulting matrix by $1/|A|$ or, equivalently, divide each element by $|A|$.

In case $|A| = 0$, the matrix A is not invertible.

EXAMPLE 2.11 Find the inverse of $A = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$.

First evaluate $|A| = 2(5) - 3(4) = 10 - 12 = -2$. Because $|A| \neq 0$, the matrix A is invertible and

$$A^{-1} = \frac{1}{-2}\begin{bmatrix} 5 & -3 \\ -4 & 2 \end{bmatrix} = \begin{bmatrix} -\frac{5}{2} & \frac{3}{2} \\ 2 & -1 \end{bmatrix}$$

Now evaluate $|B| = 1(6) - 3(2) = 6 - 6 = 0$. Because $|B| = 0$, the matrix B has no inverse.

Remark: The above property that a matrix is invertible if and only if A has a nonzero determinant is true for square matrices of any order. (See Chapter 8.)

Inverse of an *n* × *n* Matrix

Suppose A is an arbitrary n-square matrix. Finding its inverse A^{-1} reduces, as above, to finding the solution of a collection of $n \times n$ systems of linear equations. The solution of such systems and an efficient way of solving such a collection of systems is treated in Chapter 3.

2.10 Special Types of Square Matrices

This section describes a number of special kinds of square matrices.

Diagonal and Triangular Matrices

A square matrix $D = [d_{ij}]$ is *diagonal* if its nondiagonal entries are all zero. Such a matrix is sometimes denoted by

$$D = \text{diag}(d_{11}, d_{22}, \ldots, d_{nn})$$

where some or all the d_{ii} may be zero. For example,

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

are diagonal matrices, which may be represented, respectively, by

$$\text{diag}(3, -7, 2), \qquad \text{diag}(4, -5), \qquad \text{diag}(6, 0, -9, 8)$$

(Observe that patterns of 0's in the third matrix have been omitted.)

A square matrix $A = [a_{ij}]$ is *upper triangular* or simply *triangular* if all entries below the (main) diagonal are equal to 0—that is, if $a_{ij} = 0$ for $i > j$. Generic upper triangular matrices of orders 2, 3, 4 are as follows:

$$\begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix}, \qquad \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ & b_{22} & b_{23} \\ & & b_{33} \end{bmatrix}, \qquad \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ & c_{22} & c_{23} & c_{24} \\ & & c_{33} & c_{34} \\ & & & c_{44} \end{bmatrix}$$

(As with diagonal matrices, it is common practice to omit patterns of 0's.)

The following theorem applies.

THEOREM 2.5: Suppose $A = [a_{ij}]$ and $B = [b_{ij}]$ are $n \times n$ (upper) triangular matrices. Then

(i) $A + B$, kA, AB are triangular with respective diagonals:

$$(a_{11} + b_{11}, \ \ldots, \ a_{nn} + b_{nn}), \qquad (ka_{11}, \ \ldots, \ ka_{nn}), \qquad (a_{11}b_{11}, \ \ldots, \ a_{nn}b_{nn})$$

(ii) For any polynomial $f(x)$, the matrix $f(A)$ is triangular with diagonal

$$(f(a_{11}), f(a_{22}), \ldots, f(a_{nn}))$$

(iii) A is invertible if and only if each diagonal element $a_{ii} \neq 0$, and when A^{-1} exists it is also triangular.

A *lower triangular matrix* is a square matrix whose entries above the diagonal are all zero. We note that Theorem 2.5 is true if we replace ''triangular'' by either ''lower triangular'' or ''diagonal.''

Remark: A nonempty collection A of matrices is called an *algebra* (of matrices) if A is closed under the operations of matrix addition, scalar multiplication, and matrix multiplication. Clearly, the square matrices with a given order form an algebra of matrices, but so do the scalar, diagonal, triangular, and lower triangular matrices.

Special Real Square Matrices: Symmetric, Orthogonal, Normal [Optional until Chapter 12]

Suppose now A is a square matrix with real entries—that is, a real square matrix. The relationship between A and its transpose A^T yields important kinds of matrices.

(a) Symmetric Matrices

A matrix A is *symmetric* if $A^T = A$. Equivalently, $A = [a_{ij}]$ is symmetric if *symmetric elements* (mirror elements with respect to the diagonal) are equal—that is, if each $a_{ij} = a_{ji}$.

A matrix A is *skew-symmetric* if $A^T = -A$ or, equivalently, if each $a_{ij} = -a_{ji}$. Clearly, the diagonal elements of such a matrix must be zero, because $a_{ii} = -a_{ii}$ implies $a_{ii} = 0$.

(Note that a matrix A must be square if $A^T = A$ or $A^T = -A$.)

EXAMPLE 2.12 Let $A = \begin{bmatrix} 2 & -3 & 5 \\ -3 & 6 & 7 \\ 5 & 7 & -8 \end{bmatrix}, B = \begin{bmatrix} 0 & 3 & -4 \\ -3 & 0 & 5 \\ 4 & -5 & 0 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

(a) By inspection, the symmetric elements in A are equal, or $A^T = A$. Thus, A is symmetric.

(b) The diagonal elements of B are 0 and symmetric elements are negatives of each other, or $B^T = -B$. Thus, B is skew-symmetric.

(c) Because C is not square, C is neither symmetric nor skew-symmetric.

(b) Orthogonal Matrices

A real matrix A is *orthogonal* if $A^T = A^{-1}$—that is, if $AA^T = A^T A = I$. Thus, A must necessarily be square and invertible.

EXAMPLE 2.13 Let $A = \begin{bmatrix} \frac{1}{9} & \frac{8}{9} & -\frac{4}{9} \\ \frac{4}{9} & -\frac{4}{9} & -\frac{7}{9} \\ \frac{8}{9} & \frac{1}{9} & \frac{4}{9} \end{bmatrix}$. Multiplying A by A^T yields I; that is, $AA^T = I$. This means $A^T A = I$, as well. Thus, $A^T = A^{-1}$; that is, A is orthogonal.

Now suppose A is a real orthogonal 3×3 matrix with rows

$$u_1 = (a_1, a_2, a_3), \qquad u_2 = (b_1, b_2, b_3), \qquad u_3 = (c_1, c_2, c_3)$$

Because A is orthogonal, we must have $AA^T = I$. Namely,

$$AA^T = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I$$

Multiplying A by A^T and setting each entry equal to the corresponding entry in I yields the following nine equations:

$$\begin{aligned} a_1^2 + a_2^2 + a_3^2 &= 1, & a_1b_1 + a_2b_2 + a_3b_3 &= 0, & a_1c_1 + a_2c_2 + a_3c_3 &= 0 \\ b_1a_1 + b_2a_2 + b_3a_3 &= 0, & b_1^2 + b_2^2 + b_3^2 &= 1, & b_1c_1 + b_2c_2 + b_3c_3 &= 0 \\ c_1a_1 + c_2a_2 + c_3a_3 &= 0, & c_1b_1 + c_2b_2 + c_3b_3 &= 0, & c_1^2 + c_2^2 + c_3^2 &= 1 \end{aligned}$$

Accordingly, $u_1 \cdot u_1 = 1$, $u_2 \cdot u_2 = 1$, $u_3 \cdot u_3 = 1$, and $u_i \cdot u_j = 0$ for $i \neq j$. Thus, the rows u_1, u_2, u_3 are unit vectors and are orthogonal to each other.

Generally speaking, vectors $u_1, u_2, \ldots, u_m$ in $\mathbf{R}^n$ are said to form an *orthonormal set* of vectors if the vectors are unit vectors and are orthogonal to each other; that is,

$$u_i \cdot u_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

In other words, $u_i \cdot u_j = \delta_{ij}$ where δ_{ij} is the Kronecker delta function.

We have shown that the condition $AA^T = I$ implies that the rows of A form an orthonormal set of vectors. The condition $A^T A = I$ similarly implies that the columns of A also form an orthonormal set of vectors. Furthermore, because each step is reversible, the converse is true.

The above results for 3×3 matrices are true in general. That is, the following theorem holds.

THEOREM 2.6: Let A be a real matrix. Then the following are equivalent:

(a) A is orthogonal.
(b) The rows of A form an orthonormal set.
(c) The columns of A form an orthonormal set.

For $n = 2$, we have the following result (proved in Problem 2.28).

THEOREM 2.7: Let A be a real 2×2 orthogonal matrix. Then, for some real number θ,

$$A = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \qquad \text{or} \qquad A = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$$

(c) Normal Matrices

A real matrix A is *normal* if it *commutes* with its transpose A^T—that is, if $AA^T = A^TA$. If A is symmetric, orthogonal, or skew-symmetric, then A is normal. There are also other normal matrices.

EXAMPLE 2.14 Let $A = \begin{bmatrix} 6 & -3 \\ 3 & 6 \end{bmatrix}$. Then

$$AA^T = \begin{bmatrix} 6 & -3 \\ 3 & 6 \end{bmatrix}\begin{bmatrix} 6 & 3 \\ -3 & 6 \end{bmatrix} = \begin{bmatrix} 45 & 0 \\ 0 & 45 \end{bmatrix} \qquad \text{and} \qquad A^TA = \begin{bmatrix} 6 & 3 \\ -3 & 6 \end{bmatrix}\begin{bmatrix} 6 & -3 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 45 & 0 \\ 0 & 45 \end{bmatrix}$$

Because $AA^T = A^TA$, the matrix A is normal.

2.11 Complex Matrices

Let A be a complex matrix—that is, a matrix with complex entries. Recall (Section 1.7) that if $z = a + bi$ is a complex number, then $\bar{z} = a - bi$ is its conjugate. The *conjugate* of a complex matrix A, written $\bar{A}$, is the matrix obtained from A by taking the conjugate of each entry in A. That is, if $A = [a_{ij}]$, then $\bar{A} = [b_{ij}]$, where $b_{ij} = \bar{a}_{ij}$. (We denote this fact by writing $\bar{A} = [\bar{a}_{ij}]$.)

The two operations of transpose and conjugation commute for any complex matrix A, and the special notation A^H is used for the conjugate transpose of A. That is,

$$A^H = (\bar{A})^T = (\overline{A^T})$$

Note that if A is real, then $A^H = A^T$. [Some texts use A^* instead of A^H.]

EXAMPLE 2.15 Let $A = \begin{bmatrix} 2+8i & 5-3i & 4-7i \\ 6i & 1-4i & 3+2i \end{bmatrix}$. Then $A^H = \begin{bmatrix} 2-8i & -6i \\ 5+3i & 1+4i \\ 4+7i & 3-2i \end{bmatrix}$.

Special Complex Matrices: Hermitian, Unitary, Normal [Optional until Chapter 12]

Consider a complex matrix A. The relationship between A and its conjugate transpose A^H yields important kinds of complex matrices (which are analogous to the kinds of real matrices described above).

A complex matrix A is said to be *Hermitian* or *skew-Hermitian* according as to whether

$$A^H = A \qquad \text{or} \qquad A^H = -A.$$

Clearly, $A = [a_{ij}]$ is Hermitian if and only if symmetric elements are conjugate—that is, if each $a_{ij} = \bar{a}_{ji}$—in which case each diagonal element a_{ii} must be real. Similarly, if A is skew-symmetric, then each diagonal element $a_{ii} = 0$. (Note that A must be square if $A^H = A$ or $A^H = -A$.)

A complex matrix A is *unitary* if $A^HA^{-1} = A^{-1}A^H = I$—that is, if

$$A^H = A^{-1}.$$

Thus, A must necessarily be square and invertible. We note that a complex matrix A is unitary if and only if its rows (columns) form an orthonormal set relative to the dot product of complex vectors.

A complex matrix A is said to be *normal* if it commutes with A^H—that is, if

$$AA^H = A^HA$$

(Thus, A must be a square matrix.) This definition reduces to that for real matrices when A is real.

EXAMPLE 2.16 Consider the following complex matrices:

$$A = \begin{bmatrix} 3 & 1-2i & 4+7i \\ 1+2i & -4 & -2i \\ 4-7i & 2i & 5 \end{bmatrix} \qquad B = \frac{1}{2}\begin{bmatrix} 1 & -i & -1+i \\ i & 1 & 1+i \\ 1+i & -1+i & 0 \end{bmatrix} \qquad C = \begin{bmatrix} 2+3i & 1 \\ i & 1+2i \end{bmatrix}$$

(a) By inspection, the diagonal elements of A are real, and the symmetric elements $1-2i$ and $1+2i$ are conjugate, $4+7i$ and $4-7i$ are conjugate, and $-2i$ and $2i$ are conjugate. Thus, A is Hermitian.

(b) Multiplying B by B^H yields I; that is, $BB^H = I$. This implies $B^H B = I$, as well. Thus, $B^H = B^{-1}$, which means B is unitary.

(c) To show C is normal, we evaluate CC^H and $C^H C$:

$$CC^H = \begin{bmatrix} 2+3i & 1 \\ i & 1+2i \end{bmatrix}\begin{bmatrix} 2-3i & -i \\ 1 & 1-2i \end{bmatrix} = \begin{bmatrix} 14 & 4-4i \\ 4+4i & 6 \end{bmatrix}$$

and similarly $C^H C = \begin{bmatrix} 14 & 4-4i \\ 4+4i & 6 \end{bmatrix}$. Because $CC^H = C^H C$, the complex matrix C is normal.

We note that when a matrix A is real, Hermitian is the same as symmetric, and unitary is the same as orthogonal.

2.12 Block Matrices

Using a system of horizontal and vertical (dashed) lines, we can partition a matrix A into submatrices called *blocks* (or *cells*) of A. Clearly a given matrix may be divided into blocks in different ways. For example,

$$\left[\begin{array}{cc:cc:c} 1 & -2 & 0 & 1 & 3 \\ 2 & 3 & 5 & 7 & -2 \\ \hdashline 3 & 1 & 4 & 5 & 9 \\ 4 & 6 & -3 & 1 & 8 \end{array}\right], \qquad \left[\begin{array}{cc:ccc} 1 & -2 & 0 & 1 & 3 \\ \hdashline 2 & 3 & 5 & 7 & -2 \\ 3 & 1 & 4 & 5 & 9 \\ \hdashline 4 & 6 & -3 & 1 & 8 \end{array}\right], \qquad \left[\begin{array}{ccc:cc} 1 & -2 & 0 & 1 & 3 \\ 2 & 3 & 5 & 7 & -2 \\ \hdashline 3 & 1 & 4 & 5 & 9 \\ 4 & 6 & -3 & 1 & 8 \end{array}\right]$$

The convenience of the partition of matrices, say A and B, into blocks is that the result of operations on A and B can be obtained by carrying out the computation with the blocks, just as if they were the actual elements of the matrices. This is illustrated below, where the notation $A = [A_{ij}]$ will be used for a block matrix A with blocks A_{ij}.

Suppose that $A = [A_{ij}]$ and $B = [B_{ij}]$ are block matrices with the same numbers of row and column blocks, and suppose that corresponding blocks have the same size. Then adding the corresponding blocks of A and B also adds the corresponding elements of A and B, and multiplying each block of A by a scalar k multiplies each element of A by k. Thus,

$$A + B = \begin{bmatrix} A_{11}+B_{11} & A_{12}+B_{12} & \ldots & A_{1n}+B_{1n} \\ A_{21}+B_{21} & A_{22}+B_{22} & \ldots & A_{2n}+B_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ A_{m1}+B_{m1} & A_{m2}+B_{m2} & \ldots & A_{mn}+B_{mn} \end{bmatrix}$$

and

$$kA = \begin{bmatrix} kA_{11} & kA_{12} & \ldots & kA_{1n} \\ kA_{21} & kA_{22} & \ldots & kA_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ kA_{m1} & kA_{m2} & \ldots & kA_{mn} \end{bmatrix}$$

The case of matrix multiplication is less obvious, but still true. That is, suppose that $U = [U_{ik}]$ and $V = [V_{kj}]$ are block matrices such that the number of columns of each block U_{ik} is equal to the number of rows of each block V_{kj}. (Thus, each product $U_{ik}V_{kj}$ is defined.) Then

$$UV = \begin{bmatrix} W_{11} & W_{12} & \dots & W_{1n} \\ W_{21} & W_{22} & \dots & W_{2n} \\ \dots & \dots & \dots & \dots \\ W_{m1} & W_{m2} & \dots & W_{mn} \end{bmatrix}, \qquad \text{where} \quad W_{ij} = U_{i1}V_{1j} + U_{i2}V_{2j} + \dots + U_{ip}V_{pj}$$

The proof of the above formula for UV is straightforward but detailed and lengthy. It is left as an exercise (Problem 2.85).

Square Block Matrices

Let M be a block matrix. Then M is called a *square block matrix* if

(i) M is a square matrix.
(ii) The blocks form a square matrix.
(iii) The diagonal blocks are also square matrices.

The latter two conditions will occur if and only if there are the same number of horizontal and vertical lines and they are placed symmetrically.

Consider the following two block matrices:

$$A = \left[\begin{array}{cc|cc|c} 1 & 2 & 3 & 4 & 5 \\ 1 & 1 & 1 & 1 & 1 \\ \hline 9 & 8 & 7 & 6 & 5 \\ \hline 4 & 4 & 4 & 4 & 4 \\ 3 & 5 & 3 & 5 & 3 \end{array}\right] \qquad \text{and} \qquad B = \left[\begin{array}{cc|cc|c} 1 & 2 & 3 & 4 & 5 \\ 1 & 1 & 1 & 1 & 1 \\ \hline 9 & 8 & 7 & 6 & 5 \\ 4 & 4 & 4 & 4 & 4 \\ \hline 3 & 5 & 3 & 5 & 3 \end{array}\right]$$

The block matrix A is not a square block matrix, because the second and third diagonal blocks are not square. On the other hand, the block matrix B is a square block matrix.

Block Diagonal Matrices

Let $M = [A_{ij}]$ be a square block matrix such that the nondiagonal blocks are all zero matrices; that is, $A_{ij} = 0$ when $i \neq j$. Then M is called a *block diagonal matrix*. We sometimes denote such a block diagonal matrix by writing

$$M = \text{diag}(A_{11}, A_{22}, \dots, A_{rr}) \qquad \text{or} \qquad M = A_{11} \oplus A_{22} \oplus \dots \oplus A_{rr}$$

The importance of block diagonal matrices is that the algebra of the block matrix is frequently reduced to the algebra of the individual blocks. Specifically, suppose $f(x)$ is a polynomial and M is the above block diagonal matrix. Then $f(M)$ is a block diagonal matrix, and

$$f(M) = \text{diag}(f(A_{11}), f(A_{22}), \dots, f(A_{rr}))$$

Also, M is invertible if and only if each A_{ii} is invertible, and, in such a case, M^{-1} is a block diagonal matrix, and

$$M^{-1} = \text{diag}(A_{11}^{-1}, A_{22}^{-1}, \dots, A_{rr}^{-1})$$

Analogously, a square block matrix is called a *block upper triangular matrix* if the blocks below the diagonal are zero matrices and a *block lower triangular matrix* if the blocks above the diagonal are zero matrices.

EXAMPLE 2.17 Determine which of the following square block matrices are upper diagonal, lower diagonal, or diagonal:

$$A = \left[\begin{array}{cc|c} 1 & 2 & 0 \\ 3 & 4 & 5 \\ \hline 0 & 0 & 6 \end{array}\right], \qquad B = \left[\begin{array}{c|cc|c} 1 & 0 & 0 & 0 \\ \hline 2 & 3 & 4 & 0 \\ 5 & 0 & 6 & 0 \\ \hline 0 & 7 & 8 & 9 \end{array}\right], \qquad C = \left[\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & 2 & 3 \\ 0 & 4 & 5 \end{array}\right], \qquad D = \left[\begin{array}{cc|c} 1 & 2 & 0 \\ 3 & 4 & 5 \\ \hline 0 & 6 & 7 \end{array}\right]$$

(a) A is upper triangular because the block below the diagonal is a zero block.

(b) B is lower triangular because all blocks above the diagonal are zero blocks.

(c) C is diagonal because the blocks above and below the diagonal are zero blocks.

(d) D is neither upper triangular nor lower triangular. Also, no other partitioning of D will make it into either a block upper triangular matrix or a block lower triangular matrix.

SOLVED PROBLEMS

Matrix Addition and Scalar Multiplication

2.1 Given $A = \begin{bmatrix} 1 & -2 & 3 \\ 4 & 5 & -6 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & 0 & 2 \\ -7 & 1 & 8 \end{bmatrix}$, find:

(a) $A + B$, (b) $2A - 3B$.

(a) Add the corresponding elements:

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

(b) First perform the scalar multiplication and then a matrix addition:

$$2A - 3B = \begin{bmatrix} 2 & -4 & 6 \\ 8 & 10 & -12 \end{bmatrix} + \begin{bmatrix} -9 & 0 & -6 \\ 21 & -3 & -24 \end{bmatrix} = \begin{bmatrix} -7 & -4 & 0 \\ 29 & 7 & -36 \end{bmatrix}$$

(Note that we multiply B by -3 and then add, rather than multiplying B by 3 and subtracting. This usually prevents errors.)

2.2. Find x, y, z, t where $3\begin{bmatrix} x & y \\ z & t \end{bmatrix} = \begin{bmatrix} x & 6 \\ -1 & 2t \end{bmatrix} + \begin{bmatrix} 4 & x+y \\ z+t & 3 \end{bmatrix}$.

Write each side as a single equation:

$$\begin{bmatrix} 3x & 3y \\ 3z & 3t \end{bmatrix} = \begin{bmatrix} x+4 & x+y+6 \\ z+t-1 & 2t+3 \end{bmatrix}$$

Set corresponding entries equal to each other to obtain the following system of four equations:

$$3x = x+4, \qquad 3y = x+y+6, \qquad 3z = z+t-1, \qquad 3t = 2t+3$$

or $\quad 2x = 4, \qquad 2y = 6+x, \qquad 2z = t-1, \qquad t = 3$

The solution is $x = 2$, $y = 4$, $z = 1$, $t = 3$.

2.3. Prove Theorem 2.1 (i) and (v): (i) $(A+B)+C = A+(B+C)$, (v) $k(A+B) = kA+kB$.

Suppose $A = [a_{ij}]$, $B = [b_{ij}]$, $C = [c_{ij}]$. The proof reduces to showing that corresponding ij-entries in each side of each matrix equation are equal. [We prove only (i) and (v), because the other parts of Theorem 2.1 are proved similarly.]