CHAPTER 2

MATRIX ALGEBRA

In the previous chapter we introduced matrices as a shorthand device for representing systems of linear equations. Now we will see that matrices have a life of their own, first algebraically and then geometrically. The crucial new ingredient is to interpret an $m \times n$ matrix as a special sort of function that assigns to each vector $\mathbf{x} \in \mathbb{R}^n$ the product $A\mathbf{x} \in \mathbb{R}^m$.

1 Matrix Operations

Recall that an $m \times n$ matrix A is a rectangular array of mn real numbers,

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix},$$

where a_{ij} represents the *entry* in the i^{th} row and j^{th} column. We recall that two $m \times n$ matrices A and B are equal if $a_{ij} = b_{ij}$ for all $i = 1, \dots, m$ and $j = 1, \dots, n$.

We take this opportunity to warn our readers that the word *if* is ordinarily used in mathematical definitions, even though it should be the phrase *if and only if*. That is, even though we don't say so, we intend it to be understood that, for example, in this case, if $A = B$, then $a_{ij} = b_{ij}$ for all i and j. Be warned: This custom applies only to definitions, not to propositions and theorems! See the earlier discussions of *if and only if* on p. 21.

A has m row vectors,

$$\begin{aligned} \mathbf{A}_1 &= (a_{11}, \dots, a_{1n}), \\ \mathbf{A}_2 &= (a_{21}, \dots, a_{2n}), \\ &\vdots \\ \mathbf{A}_m &= (a_{m1}, \dots, a_{mn}), \end{aligned}$$

which are vectors in $\mathbb{R}^n$, and n column vectors,

$$\mathbf{a}_1 = \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} a_{12} \\ \vdots \\ a_{m2} \end{bmatrix}, \quad \dots, \quad \mathbf{a}_n = \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix},$$

which are, correspondingly, vectors in $\mathbb{R}^m$.

We denote by O the *zero matrix*, the $m \times n$ matrix all of whose entries are 0. We also introduce the notation $\mathcal{M}_{m\times n}$ for the set of all $m \times n$ matrices. For future reference, we call a matrix *square* if $m = n$ (i.e., it has equal numbers of rows and columns). In the case of a square matrix, we refer to the entries a_{ii}, $i = 1, \dots, n$, as *diagonal* entries.

Definition. Let A be an $n \times n$ (square) matrix with entries a_{ij} for $i = 1, \dots, n$ and $j = 1, \dots, n$.

1. We call A *diagonal* if every nondiagonal entry is zero, i.e., if $a_{ij} = 0$ whenever $i \neq j$.
2. We call A *upper triangular* if all of the entries below the diagonal are zero, i.e., if $a_{ij} = 0$ whenever $i > j$.
3. We call A *lower triangular* if all of the entries above the diagonal are zero, i.e., if $a_{ij} = 0$ whenever $i < j$.

Let's now consider various algebraic operations we can perform on matrices. Given an $m \times n$ matrix A, the simplest algebraic manipulation is to multiply every entry of A by a real number c (*scalar multiplication*). If A is the matrix with entries a_{ij} ($i = 1, \dots, m$ and $j = 1, \dots, n$), then cA is the matrix whose entries are ca_{ij}:

$$cA = c\begin{bmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} ca_{11} & \dots & ca_{1n} \\ ca_{21} & \dots & ca_{2n} \\ \vdots & \ddots & \vdots \\ ca_{m1} & \dots & ca_{mn} \end{bmatrix}.$$

Next comes *addition of matrices*. Given $m \times n$ matrices A and B, we define their sum entry by entry. In symbols, when

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} b_{11} & \dots & b_{1n} \\ b_{21} & \dots & b_{2n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \dots & b_{mn} \end{bmatrix},$$

we define

$$A + B = \begin{bmatrix} a_{11}+b_{11} & \dots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & \dots & a_{2n}+b_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1}+b_{m1} & \dots & a_{mn}+b_{mn} \end{bmatrix}.$$

It is important to understand that when we refer to the set of all $m \times n$ matrices, $\mathcal{M}_{m\times n}$, we have not specified the positive integers m and n. They can be chosen arbitrarily. However, when we say that $A, B \in \mathcal{M}_{m\times n}$, we mean that A and B must have the same "shape," i.e., the same number of rows (m) and the same number of columns (n).

EXAMPLE 1

Let $c = -2$ and

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & -2 \\ 4 & -1 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 6 & 4 & -1 \\ -3 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix}.$$

Then

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

and neither sum $A + C$ nor $B + C$ makes sense, because C has a different shape from A and B. ▲

We leave it to the reader to check that scalar multiplication of matrices and matrix addition satisfy the same list of properties we gave in Exercise 1.1.28 for scalar multiplication of vectors and vector addition. We list them here for reference.

Proposition 1.1. *Let $A, B, C \in \mathcal{M}_{m\times n}$ and let $c, d \in \mathbb{R}$.*

1. $A + B = B + A$.
2. $(A + B) + C = A + (B + C)$.
3. $\mathrm{O} + A = A$.
4. *There is a matrix $-A$ so that $A + (-A) = \mathrm{O}$.*
5. $c(dA) = (cd)A$.
6. $c(A + B) = cA + cB$.
7. $(c + d)A = cA + dA$.
8. $1A = A$.

Proof. Left to the reader in Exercise 3. □

To understand these properties, one might simply examine corresponding entries of the appropriate matrices and use the relevant properties of real numbers to see why they are equal. A more elegant approach is the following: We can encode an $m \times n$ matrix as a vector in $\mathbb{R}^{mn}$, for example,

$$\begin{bmatrix} 1 & -1 \\ 2 & 3 \\ -5 & 4 \end{bmatrix} \in \mathcal{M}_{3\times 2} \rightsquigarrow (1, -1, 2, 3, -5, 4) \in \mathbb{R}^6,$$

and you can check that scalar multiplication and addition of matrices correspond exactly to scalar multiplication and addition of vectors. We will make this concept more precise in Section 6 of Chapter 3.

The real power of matrices comes from the operation of matrix multiplication. Just as we can compute a dot product of two vectors in $\mathbb{R}^n$, ending up with a scalar, we shall see that we can multiply matrices of appropriate shapes:

$$\mathcal{M}_{m\times n} \times \mathcal{M}_{n\times p} \to \mathcal{M}_{m\times p}.$$

In particular, when $m = n = p$ (so that our matrices are square and of the same size), we have a way of combining two $n \times n$ matrices to obtain another $n \times n$ matrix.

Definition. Let A be an $m \times n$ matrix and B an $n \times p$ matrix. Their product AB is an $m \times p$ matrix whose ij-entry is

$$(AB)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj};$$

that is, the dot product of the i^{th} row vector of A and the j^{th} column vector of B, both of which are vectors in $\mathbb{R}^n$. Graphically, we have

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & \vdots & & \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ & \vdots & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{11} & & b_{1j} & & b_{1p} \\ b_{21} & \cdots & b_{2j} & \cdots & b_{2p} \\ \vdots & & \vdots & & \vdots \\ b_{n1} & & b_{nj} & & b_{np} \end{bmatrix} = \begin{bmatrix} \cdots & \cdots & \cdots & \cdots & \cdots \\ & & \vdots & & \\ \cdots & \cdots & (AB)_{ij} & \cdots & \cdots \\ & & \vdots & & \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}.$$

We reiterate that in order for the product AB to be defined, the number of *columns* of A must equal the number of *rows* of B.

Recall that in Section 4 of Chapter 1 we defined the product of an $m \times n$ matrix A with a vector $\mathbf{x} \in \mathbb{R}^n$. The definition we just gave generalizes that if we think of an $n \times p$ matrix B as a collection of p column vectors. In particular,

$\boxed{\text{The } j^{\text{th}} \text{ column of } AB \text{ is the product of } A \text{ with the } j^{\text{th}} \text{ column vector of } B.}$

EXAMPLE 2

Note that this definition is compatible with our definition in Chapter 1 of the multiplication of an $m \times n$ matrix with a column vector in $\mathbb{R}^n$ (an $n \times 1$ matrix). For example, if

$$A = \begin{bmatrix} 1 & 3 \\ 2 & -1 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} 4 & 1 & 0 & -2 \\ -1 & 1 & 5 & 1 \end{bmatrix},$$

then

$$A\mathbf{x} = \begin{bmatrix} 1 & 3 \\ 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 9 \\ 3 \end{bmatrix}, \quad \text{and}$$

$$AB = \begin{bmatrix} 1 & 3 \\ 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 4 & 10 & -2 \\ -1 & 15 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 4 & 15 & 1 \\ 9 & 1 & -5 & -5 \\ 3 & 2 & 5 & -1 \end{bmatrix}.$$

Notice also that the product BA does not make sense: B is a 2×4 matrix and A is 3×2, and $4 \neq 3$. ▲

The preceding example brings out an important point about the nature of matrix multiplication: It can happen that the matrix product AB is defined and the product BA is not. Now if A is an $m \times n$ matrix and B is an $n \times m$ matrix, then both products AB and BA make sense: AB is $m \times m$ and BA is $n \times n$. Notice that these are both square matrices, but of different sizes.

EXAMPLE 3

To see an extreme example of this, consider the 1×3 matrix $A = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$ and the 3×1 matrix $B = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$. Then

$$AB = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \end{bmatrix}, \quad \text{whereas}$$

$$BA = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ -1 & -2 & -3 \\ 2 & 4 & 6 \end{bmatrix}.$$

▲

Even if we start with both A and B as $n \times n$ matrices, the products AB and BA have the same shape but need not be equal.

EXAMPLE 4

Let

$$A = \begin{bmatrix} 1 & 2 \\ -3 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix}.$$

Then

$$AB = \begin{bmatrix} 1 & 0 \\ 4 & 0 \end{bmatrix}, \quad \text{whereas} \quad BA = \begin{bmatrix} -1 & -2 \\ 1 & 2 \end{bmatrix}.$$

▲

When—and only when—A is a square matrix, we can multiply A by itself, obtaining $A^2 = AA$, $A^3 = A^2A = AA^2$, etc. In the last examples of Chapter 1, Section 6, the vectors $\mathbf{x}_k$ are obtained from the initial vector $\mathbf{x}_0$ by repeatedly multiplying by the matrix A, so that $\mathbf{x}_k = A^k\mathbf{x}_0$.

EXAMPLE 5

There is an interesting way to interpret matrix powers in terms of directed graphs. Starting with the matrix

$$A = \begin{bmatrix} 0 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix},$$

we draw a graph with 3 nodes (vertices) and a_{ij} directed edges (paths) from node i to node j, as shown in Figure 1.1. For example, there are 2 edges from node 1 to node 2 and none from node 3 to node 2. If we multiply a_{ij} by a_{jk}, we get the number of two-step paths from node i to node k passing through node j. Thus, in this case, the sum

$$a_{i1}a_{1k} + a_{i2}a_{2k} + a_{i3}a_{3k}$$

gives all the two-step paths from node i to node k. For example, the 13-entry of A^2,

$$(A^2)_{13} = a_{11}a_{13} + a_{12}a_{23} + a_{13}a_{33} = (0)(1) + (2)(1) + (1)(1) = 3,$$

gives the number of two-step paths from node 1 to node 3. With a bit of thought, the reader will convince herself that the ij-entry of A^n is the number of n-step directed paths from node i to node j.

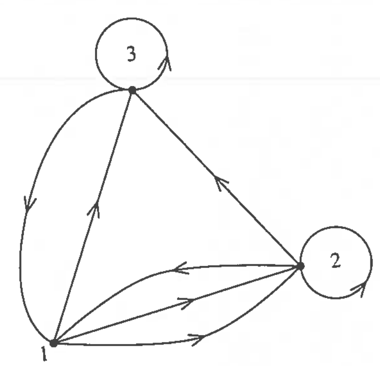

FIGURE 1.1

We calculate

$$A^2 = \begin{bmatrix} 0 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 2 & 3 \\ 2 & 3 & 3 \\ 1 & 2 & 2 \end{bmatrix},$$

$$A^3 = \begin{bmatrix} 5 & 8 & 8 \\ 6 & 7 & 8 \\ 4 & 4 & 5 \end{bmatrix}, \quad \text{and} \quad \dots$$

$$A^7 = \begin{bmatrix} 272 & 338 & 377 \\ 273 & 337 & 377 \\ 169 & 208 & 233 \end{bmatrix}.$$

In particular, there are 169 seven-step paths from node 3 to node 1. ▲

We have seen that, in general, matrix multiplication is *not* commutative. However, it does have the following crucial properties. Let I_n denote the $n \times n$ matrix with 1's on the diagonal and 0's elsewhere, as illustrated on p. 61.

Proposition 1.2. *Let A and A' be $m \times n$ matrices, let B and B' be $n \times p$ matrices, let C be a $p \times q$ matrix, and let c be a scalar. Then*

1. $AI_n = A = I_mA$. *For this reason, I_n is called the $n \times n$* identity matrix.
2. $(A + A')B = AB + A'B$ *and* $A(B + B') = AB + AB'$. *This is the* distributive *property of matrix multiplication over matrix addition.*
3. $(cA)B = c(AB) = A(cB)$.
4. $(AB)C = A(BC)$. *This is the* associative *property of matrix multiplication.*

Proof. We prove the associative property and leave the rest to the reader in Exercise 4. Note first of all that there is hope: AB is an $m \times p$ matrix and C is a $p \times q$ matrix, so $(AB)C$ will be an $m \times q$ matrix; similarly, A is an $m \times n$ matrix and BC is a $n \times q$ matrix, so $A(BC)$ will be an $m \times q$ matrix. Associativity amounts to the statement that

$$(AB)\mathbf{c} = A(B\mathbf{c})$$

for any column vector $\mathbf{c}$ of the matrix C: To calculate the j^{th} column of $(AB)C$ we multiply AB by the j^{th} column of C; to calculate the j^{th} column of $A(BC)$ we multiply A by the j^{th} column of BC, which, in turn, is the product of B with the j^{th} column of C.

Letting $\mathbf{b}_1, \dots, \mathbf{b}_p$ denote the column vectors of B, we recall (see the crucial observation (∗) on p. 53) that $B\mathbf{c}$ is the linear combination $c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_p\mathbf{b}_p$, and so (using Proposition 5.2 of Chapter 1)

$$\begin{aligned} A(B\mathbf{c}) &= A(c_1\mathbf{b}_1 + c_2\mathbf{b}_2 + \cdots + c_p\mathbf{b}_p) = c_1(A\mathbf{b}_1) + c_2(A\mathbf{b}_2) + \cdots + c_p(A\mathbf{b}_p) \\ &= c_1(\text{first column of } AB) + c_2(\text{second column of } AB) \\ &\quad + \cdots + c_p(p^{\text{th}} \text{ column of } AB) \\ &= (AB)\mathbf{c}. \end{aligned}$$

□

There is an important conceptual point underlying this computation, as we now study. Through Chapter 1, we thought of matrices simply as an algebraic shorthand for dealing with systems of linear equations. However, we can interpret matrices as functions, hence imparting to them a geometric interpretation and explaining the meaning of matrix multiplication.

Multiplying the $m \times n$ matrix A by vectors $\mathbf{x} \in \mathbb{R}^n$ defines a function

$$\mu_A : \mathbb{R}^n \to \mathbb{R}^m, \quad \text{given by} \quad \mu_A(\mathbf{x}) = A\mathbf{x}.$$

The function μ_A has *domain* $\mathbb{R}^n$ and *range* $\mathbb{R}^m$, and we often say that "μ_A maps $\mathbb{R}^n$ to $\mathbb{R}^m$."

A function $f : X \to Y$ is a "rule" that assigns to each element x of the domain X an element $f(x)$ of the range Y. We refer to $f(x)$ as the *value* of the function at x. We can think of a function as a machine that turns raw ingredients (inputs) into products (outputs), depicted by a diagram such as on the left in Figure 1.2. In high school mathematics and calculus classes, we tend to visualize a function f by means of its graph, the set of ordered pairs (x, y) with $y = f(x)$. The graph must pass the "vertical line test": For each $x = x_0$ in X, there must be exactly one point (x_0, y) among the ordered pairs.

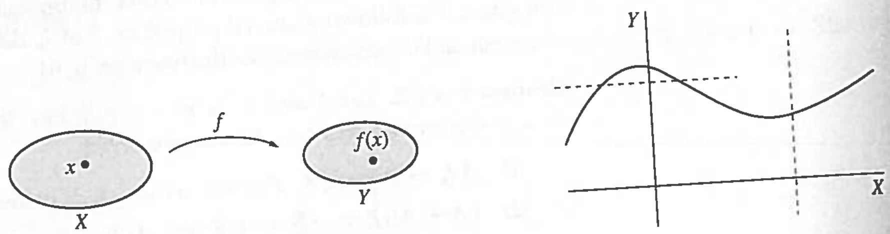

FIGURE 1.2

We say the function is one-to-one if the graph passes the "horizontal line test": For each $y = y_0 \in Y$, there is *at most* one point (x, y_0) among the ordered pairs. The function whose graph is pictured on the right in Figure 1.2 is not one-to-one. More formally, $f : X \to Y$ is *one-to-one* (or injective) if, for $a, b \in X$, the only way we can have $f(a) = f(b)$ is with $a = b$.

Another term that appears frequently is this: We say f is *onto* (or surjective) if every $y \in Y$ is of the form $y = f(x)$ for (at least one) $x \in X$. That is to say, f is onto if the set of all its values (often called the *image* of f) is all of Y. When we were considering linear equations $A\mathbf{x} = \mathbf{b}$ in Chapter 1, we found constraint equations that $\mathbf{b}$ must satisfy in order for the equation to be consistent. Vectors $\mathbf{b}$ satisfying those constraint equations are in the image of μ_A. The mapping μ_A is onto precisely when there are no constraint equations for consistency.

Last, a function $f : X \to Y$ that is both one-to-one and onto is often called a one-to-one correspondence between X and Y (or a bijection). We saw in Section 5 of Chapter 1 that $\mu_A : \mathbb{R}^n \to \mathbb{R}^n$ is one-to-one and onto precisely when A is nonsingular.

As we just saw in proving associativity of matrix multiplication, for an $m \times n$ matrix A and an $n \times p$ matrix B,

$$(AB)\mathbf{c} = A(B\mathbf{c})$$

for every vector $\mathbf{c} \in \mathbb{R}^p$. We can now rewrite this as

$$\mu_{AB}(\mathbf{c}) = \mu_A(\mu_B(\mathbf{c})) = (\mu_A \circ \mu_B)(\mathbf{c}),$$

where the latter notation denotes composition of functions. Of course, this formula is the real motivation for defining matrix multiplication as we did. In fact, one might define the matrix product as a composition of functions and then derive the computational formula. Now, we know that *composition of functions is associative* (even though it is not commutative):

$$(f \circ g) \circ h = f \circ (g \circ h),$$

from which we infer that

$$(\mu_A \circ \mu_B) \circ \mu_C = \mu_A \circ (\mu_B \circ \mu_C), \quad \text{and so}$$
$$\mu_{(AB)C} = \mu_{A(BC)}; \quad \text{that is,}$$
$$(AB)C = A(BC).$$

This is how one should understand matrix multiplication and its associativity.

Remark. Mathematicians will often express the rule $\mu_{AB} = \mu_A \circ \mu_B$ schematically by the following diagram:

$$\begin{tikzcd} \mathbb{R}^p \arrow[r, "\mu_B"] \arrow[rr, bend right, "\mu_{AB}"'] & \mathbb{R}^n \arrow[r, "\mu_A"] & \mathbb{R}^m \end{tikzcd}$$

We will continue to explore the interpretation of matrices as functions in the next section.

Exercises 2.1

1. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $B = \begin{bmatrix} 2 & 1 \\ 4 & 3 \end{bmatrix}$, $C = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$, and $D = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 2 & 3 \end{bmatrix}$. Calculate each of the following expressions or explain why it is not defined.

a. $A + B$	d. $C + D$	*g. AC	j. DB
*b. $2A - B$	e. AB	*h. CA	*k. CD
c. $A - C$	*f. BA	i. BD	*l. DC

2. Let $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} -1 & 3 \\ 1 & -3 \end{bmatrix}$. Show that $AB = O$ but $BA \neq O$. Explain this result geometrically.
3. Prove Proposition 1.1. While you're at it, prove (using these properties) that for any $A \in \mathcal{M}_{m \times n}$, $0A = O$.
4. a. Prove the remainder of Proposition 1.2.
   b. Interpret parts **1**, **2**, and **3** of Proposition 1.2 in terms of properties of functions.
   c. Suppose Charlie has carefully proved the first statement in part **2** and offers the following justification of the second: Since $(B + B')A = BA + B'A$, we now have $A(B + B') = (B + B')A = BA + B'A = AB + AB' = A(B + B')$. Decide whether he is correct.
5. *a. If A is an $m \times n$ matrix and $A\mathbf{x} = \mathbf{0}$ for all $\mathbf{x} \in \mathbb{R}^n$, show that $A = O$.
   b. If A and B are $m \times n$ matrices and $A\mathbf{x} = B\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$, show that $A = B$.
6. Prove or give a counterexample. Assume all the matrices are $n \times n$.
   a. If $AB = CB$ and $B \neq O$, then $A = C$.
   b. If $A^2 = A$, then $A = O$ or $A = I$.

c. $(A+B)(A-B) = A^2 - B^2$.
d. If $AB = CB$ and B is nonsingular, then $A = C$.
e. If $AB = BC$ and B is nonsingular, then $A = C$.

In the box on p. 52, we suggested that in such a problem you might try $n = 1$ to get intuition. Well, if we have real numbers a, b, and c satisfying $ab = cb$, then $ab - cb = (a-c)b = 0$, so $b = 0$ or $a = c$. Similarly, if $a^2 = a$, then $a^2 - a = a(a-1) = 0$, so $a = 0$ or $a = 1$, and so on. So, once again, it's not clear that the case $n = 1$ gives much insight into the general case. But it might lead us to the right question: Is it true for $n \times n$ matrices that $AB = \mathrm{O}$ implies $A = \mathrm{O}$ or $B = \mathrm{O}$?

To answer this question, you might either play around with numerical examples (e.g., with 2×2 matrices) or interpret this matrix product geometrically: What does it say about the relation between the rows of A and the columns of B?

7. Find all 2×2 matrices $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ satisfying

a. $A^2 = I_2$ *b. $A^2 = \mathrm{O}$ c. $A^2 = -I_2$

8. For each of the following matrices A, find a formula for A^k for positive integers k. (If you know how to do proof by induction, please do.)

a. $A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$ b. $A = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix}$ c. $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$

9. (Block multiplication) We can think of an $(m+n) \times (m+n)$ matrix as being decomposed into "blocks," and thinking of these blocks as matrices themselves, we can form products and sums appropriately. Suppose A and A' are $m \times m$ matrices, B and B' are $m \times n$ matrices, C and C' are $n \times m$ matrices, and D and D' are $n \times n$ matrices. Verify the following formula for the product of "block" matrices:

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right]\left[\begin{array}{c|c} A' & B' \\ \hline C' & D' \end{array}\right] = \left[\begin{array}{c|c} AA' + BC' & AB' + BD' \\ \hline CA' + DC' & CB' + DD' \end{array}\right].$$

10. Suppose A and B are nonsingular $n \times n$ matrices. Prove that AB is nonsingular.

Although it is tempting to try to show that the reduced echelon form of AB is the identity matrix, there is no direct way to do this. As is the case in most non-numerical problems regarding nonsingularity, you should remember that AB is nonsingular precisely when the only solution of $(AB)\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$.

11. ♯a. Suppose $A \in \mathcal{M}_{m\times n}$, $B \in \mathcal{M}_{n\times m}$, and $BA = I_n$. Prove that if for some $\mathbf{b} \in \mathbb{R}^m$ the equation $A\mathbf{x} = \mathbf{b}$ has a solution, then that solution is unique.

b. Suppose $A \in \mathcal{M}_{m\times n}$, $C \in \mathcal{M}_{n\times m}$, and $AC = I_m$. Prove that the system $A\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b} \in \mathbb{R}^m$.

To show that *if* a solution exists, then it is unique, one approach (which works well here) is to *suppose* that $\mathbf{x}$ satisfies the equation and find a formula that determines it. Another approach is to assume that $\mathbf{x}$ and $\mathbf{y}$ are both solutions and then use the equations to prove that $\mathbf{x} = \mathbf{y}$.

To prove that a solution exists, the direct approach (which works here) is to find *some* $\mathbf{x}$ that works—even if that means guessing. A more subtle approach to existence questions involves proof by contradiction (see the box on p. 18): Assume there is no solution, and deduce from this assumption something that is known to be false.

♯c. Suppose $A \in \mathcal{M}_{m\times n}$ and $B, C \in \mathcal{M}_{n\times m}$ are matrices that satisfy $BA = I_n$ and $AC = I_m$. Prove that $B = C$.

12. An $n \times n$ matrix is called a *permutation matrix* if it has a single 1 in each row and column and all its remaining entries are 0.

a. Write down all the 2×2 permutation matrices. How many are there?
b. Write down all the 3×3 permutation matrices. How many are there?
c. Show that the product of two permutation matrices is again a permutation matrix. Do they commute?
d. Prove that every permutation matrix is nonsingular.
e. If A is an $n \times n$ matrix and P is an $n \times n$ permutation matrix, describe the columns of AP and the rows of PA.

13. Find matrices A so that

a. $A \neq \mathrm{O}$, but $A^2 = \mathrm{O}$
b. $A^2 \neq \mathrm{O}$, but $A^3 = \mathrm{O}$

Can you make a conjecture about matrices satisfying $A^{n-1} \neq \mathrm{O}$ but $A^n = \mathrm{O}$?

14. Find all 2×2 matrices A that commute with all 2×2 matrices B. That is, if $AB = BA$ for all $B \in \mathcal{M}_{2\times 2}$, what are the possible matrices that A can be?

15. (The binomial theorem for matrices) Suppose A and B are $n \times n$ matrices with the property that $AB = BA$. Prove that for any positive integer k, we have

$$\begin{aligned}(A+B)^k &= \sum_{i=0}^{k} \frac{k!}{i!(k-i)!} A^{k-i}B^i \\ &= A^k + kA^{k-1}B + \frac{k(k-1)}{2}A^{k-2}B^2 + \frac{k(k-1)(k-2)}{6}A^{k-3}B^3 \\ &\quad + \cdots + kAB^{k-1} + B^k.\end{aligned}$$

Show that the result is false when $AB \neq BA$.

2 Linear Transformations: An Introduction

The function μ_A we defined at the end of Section 1 is a prototype of the functions one studies in linear algebra, called *linear transformations*. We shall explore them in greater detail in Chapter 4, but here we want to familiarize ourselves with a number of examples. First, a definition:

Definition. A function $T: \mathbb{R}^n \to \mathbb{R}^m$ is called a *linear transformation* (or *linear map*) if it satisfies

(i) $T(\mathbf{x}+\mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

(ii) $T(c\mathbf{x}) = cT(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^n$ and all scalars c.

These are often called the *linearity properties*.

EXAMPLE 1

Here are a few examples of functions, some linear, some not.

(a) Consider the function $T: \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 \\ x_1 \end{bmatrix}$. Let's decide whether it satisfies the two properties of a linear map.

(i)

$$\begin{aligned} T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) &= T\left(\begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix}\right) \\ &= \begin{bmatrix} (x_1+y_1)+(x_2+y_2) \\ (x_1+y_1) \end{bmatrix} = \begin{bmatrix} (x_1+x_2)+(y_1+y_2) \\ x_1+y_1 \end{bmatrix} \\ &= \begin{bmatrix} x_1+x_2 \\ x_1 \end{bmatrix} + \begin{bmatrix} y_1+y_2 \\ y_1 \end{bmatrix} = T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) + T\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) \end{aligned}$$

(ii)

$$\begin{aligned} T\left(c\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) &= T\left(\begin{bmatrix} cx_1 \\ cx_2 \end{bmatrix}\right) = \begin{bmatrix} cx_1 + cx_2 \\ cx_1 \end{bmatrix} \\ &= c\begin{bmatrix} x_1 + x_2 \\ x_1 \end{bmatrix} = cT\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) \quad \text{for all scalars } c. \end{aligned}$$

Thus, T is a linear map.

It is important to remember that we have to check that the equation $T(\mathbf{x}+\mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$ holds for *all* vectors $\mathbf{x}$ and $\mathbf{y}$, so the argument must be an algebraic one using variables. Similarly, we must show $T(c\mathbf{x}) = cT(\mathbf{x})$ for all vectors $\mathbf{x}$ and all scalars c. It is not enough to check a few cases.

(b) What about the function $T: \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ 2 \end{bmatrix}$? Here we can see that both properties fail, but we only need to provide evidence that *one* fails. For example, $T\left(3\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = T\left(\begin{bmatrix} 3 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \neq 3\begin{bmatrix} 1 \\ 2 \end{bmatrix}$, which is what $3T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$ would be. The reader can also try checking whether

$$T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right).$$

Just a reminder: To check that a multi-part (in this case, two-part) definition *holds*, we must check each condition. However, to show that a multi-part definition *fails*, we only need to show that *one* of the criteria does not hold.

(c) We learned in Section 2 of Chapter 1 to project one vector onto another. We now think of this as defining a function: Let $\mathbf{a} \in \mathbb{R}^2$ be fixed and let $T: \mathbb{R}^2 \to \mathbb{R}^2$ be given by $T(\mathbf{x}) = \text{proj}_{\mathbf{a}}\mathbf{x}$. One can give a geometric argument that this is a linear map (see Exercise 15), but we will use our earlier formula from p. 22 to establish this. Since

$$\text{proj}_{\mathbf{a}}\mathbf{x} = \frac{\mathbf{x}\cdot\mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a},$$

we have

(i) $T(\mathbf{x}+\mathbf{y}) = \dfrac{(\mathbf{x}+\mathbf{y})\cdot\mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a} = \dfrac{\mathbf{x}\cdot\mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a} + \dfrac{\mathbf{y}\cdot\mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a} = T(\mathbf{x}) + T(\mathbf{y})$, and

(ii) $T(c\mathbf{x}) = \dfrac{(c\mathbf{x})\cdot\mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a} = c\dfrac{\mathbf{x}\cdot\mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a} = cT(\mathbf{x})$.

Notice that if we replace $\mathbf{a}$ with a nonzero scalar multiple of $\mathbf{a}$, the map T doesn't change. For this reason, we will refer to $T = \text{proj}_{\mathbf{a}}$ as the *projection of* $\mathbb{R}^2$ *onto the line* ℓ, where ℓ is the line spanned by $\mathbf{a}$. We will denote this mapping by P_ℓ.

(d) It follows from Exercise 1.4.13 (see also Proposition 5.2 of Chapter 1) that for any $m \times n$ matrix A, the function $\mu_A: \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. ▲

EXAMPLE 2

Expanding on the previous example, we consider the linear transformations $\mu_A: \mathbb{R}^2 \to \mathbb{R}^2$ for some specific 2×2 matrices A and give geometric interpretations of these maps.

(a) If $A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = O$ is the zero matrix, then $A\mathbf{x} = \mathbf{0}$ for all $\mathbf{x} \in \mathbb{R}^2$, so μ_A sends every vector in $\mathbb{R}^2$ to the zero vector $\mathbf{0}$. If $B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$ is the 2×2 identity matrix, then $B\mathbf{x} = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^2$. The function μ_B is the *identity map* from $\mathbb{R}^2$ to $\mathbb{R}^2$.

(b) Consider the linear transformation $T: \mathbb{R}^2 \to \mathbb{R}^2$ defined by multiplication by the matrix

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

The effect of T is pictured in Figure 2.1. One might slide a deck of cards in this fashion, and such a motion is called a *shear*.

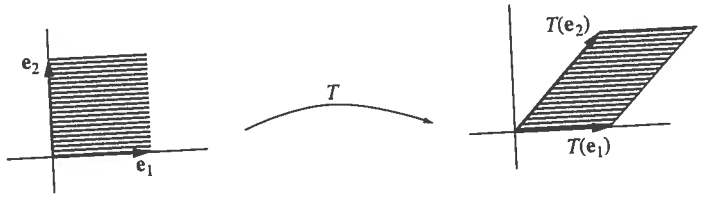

FIGURE 2.1

(c) Let

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

Then we have

$$A\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_1 \end{bmatrix},$$

and we see in Figure 2.2 that $A\mathbf{x}$ is obtained by rotating $\mathbf{x}$ an angle of $\pi/2$ counterclockwise.

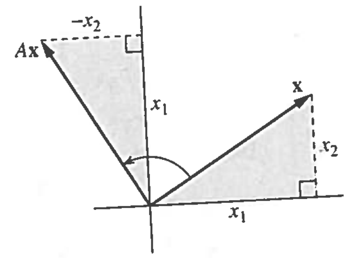

FIGURE 2.2

(d) Let

$$B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Then we have

$$B\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_1 \end{bmatrix},$$

as shown in Figure 2.3. We see that $B\mathbf{x}$ is the "mirror image" of the vector $\mathbf{x}$, reflecting across the "mirror" $x_1 = x_2$. In general, we say $T: \mathbb{R}^2 \to \mathbb{R}^2$ is given by *reflection* across a line ℓ if, for every $\mathbf{x} \in \mathbb{R}^2$, $T(\mathbf{x})$ has the same length as $\mathbf{x}$ and the two vectors make the same angle with ℓ.[1]

(e) Continuing with the matrices A and B from parts c and d, respectively, let's consider the function $\mu_{AB}: \mathbb{R}^2 \to \mathbb{R}^2$. Recalling that $\mu_{AB} = \mu_A \circ \mu_B$, we have the situation shown in Figure 2.4. The picture suggests that μ_{AB} is the linear transformation that gives reflection across the vertical axis, $x_1 = 0$. To be sure, we can compute algebraically:

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

[1] Strictly speaking, if the angle from ℓ to $\mathbf{x}$ is θ, then the angle from ℓ to $T(\mathbf{x})$ should be $-\theta$.

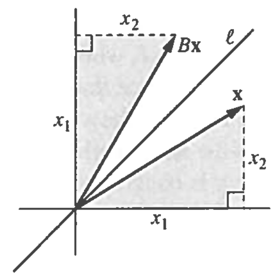

FIGURE 2.3

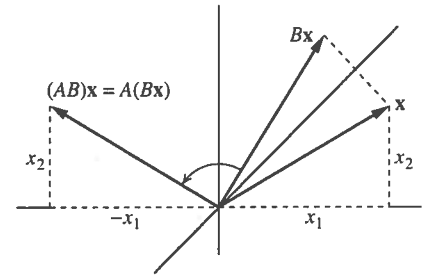

FIGURE 2.4

and so

$$(AB)\mathbf{x} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -x_1 \\ x_2 \end{bmatrix}.$$

This is indeed the formula for the reflection across the vertical axis. So we have seen that the function $\mu_A \circ \mu_B$—the composition of the reflection about the line $x_1 = x_2$ and a rotation through an angle of $\pi/2$—is the reflection across the line $x_1 = 0$. On the other hand, as indicated in Figure 2.5, the function $\mu_B \circ \mu_A = \mu_{BA}: \mathbb{R}^2 \to \mathbb{R}^2$, as we leave it to the reader to check, is the reflection across the line $x_2 = 0$. ▲

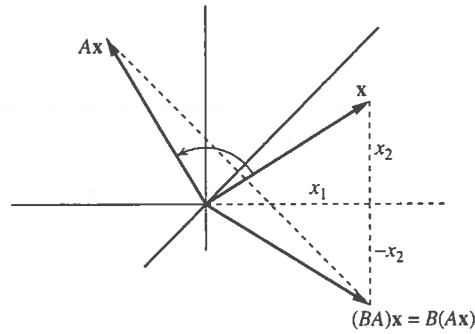

FIGURE 2.5

EXAMPLE 3

Continuing Example 2(d), if ℓ is a line in $\mathbb{R}^2$ through the origin, the *reflection across* ℓ is the map $R_\ell: \mathbb{R}^2 \to \mathbb{R}^2$ that sends $\mathbf{x}$ to its "mirror image" in ℓ. We begin by writing $\mathbf{x} = \mathbf{x}^{\parallel} + \mathbf{x}^{\perp}$, where $\mathbf{x}^{\parallel}$ is parallel to ℓ and $\mathbf{x}^{\perp}$ is orthogonal to ℓ, as in Section 2 of Chapter 1. Then, as we see in Figure 2.6,

$$R_\ell(\mathbf{x}) = \mathbf{x}^{\parallel} - \mathbf{x}^{\perp} = \mathbf{x}^{\parallel} - (\mathbf{x} - \mathbf{x}^{\parallel}) = 2\mathbf{x}^{\parallel} - \mathbf{x}.$$

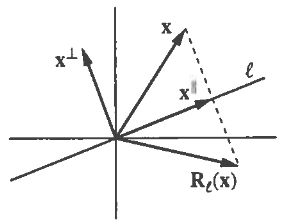

FIGURE 2.6

Using the notation of Example 1(c), we have $\mathbf{x}^{\parallel} = P_\ell(\mathbf{x})$, and so $R_\ell(\mathbf{x}) = 2P_\ell(\mathbf{x}) - \mathbf{x}$, or, in functional notation, $R_\ell = 2P_\ell - I$, where $I : \mathbb{R}^2 \to \mathbb{R}^2$ is the identity map. One can now use the result of Exercise 11 to deduce that R_ℓ is a linear map.

It is worth noting that $R_\ell(\mathbf{x})$ is the vector on the other side of ℓ from $\mathbf{x}$ that has the same length as $\mathbf{x}$ and makes the same angle with ℓ as $\mathbf{x}$ does. In particular, the right triangle with leg $\mathbf{x}^{\parallel}$ and hypotenuse $\mathbf{x}$ is congruent to the right triangle with leg $\mathbf{x}^{\parallel}$ and hypotenuse $R_\ell(\mathbf{x})$. This observation leads to a geometric argument that reflection across ℓ is indeed a linear transformation (see Exercise 15). ▲

EXAMPLE 4

We conclude this discussion with a few examples of linear transformations from $\mathbb{R}^3$ to $\mathbb{R}^3$.

(a) Let

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

Because μ_A leaves the x_2x_3-plane fixed and sends $(1, 0, 0)$ to $(-1, 0, 0)$, we see that $A\mathbf{x}$ is obtained by reflecting $\mathbf{x}$ across the x_2x_3-plane.

(b) Let

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

Then we have

$$B\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -x_2 \\ x_1 \\ x_3 \end{bmatrix}.$$

We see that μ_B leaves the x_3-axis fixed and rotates the x_1x_2-plane through an angle of $\pi/2$. Thus, μ_B rotates an arbitrary vector $\mathbf{x} \in \mathbb{R}^3$ an angle of $\pi/2$ about the x_3-axis, as pictured in Figure 2.7.

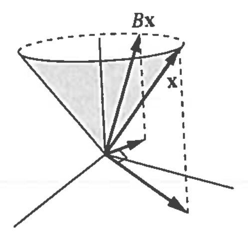

FIGURE 2.7

(c) Let $\mathbf{a} = (1, 1, 1)$. For any $\mathbf{x} \in \mathbb{R}^3$, we know that the projection of $\mathbf{x}$ onto $\mathbf{a}$ is given by

$$\text{proj}_{\mathbf{a}}\mathbf{x} = \frac{\mathbf{x} \cdot \mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a} = \frac{1}{3}(x_1 + x_2 + x_3)\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \\ x_1 + x_2 + x_3 \end{bmatrix}.$$

Thus, if we define the matrix

$$C = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

we have $\text{proj}_{\mathbf{a}}\mathbf{x} = C\mathbf{x}$. In particular, $\text{proj}_{\mathbf{a}}$ is the linear transformation μ_C. As we did earlier in $\mathbb{R}^2$, we can also denote this linear map by P_ℓ, where ℓ is the line spanned by $\mathbf{a}$. ▲

2.1 The Standard Matrix of a Linear Transformation

When we examine the previous examples, we find a geometric meaning of the *column* vectors of the matrices. As we know, when we multiply a matrix by a vector, we get the appropriate linear combination of the columns of the matrix. In particular,

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix}.$$

And so we see that the first column of A is the vector $A\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \mu_A\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)$ and the second column of A is the vector $A\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \mu_A\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)$. Turning this observation on its head, we note that we can find the matrix A (and hence the linear map μ_A) by finding the two vectors $\mu_A\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)$ and $\mu_A\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)$. This seems surprising at first, as the function μ_A is completely determined by what it does to only two (nonparallel) vectors in $\mathbb{R}^2$.

This is, in fact, a general property of linear maps. If, for example, $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation, then for any $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$, we write

$$\mathbf{x} = x_1\begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2\begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

and so, by the linearity properties,

$$\begin{aligned} T(\mathbf{x}) &= T\left(x_1\begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \\ &= x_1 T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + x_2 T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right). \end{aligned}$$

That is, once we know the two vectors $\mathbf{v}_1 = T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)$ and $\mathbf{v}_2 = T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)$, we can determine $T(\mathbf{x})$ for every $\mathbf{x} \in \mathbb{R}^2$. Indeed, if we create a 2×2 matrix by inserting $\mathbf{v}_1$ as the first column and $\mathbf{v}_2$ as the second column, then it follows from what we've done that $T = \mu_A$. Specifically, if

$$\mathbf{v}_1 = \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} \quad \text{and} \quad \mathbf{v}_2 = \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix},$$

then we obtain

$$A = \begin{bmatrix} | & | \\ \mathbf{v}_1 & \mathbf{v}_2 \\ | & | \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

A is called *the standard matrix* for T. (This entire discussion works more generally for linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$, but we will postpone that to Chapter 4.)

WARNING In order to apply the procedure we have just outlined, one *must* know in advance that the given function T is *linear*. If it is not, the matrix A constructed in this manner will *not* reproduce the original function T.

EXAMPLE 5

Let ℓ be the line in $\mathbb{R}^2$ spanned by $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and let $P_\ell\colon \mathbb{R}^2 \to \mathbb{R}^2$ be the projection onto ℓ. We checked in Example 1(c) that this is a linear map. Thus, we can find the standard matrix for P_ℓ. To do this, we compute

$$P\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \frac{1}{5}\begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{and} \quad P\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \frac{2}{5}\begin{bmatrix} 1 \\ 2 \end{bmatrix},$$

so the standard matrix representing P_ℓ is

$$A = \frac{1}{5}\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}.$$

Since we know that reflection across ℓ is given by $R_\ell = 2P_\ell - I$, the standard matrix for R_ℓ will be

$$B = 2 \cdot \frac{1}{5}\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} -3 & 4 \\ 4 & 3 \end{bmatrix}.$$

We ask the reader to find the matrix for reflection across a general line in Exercise 12. ▲

EXAMPLE 6

Generalizing Example 2(c), we consider the matrix

$$A_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

We see that

$$A_\theta\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \quad \text{and} \quad A_\theta\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix},$$

as pictured in Figure 2.8. Thus, the function μ_{A_θ} rotates $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ through the angle θ, and we strongly suspect that $\mu_{A_\theta}(\mathbf{x}) = A_\theta\mathbf{x}$ should be the vector obtained by rotating through angle θ. We leave it to the reader to check in Exercise 8 that this is the case, and we call A_θ a *rotation matrix*.

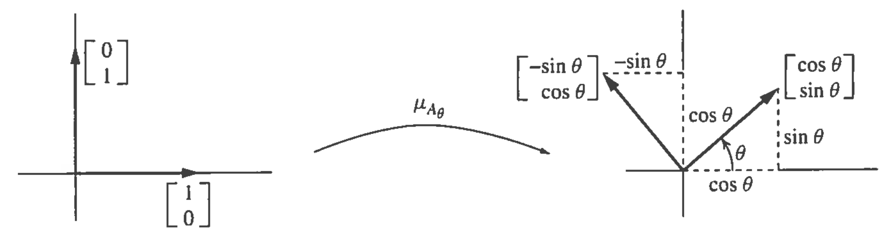

FIGURE 2.8

On the other hand, we could equally well have started with the map $T_\theta\colon \mathbb{R}^2 \to \mathbb{R}^2$ defined by rotating each vector counterclockwise by the angle θ. To take a geometric definition, the length of $T_\theta(\mathbf{x})$ is the same as the length of $\mathbf{x}$, and the angle between them is θ. Why is this map linear? Here is a detailed geometric justification. It is clear that if we rotate $\mathbf{x}$ and then multiply by a scalar c, we get the same result as rotating the vector $c\mathbf{x}$ (officially, the vector has the right length and makes the right angle with $c\mathbf{x}$). Now, as indicated in Figure 2.9, since the angle between $T_\theta(\mathbf{x})$ and $T_\theta(\mathbf{y})$ equals the angle between $\mathbf{x}$ and $\mathbf{y}$ (why?) and since lengths are preserved, it follows from the side-angle-side congruence theorem that the shaded triangles are congruent, and hence the parallelogram spanned by $\mathbf{x}$ and $\mathbf{y}$ is congruent to the parallelogram spanned by $T_\theta(\mathbf{x})$ and $T_\theta(\mathbf{y})$. The angle between $T_\theta(\mathbf{x}) + T_\theta(\mathbf{y})$ and $T_\theta(\mathbf{x})$ is the same as the angle between $\mathbf{x}+\mathbf{y}$ and $\mathbf{x}$, so, by simple arithmetic (do it!), the angle between $T_\theta(\mathbf{x}) + T_\theta(\mathbf{y})$ and $\mathbf{x}+\mathbf{y}$ is θ. Again because the parallelograms are congruent, $T_\theta(\mathbf{x}) + T_\theta(\mathbf{y})$ has the same length as $\mathbf{x}+\mathbf{y}$, hence the same length as $T_\theta(\mathbf{x}+\mathbf{y})$, and so the vectors $T_\theta(\mathbf{x}) + T_\theta(\mathbf{y})$ and $T_\theta(\mathbf{x}+\mathbf{y})$ must be equal. Whew!

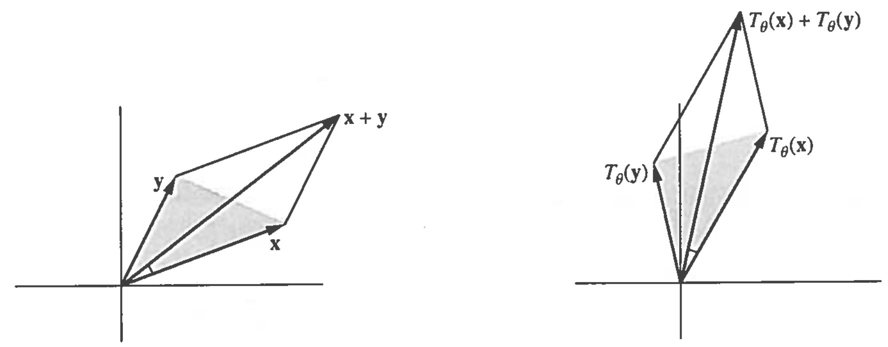

FIGURE 2.9

A natural question to ask is this: What is the product $A_\theta A_\phi$? The answer should be quite clear if we think of this as the composition of functions $\mu_{A_\theta A_\phi} = \mu_{A_\theta} \circ \mu_{A_\phi}$. We leave this to Exercise 7. ▲

EXAMPLE 7

The geometric interpretation of a given linear transformation is not always easy to determine just by looking at the matrix. For example, if we let

$$A = \begin{bmatrix} \frac{1}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{4}{5} \end{bmatrix},$$

then we might observe that for every $\mathbf{x} \in \mathbb{R}^2$, $A\mathbf{x}$ is a scalar multiple of the vector $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ (why?). From our past experience, what does this suggest? As a clue to understanding the associated linear transformation, we might try calculating A^2, and we find that $A^2 = A$; it follows that $A^n = A$ for all positive integers n (why?). What is the geometric explanation?

With some care we can unravel the mystery:

$$A\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{5}(x_1 + 2x_2) \\ \frac{2}{5}(x_1 + 2x_2) \end{bmatrix} = \frac{x_1 + 2x_2}{5}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \frac{\mathbf{x}\cdot(1,2)}{\|(1,2)\|^2}\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

is the projection of $\mathbf{x}$ onto the line spanned by $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$. (Of course, if one remembers Example 5, this was really no mystery.) This explains why $A^2\mathbf{x} = A\mathbf{x}$ for every $\mathbf{x} \in \mathbb{R}^2$: $A^2\mathbf{x} = A(A\mathbf{x})$, and once we've projected the vector $\mathbf{x}$ onto the line, it stays put. ▲

Exercises 2.2

1. Suppose that $T: \mathbb{R}^3 \to \mathbb{R}^2$ is a linear transformation and that

$$T\left(\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \quad\text{and}\quad T\left(\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 0 \end{bmatrix}.$$

Compute $T\left(2\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}\right)$, $T\left(\begin{bmatrix} 3 \\ 6 \\ 3 \end{bmatrix}\right)$, and $T\left(\begin{bmatrix} -1 \\ 3 \\ 0 \end{bmatrix}\right)$.

2. Suppose that $T: \mathbb{R}^3 \to \mathbb{R}^2$ is defined by $T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} x_1 + 2x_2 + x_3 \\ 3x_1 - x_2 - x_3 \end{bmatrix}$. Find a matrix A so that $T = \mu_A$.

3. Suppose $T: \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation. In each case, use the information provided to find the standard matrix A for T.
   *a. $T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$ and $T\left(\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$
   b. $T\left(\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 5 \\ 3 \end{bmatrix}$ and $T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$
   c. $T\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$ and $T\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

*4. Determine whether each of the following functions is a linear transformation. If so, provide a proof; if not, explain why.
   a. $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + 2x_2 \\ x_2^2 \end{bmatrix}$
   b. $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + 2x_2 \\ 0 \end{bmatrix}$

   c. $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = x_1 - x_2$
   d. $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} |x_2| \\ 3x_1 \end{bmatrix}$
   e. $T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + 2x_2 \\ x_2 \\ -x_1 + 3x_2 \end{bmatrix}$
   f. $T: \mathbb{R}^n \to \mathbb{R}$ given by $T(\mathbf{x}) = \|\mathbf{x}\|$

5. Give 2×2 matrices A so that for any $\mathbf{x} \in \mathbb{R}^2$ we have, respectively:
   a. $A\mathbf{x}$ is the vector whose components are, respectively, the sum and difference of the components of $\mathbf{x}$.
   *b. $A\mathbf{x}$ is the vector obtained by projecting $\mathbf{x}$ onto the line $x_1 = x_2$ in $\mathbb{R}^2$.
   c. $A\mathbf{x}$ is the vector obtained by first reflecting $\mathbf{x}$ across the line $x_1 = 0$ and then reflecting the resulting vector across the line $x_2 = x_1$.
   d. $A\mathbf{x}$ is the vector obtained by projecting $\mathbf{x}$ onto the line $2x_1 - x_2 = 0$.
   *e. $A\mathbf{x}$ is the vector obtained by first projecting $\mathbf{x}$ onto the line $2x_1 - x_2 = 0$ and then rotating the resulting vector $\pi/2$ counterclockwise.
   f. $A\mathbf{x}$ is the vector obtained by first rotating $\mathbf{x}$ an angle of $\pi/2$ counterclockwise and then projecting the resulting vector onto the line $2x_1 - x_2 = 0$.

*6. Let $T: \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation defined by rotating the plane $\pi/2$ counterclockwise; let $S: \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation defined by reflecting the plane across the line $x_1 + x_2 = 0$.
   a. Give the standard matrices representing S and T.
   b. Give the standard matrix representing $T \circ S$.
   c. Give the standard matrix representing $S \circ T$.

7. a. Calculate $A_\theta A_\phi$ and $A_\phi A_\theta$. (Recall the definition of the rotation matrix on p. 98.)
   b. Use your answer to part a to derive the addition formulas for sine and cosine.

8. Let A_θ be the rotation matrix defined on p. 98, $0 \le \theta \le \pi$. Prove that
   a. $\|A_\theta \mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{R}^2$.
   b. the angle between $\mathbf{x}$ and $A_\theta\mathbf{x}$ is θ.
   These properties characterize a rotation of the plane through angle θ.

9. Let ℓ be the line spanned by $\mathbf{a} \in \mathbb{R}^2$, and let $R_\ell: \mathbb{R}^2 \to \mathbb{R}^2$ be the linear map defined by reflection across ℓ. Using the formula $R_\ell(\mathbf{x}) = \mathbf{x}^\parallel - \mathbf{x}^\perp$ given in Example 3, verify that
   a. $\|R_\ell(\mathbf{x})\| = \|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{R}^2$.
   b. $R_\ell(\mathbf{x}) \cdot \mathbf{a} = \mathbf{x} \cdot \mathbf{a}$ for all $\mathbf{x} \in \mathbb{R}^2$; i.e., the angle between $\mathbf{x}$ and ℓ is the same as the angle between $R_\ell(\mathbf{x})$ and ℓ.

10. Let $T: \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Prove the following:
   a. $T(\mathbf{0}) = \mathbf{0}$
   b. $T(a\mathbf{u} + b\mathbf{v}) = aT(\mathbf{u}) + bT(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ and all scalars a and b

11. a. Prove that if $T: \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation and c is any scalar, then the function $cT: \mathbb{R}^n \to \mathbb{R}^m$ defined by $(cT)(\mathbf{x}) = cT(\mathbf{x})$ (i.e., the scalar c times the vector $T(\mathbf{x})$) is also a linear transformation.
   b. Prove that if $S: \mathbb{R}^n \to \mathbb{R}^m$ and $T: \mathbb{R}^n \to \mathbb{R}^m$ are linear transformations, then the function $S + T: \mathbb{R}^n \to \mathbb{R}^m$ defined by $(S+T)(\mathbf{x}) = S(\mathbf{x}) + T(\mathbf{x})$ is also a linear transformation.
   c. Prove that if $S: \mathbb{R}^m \to \mathbb{R}^p$ and $T: \mathbb{R}^n \to \mathbb{R}^m$ are linear transformations, then the function $S \circ T: \mathbb{R}^n \to \mathbb{R}^p$ is also a linear transformation.

12. a. Let ℓ be the line spanned by $\begin{bmatrix}\cos\theta\\ \sin\theta\end{bmatrix}$. Show that the standard matrix for R_ℓ is

$$R = \begin{bmatrix}\cos 2\theta & \sin 2\theta\\ \sin 2\theta & -\cos 2\theta\end{bmatrix}$$

by using Figure 2.10 and basic geometry to find the reflections of (1, 0) and (0, 1).

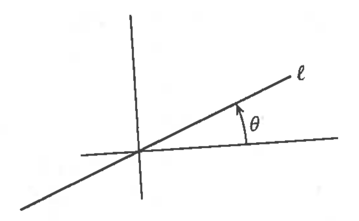

FIGURE 2.10

b. Derive this formula for R by using $R_\ell = 2P_\ell - I$ (see Example 3).

c. Letting A_θ be the rotation matrix defined on p. 98, check that

$$A_{2\theta}\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix} = R = A_\theta\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}A_{(-\theta)}.$$

d. Give geometric interpretations of these equalities.

13. Let ℓ be a line through the origin in $\mathbb{R}^2$.
a. Show that $P_\ell^2 = P_\ell \circ P_\ell = P_\ell$.
b. Show that $R_\ell^2 = R_\ell \circ R_\ell = I$.

14. Let ℓ_1 be the line through the origin in $\mathbb{R}^2$ making angle α with the x_1-axis, and let ℓ_2 be the line through the origin in $\mathbb{R}^2$ making angle β with the x_1-axis. Find $R_{\ell_2} \circ R_{\ell_1}$. (*Hint:* One approach is to use the matrix for reflection found in Exercise 12.)

15. Let $\ell \subset \mathbb{R}^2$ be a line through the origin.
a. Give a geometric argument that reflection across ℓ, the function $R_\ell\colon \mathbb{R}^2 \to \mathbb{R}^2$, is a linear transformation. (*Hint:* Consider the right triangles formed by $\mathbf{x}$ and $\mathbf{x}^\parallel$, $\mathbf{y}$ and $\mathbf{y}^\parallel$, and $\mathbf{x}+\mathbf{y}$ and $\mathbf{x}^\parallel + \mathbf{y}^\parallel$.)
b. Give a geometric argument that projection onto ℓ, the function $P_\ell\colon \mathbb{R}^2 \to \mathbb{R}^2$, is a linear transformation.

3 Inverse Matrices

Given an $m \times n$ matrix A, we are sometimes faced with the task of solving the equation $A\mathbf{x} = \mathbf{b}$ for several different values of $\mathbf{b} \in \mathbb{R}^m$. To accomplish this, it would be convenient to have an $n \times m$ matrix B satisfying $AB = I_m$: Taking $\mathbf{x} = B\mathbf{b}$, we will then have $A\mathbf{x} = A(B\mathbf{b}) = (AB)\mathbf{b} = I_m\mathbf{b} = \mathbf{b}$. This leads us to the following definition.

Definition. Given an $m \times n$ matrix A, an $n \times m$ matrix B is called a *right inverse* of A if $AB = I_m$. Similarly, an $n \times m$ matrix C is called a *left inverse* of A if $CA = I_n$.

Note the symmetry here: If B is a right inverse of A, then A is a left inverse of B, and vice versa. Also, thinking in terms of linear transformations, if B is a right inverse of A, for example, then $\mu_A \circ \mu_B$ is the identity mapping from $\mathbb{R}^m$ to $\mathbb{R}^m$.

EXAMPLE 1

Let $A = \begin{bmatrix}2 & -1 & 0\\ 1 & -2 & 2\end{bmatrix}$ and $B = \begin{bmatrix}0 & -1\\ -1 & -2\\ -1 & -1\end{bmatrix}$. Then

$$AB = \begin{bmatrix}2 & -1 & 0\\ 1 & -2 & 2\end{bmatrix}\begin{bmatrix}0 & -1\\ -1 & -2\\ -1 & -1\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix},$$

and so B is a right inverse of A (and A is a left inverse of B). Notice, however, that

$$BA = \begin{bmatrix}0 & -1\\ -1 & -2\\ -1 & -1\end{bmatrix}\begin{bmatrix}2 & -1 & 0\\ 1 & -2 & 2\end{bmatrix} = \begin{bmatrix}-1 & 2 & 2\\ -4 & 5 & -4\\ -3 & 3 & -2\end{bmatrix},$$

which is nothing like I_3. ▲

We observed earlier that if A has a right inverse, then we can always solve $A\mathbf{x} = \mathbf{b}$; i.e., this equation is consistent for every $\mathbf{b} \in \mathbb{R}^m$. On the other hand, if A has a left inverse, C, then a solution, *if it exists*, must be unique: If $A\mathbf{x} = \mathbf{b}$, then $C(A\mathbf{x}) = C\mathbf{b}$, and so $\mathbf{x} = I_n\mathbf{x} = (CA)\mathbf{x} = C(A\mathbf{x}) = C\mathbf{b}$. Thus, provided $\mathbf{x}$ is a solution of $A\mathbf{x} = \mathbf{b}$, then $\mathbf{x}$ must equal $C\mathbf{b}$, but maybe there aren't any solutions at all. To verify that $C\mathbf{b}$ is in fact a solution, we must calculate $A(C\mathbf{b})$ and see whether it is equal to $\mathbf{b}$. Of course, by associativity, this can be rewritten as $(AC)\mathbf{b} = \mathbf{b}$. This may or may not happen, but we do observe that if we want the vector $C\mathbf{b}$ to be a solution of $A\mathbf{x} = \mathbf{b}$ for *every* choice of $\mathbf{b} \in \mathbb{R}^m$, then we will need to have $AC = I_m$; i.e., we will need C to be both a left inverse *and* a right inverse of A. (This might be a good time to review the discussion of solving equations in the blue box on p. 23.)

We recall from Chapter 1 that, given the $m \times n$ matrix A, the equation $A\mathbf{x} = \mathbf{b}$ is *consistent* for all $\mathbf{b} \in \mathbb{R}^m$ precisely when the echelon form of A has no rows of 0's, i.e., when the rank of A is equal to m, the number of rows of A. On the other hand, the equation $A\mathbf{x} = \mathbf{b}$ has a *unique* solution precisely when the rank of A is equal to n, the number of columns of A. Summarizing, we have the following proposition.

Proposition 3.1. *If the $m \times n$ matrix A has a right inverse, then the rank of A must be m, and if A has a left inverse, then its rank must be n. Thus, if A has both a left inverse and a right inverse, it must be square ($n \times n$) with rank n.*

Now suppose A is a square, $n \times n$, matrix with right inverse B and left inverse C, so that

$$AB = I_n = CA.$$

Then, exploiting associativity of matrix multiplication, we have

$$(*) \qquad C = CI_n = C(AB) = (CA)B = I_nB = B.$$

That is, if A has both a left inverse and a right inverse, they must be equal. This leads us to the following definition.

Definition. An $n \times n$ matrix A is *invertible* if there is an $n \times n$ matrix B satisfying
$$AB = I_n \quad \text{and} \quad BA = I_n.$$
The matrix B is usually denoted A^{-1} (read "A inverse").

Remark. Note that if $B = A^{-1}$, then it is also the case that $A = B^{-1}$. We also note from equation (∗) that the inverse is unique: If B and C are both inverses of A, then, in particular, $AB = I_n$ and $CA = I_n$, so $B = C$.

EXAMPLE 2

Let
$$A = \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix}.$$
Then $AB = I_2$ and $BA = I_2$, so B is the inverse matrix of A. ▲

It is a consequence of our earlier discussion that if A is an invertible $n \times n$ matrix, then $A\mathbf{x} = \mathbf{c}$ has a unique solution for every $\mathbf{c} \in \mathbb{R}^n$, and so it follows from Proposition 5.5 of Chapter 1 that A must be nonsingular. What about the converse? If A is nonsingular, must A be invertible? Well, if A is nonsingular, we know that every equation $A\mathbf{x} = \mathbf{c}$ has a unique solution. In particular, if $\mathbf{e}_j = (0, \dots, 0, 1, 0, \dots, 0)$ is the vector with all entries 0 except for a 1 in the j^{th} slot, there is a unique vector $\mathbf{b}_j$ that solves $A\mathbf{b}_j = \mathbf{e}_j$. If we let B be the $n \times n$ matrix whose column vectors are $\mathbf{b}_1, \dots, \mathbf{b}_n$, then we have
$$AB = A\begin{bmatrix} | & | & & | \\ \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \\ | & | & & | \end{bmatrix} = I_n.$$
This suggests that the matrix we've constructed should be the inverse matrix of A. But we need to know that $BA = I_n$ as well. Here is a very elegant way to understand why this is so. We can find the matrix B by forming the giant augmented matrix (see Exercise 1.4.7)
$$\left[\begin{array}{c|ccc} A & \mathbf{e}_1 & \cdots & \mathbf{e}_n \end{array}\right] = \left[\begin{array}{c|c} A & I_n \end{array}\right]$$
and using Gaussian elimination to obtain the reduced echelon form
$$\left[\begin{array}{c|c} I_n & B \end{array}\right].$$
(Note that the reduced echelon form of A must be I_n because A is nonsingular.) Now here is the tricky part: By reversing the row operations, we find that the augmented matrix
$$\left[\begin{array}{c|c} B & I_n \end{array}\right]$$
is transformed to
$$\left[\begin{array}{c|c} I_n & A \end{array}\right].$$
This says that $BA = I_n$, which is what we needed to check. In conclusion, we have proved the following theorem.

Theorem 3.2. *An $n \times n$ matrix is nonsingular if and only if it is invertible.*

Note that Gaussian elimination will also let us know when A is not invertible: If we come to a row of 0's while we are reducing A to echelon form, then, of course, A is singular and so it cannot be invertible.

EXAMPLE 3

We wish to determine the inverse of the matrix
$$A = \begin{bmatrix} 1 & -1 & 1 \\ 2 & -1 & 0 \\ 1 & -2 & 2 \end{bmatrix}$$
(if it exists). We apply Gaussian elimination to the augmented matrix:
$$\left[\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 2 & -1 & 0 & 0 & 1 & 0 \\ 1 & -2 & 2 & 0 & 0 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -2 & -2 & 1 & 0 \\ 0 & -1 & 1 & -1 & 0 & 1 \end{array}\right]$$
$$\rightsquigarrow \left[\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -2 & -2 & 1 & 0 \\ 0 & 0 & -1 & -3 & 1 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|ccc} 1 & -1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -2 & -2 & 1 & 0 \\ 0 & 0 & 1 & 3 & -1 & -1 \end{array}\right]$$
$$\rightsquigarrow \left[\begin{array}{ccc|ccc} 1 & -1 & 0 & -2 & 1 & 1 \\ 0 & 1 & 0 & 4 & -1 & -2 \\ 0 & 0 & 1 & 3 & -1 & -1 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 2 & 0 & -1 \\ 0 & 1 & 0 & 4 & -1 & -2 \\ 0 & 0 & 1 & 3 & -1 & -1 \end{array}\right]$$
Since we have determined that A is nonsingular, it follows that
$$A^{-1} = \begin{bmatrix} 2 & 0 & -1 \\ 4 & -1 & -2 \\ 3 & -1 & -1 \end{bmatrix}.$$
(The reader should check our arithmetic by multiplying AA^{-1} or $A^{-1}A$.) ▲

EXAMPLE 4

It is convenient to derive a formula for the inverse (when it exists) of a general 2×2 matrix
$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

We assume $a \neq 0$ to start with. Then

$$\left[\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ c & d & 0 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & d - \frac{bc}{a} & -\frac{c}{a} & 1 \end{array}\right] \quad (\text{assuming } ad - bc \neq 0)$$

$$\rightsquigarrow \left[\begin{array}{cc|cc} 1 & \frac{b}{a} & \frac{1}{a} & 0 \\ 0 & 1 & -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|cc} 1 & 0 & \frac{1}{a} - \frac{b}{a}\left(-\frac{c}{ad-bc}\right) & -\frac{b}{a}\frac{a}{ad-bc} \\ 0 & 1 & -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right]$$

$$= \left[\begin{array}{cc|cc} 1 & 0 & \frac{d}{ad-bc} & -\frac{b}{ad-bc} \\ 0 & 1 & -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right],$$

and so we see that, provided $ad - bc \neq 0$,

$$A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

As a check, we have

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = I_2 = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Of course, we have derived this assuming $a \neq 0$, but the reader can check easily that the formula works fine even when $a = 0$. We do see, however, from the row reduction that

$$\boxed{\begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ is nonsingular } \iff ad - bc \neq 0,}$$

because if $ad - bc = 0$, then we get a row of 0's in the echelon form of A. ▲

EXAMPLE 5

It follows immediately from Example 4 that for our rotation matrix

$$A_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \quad \text{we have} \quad A_\theta^{-1} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

Since $\cos(-\theta) = \cos\theta$ and $\sin(-\theta) = -\sin\theta$, we see that this is the matrix $A_{(-\theta)}$. If we think about the corresponding functions μ_{A_θ} and $\mu_{A_{(-\theta)}}$, this result becomes obvious: To invert (or "undo") a rotation through angle θ, we must rotate through angle $-\theta$. ▲

By now it may have occurred to the reader that for square matrices, a one-sided inverse must actually be a true inverse. We formalize this observation here.

Corollary 3.3. *If A and B are $n \times n$ matrices satisfying $BA = I_n$, then $B = A^{-1}$ and $A = B^{-1}$.*

Proof. If $A\mathbf{x} = \mathbf{0}$, then $\mathbf{x} = (BA)\mathbf{x} = B(A\mathbf{x}) = \mathbf{0}$, so, by Proposition 5.5 of Chapter 1, A is nonsingular. According to Theorem 3.2, A is therefore invertible. Since A has an inverse matrix, A^{-1}, we deduce that[2]

$$\begin{aligned} BA &= I_n \\ &\Downarrow \text{ multiplying both sides of the equation by } A^{-1} \text{ on the right} \\ (BA)A^{-1} &= I_n A^{-1} \\ &\Downarrow \text{ using the associative property} \\ B(AA^{-1}) &= A^{-1} \\ &\Downarrow \text{ using the definition of } A^{-1} \\ B &= A^{-1}, \end{aligned}$$

as desired. Because $AB = I_n$ and $BA = I_n$, it now follows that $A = B^{-1}$, as well. □

EXAMPLE 6

We can use Gaussian elimination to find a *right* inverse of an $m \times n$ matrix A, so long as the rank of A is equal to m. The fact that we have free variables when $m < n$ will give many choices of right inverse. For example, taking

$$A = \begin{bmatrix} 1 & -1 & 1 \\ 2 & -1 & 0 \end{bmatrix},$$

we apply Gaussian elimination to the augmented matrix

$$\left[\begin{array}{ccc|cc} 1 & -1 & 1 & 1 & 0 \\ 2 & -1 & 0 & 0 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|cc} 1 & -1 & 1 & 1 & 0 \\ 0 & 1 & -2 & -2 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|cc} 1 & 0 & -1 & -1 & 1 \\ 0 & 1 & -2 & -2 & 1 \end{array}\right].$$

From this we see that the general solution of $A\mathbf{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is

$$\mathbf{x} = \begin{bmatrix} -1 \\ -2 \\ 0 \end{bmatrix} + s\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

and the general solution of $A\mathbf{x} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ is

$$\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}.$$

If we take $s = t = 0$, we get the right inverse

$$B = \begin{bmatrix} -1 & 1 \\ -2 & 1 \\ 0 & 0 \end{bmatrix},$$

[2]We are writing the "implies" symbol ($\Longrightarrow$) vertically so that we can indicate the reasoning in each step.

but we could take, say, $s = 1$ and $t = -1$ to obtain another right inverse,

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

Finding a left inverse is a bit trickier. You can sometimes do it with a little guesswork, or you can set up a large system of equations to solve (thinking of the entries of the left inverse as the unknowns), but we will discuss a more systematic approach in the next section. ▲

We end this discussion with a very important observation.

Proposition 3.4. *Suppose A and B are invertible $n \times n$ matrices. Then their product AB is invertible, and*

$$(AB)^{-1} = B^{-1}A^{-1}.$$

Remark. Some people refer to this result rather endearingly as the "shoe-sock theorem," for to undo (invert) the process of putting on one's socks and then one's shoes, one must first remove the shoes and then remove the socks.

Proof. To prove the matrix AB is invertible, we need only check that the candidate for the inverse works. That is, we need to check that

$$(AB)(B^{-1}A^{-1}) = I_n \quad \text{and} \quad (B^{-1}A^{-1})(AB) = I_n.$$

But these follow immediately from associativity:

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AI_nA^{-1} = AA^{-1} = I_n, \quad \text{and}$$
$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}I_nB = B^{-1}B = I_n.$$ □

Exercises 2.3

1. Use Gaussian elimination to find A^{-1} (if it exists):

*a. $A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}$ d. $A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 2 \\ 0 & 1 & 2 \end{bmatrix}$ f. $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$

b. $A = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$ *e. $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \\ -1 & 3 & 1 \end{bmatrix}$ *g. $A = \begin{bmatrix} 2 & 3 & 4 \\ 2 & 1 & 1 \\ -1 & 1 & 2 \end{bmatrix}$

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

2. In each case, given A and $\mathbf{b}$,

(i) Find A^{-1}.

(ii) Use your answer to (i) to solve $A\mathbf{x} = \mathbf{b}$.

(iii) Use your answer to (ii) to express $\mathbf{b}$ as a linear combination of the columns of A.

a. $A = \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ *b. $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 3 \\ 3 & 2 & 2 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$

c. $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}$ *d. $A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 1 & 4 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 2 \\ 0 \\ 1 \\ 1 \end{bmatrix}$

*3. Suppose A is an $n \times n$ matrix and B is an invertible $n \times n$ matrix. Simplify the following.
a. $(BAB^{-1})^2$
b. $(BAB^{-1})^n$ (n a positive integer)
c. $(BAB^{-1})^{-1}$ (what additional assumption is required here?)

*4. Suppose A is an invertible $n \times n$ matrix and $\mathbf{x} \in \mathbb{R}^n$ satisfies $A\mathbf{x} = 7\mathbf{x}$. Calculate $A^{-1}\mathbf{x}$.

5. If P is a permutation matrix (see Exercise 2.1.12 for the definition), show that P is invertible and find P^{-1}.

6. a. Give another right inverse of the matrix A in Example 6.
b. Find two right inverses of the matrix $A = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$.
c. Find two right inverses of the matrix $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \end{bmatrix}$.

7. a. Give a matrix that has a left inverse but no right inverse.
b. Give a matrix that has a right inverse but no left inverse.
c. Find two left inverses of the matrix $A = \begin{bmatrix} 1 & 2 \\ 0 & -1 \\ 1 & 1 \end{bmatrix}$.

*8. Suppose A is a square matrix satisfying the equation $A^3 - 3A + I = O$. Show that A is invertible. (*Hint:* Can you give an explicit formula for A^{-1}?)

9. Suppose A is a square matrix satisfying the equation $A^3 - 2I = O$. Prove that A and $A - I$ are both invertible. (*Hint:* Give explicit formulas for their inverses. In the second case, a little trickery will be necessary: Start by factoring $x^3 - 1$.)

10. Suppose A is an $n \times n$ matrix with the property that $A - I$ is invertible.
a. For any $k = 1, 2, 3, \ldots$, give a formula for $(A - I)^{-1}(A^{k+1} - I)$. (*Hint:* Think about simplifying $\dfrac{x^{k+1} - 1}{x - 1}$ for $x \neq 1$.)
b. Use your answer to part a to find the number of paths of length ≤ 6 from node 1 to node 3 in Example 5 in Section 1.

11. Suppose A and B are $n \times n$ matrices. Prove that if AB is nonsingular, then both A and B are nonsingular. (*Hint:* First show that B is nonsingular; then use Theorem 3.2 and Proposition 3.4.)

12. Suppose A is an invertible $m \times m$ matrix and B is an invertible $n \times n$ matrix. (See Exercise 2.1.9 for the notion of block multiplication.)
a. Show that the matrix

$$\left[\begin{array}{c|c} A & O \\ \hline O & B \end{array}\right]$$

is invertible and give a formula for its inverse.

b. Suppose C is an arbitrary $m \times n$ matrix. Is the matrix

$$\left[\begin{array}{c|c} A & C \\ \hline O & B \end{array}\right]$$

invertible?

13. Suppose A is an invertible matrix and A^{-1} is known.

a. Suppose B is obtained from A by switching two columns. How can we find B^{-1} from A^{-1}? (*Hint:* Since $A^{-1}A = I$, we know the dot products of the *rows* of A^{-1} with the *columns* of A. So rearranging the columns of A to make B, we should be able to suitably rearrange the rows of A^{-1} to make B^{-1}.)

b. Suppose B is obtained from A by multiplying the j^{th} column by a nonzero scalar. How can we find B^{-1} from A^{-1}?

c. Suppose B is obtained from A by adding a scalar multiple of one column to another. How can we find B^{-1} from A^{-1}?

d. Suppose B is obtained from A by replacing the j^{th} column by a different vector. Assuming B is still invertible, how can we find B^{-1} from A^{-1}?

14. Let A be an $m \times n$ matrix.

a. Assume the rank of A is m and B is a right inverse of A. Show that B' is another right inverse of A if and only if $A(B - B') = O$ and that this occurs if and only if every column of $B - B'$ is orthogonal to every row of A.

b. Assume the rank of A is n and C is a left inverse of A. Show that C' is another left inverse of A if and only if $(C - C')A = O$ and that this occurs if and only if every row of $C - C'$ is orthogonal to every column of A.

15. Suppose A is an $m \times n$ matrix with a *unique* right inverse B. Prove that $m = n$ and that A is invertible.

16. Suppose A is an $n \times n$ matrix satisfying $A^{10} = O$. Prove that the matrix $I_n - A$ is invertible. (*Hint:* As a warm-up, try assuming $A^2 = O$.)

4 Elementary Matrices: Rows Get Equal Time

So far we have focused on interpreting matrix multiplication in terms of columns—that is, on the fact that the j^{th} column of AB is the product of A with the j^{th} column vector of B. But equally relevant is the following observation:

> The i^{th} row of AB is the product of the i^{th} row vector of A with B.

Just as multiplying the matrix A by a column vector $\mathbf{x}$ on the right,

$$\begin{bmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

gives us the linear combination $x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$ of the columns of A, the reader

can easily check that multiplying A on the *left* by the row vector $[x_1\ x_2\ \cdots\ x_m]$,

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_m \end{bmatrix} \begin{bmatrix} — & \mathbf{A}_1 & — \\ — & \mathbf{A}_2 & — \\ & \vdots & \\ — & \mathbf{A}_m & — \end{bmatrix},$$

yields the linear combination $x_1\mathbf{A}_1 + x_2\mathbf{A}_2 + \cdots + x_m\mathbf{A}_m$ of the *rows* of A.

It should come as no surprise, then, that we can perform row operations on a matrix A by multiplying on the *left* by appropriately chosen matrices. For example, if

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix},$$

$$E_1 = \begin{bmatrix} & 1 & \\ 1 & & \\ & & 1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 4 \end{bmatrix}, \quad \text{and} \quad E_3 = \begin{bmatrix} 1 & & \\ -2 & 1 & \\ & & 1 \end{bmatrix},$$

then

$$E_1A = \begin{bmatrix} 3 & 4 \\ 1 & 2 \\ 5 & 6 \end{bmatrix}, \quad E_2A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 20 & 24 \end{bmatrix}, \quad \text{and} \quad E_3A = \begin{bmatrix} 1 & 2 \\ 1 & 0 \\ 5 & 6 \end{bmatrix}.$$

Here we establish the custom that when it is clearer to do so, we indicate 0 entries in a matrix by blank spaces.

Such matrices that give corresponding elementary row operations are called *elementary matrices*. Note that each elementary matrix differs from the identity matrix only in a small way.

(i) To interchange rows i and j, we should multiply by an elementary matrix of the form

$$\begin{array}{c} \\ i \rightarrow \\ \\ j \rightarrow \\ \\ \end{array} \overset{\qquad\quad i \downarrow \qquad j \downarrow}{\begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & \cdots & 0 & \cdots & 1 & \cdots & \\ & & & \ddots & & & \\ & \cdots & 1 & \cdots & 0 & \cdots & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix}}.$$

(ii) To multiply row i by a scalar c, we should multiply by an elementary matrix of the form

$$\begin{array}{cc} & \begin{matrix} & & & \overset{i}{\downarrow} & & & \end{matrix} \\ \begin{matrix} \\ \\ \\ i \rightarrow \\ \\ \\ \\ \end{matrix} & \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & c & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} \end{array}.$$

(iii) To add c times row i to row j, we should multiply by an elementary matrix of the form

$$\begin{array}{cc} & \begin{matrix} & & \overset{i}{\downarrow} & & \overset{j}{\downarrow} & & \end{matrix} \\ \begin{matrix} \\ \\ i \rightarrow \\ \\ j \rightarrow \\ \\ \\ \end{matrix} & \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & \ddots & & & \\ & \cdots & c & \cdots & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} \end{array}.$$

Here's an easy way to remember the form of these matrices: *Each elementary matrix is obtained by performing the corresponding elementary row operation on the identity matrix.*

EXAMPLE 1

Let $A = \begin{bmatrix} 4 & 3 & 5 \\ 1 & 2 & 5 \end{bmatrix}$. We put A in reduced echelon form by the following sequence of row operations:

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

These steps correspond to multiplying, in sequence from right to left, by the elementary matrices

$$E_1 = \begin{bmatrix} & 1 \\ 1 & \end{bmatrix}, \quad E_2 = \begin{bmatrix} 1 & \\ -4 & 1 \end{bmatrix}, \quad E_3 = \begin{bmatrix} 1 & \\ & -\frac{1}{5} \end{bmatrix}, \quad E_4 = \begin{bmatrix} 1 & -2 \\ & 1 \end{bmatrix}.$$

Now the reader can check that

$$E = E_4E_3E_2E_1 = \begin{bmatrix} 1 & -2 \\ & 1 \end{bmatrix} \begin{bmatrix} 1 & \\ & -\frac{1}{5} \end{bmatrix} \begin{bmatrix} 1 & \\ -4 & 1 \end{bmatrix} \begin{bmatrix} & 1 \\ 1 & \end{bmatrix} = \begin{bmatrix} \frac{2}{5} & -\frac{3}{5} \\ -\frac{1}{5} & \frac{4}{5} \end{bmatrix}$$

and, indeed,

$$EA = \begin{bmatrix} \frac{2}{5} & -\frac{3}{5} \\ -\frac{1}{5} & \frac{4}{5} \end{bmatrix} \begin{bmatrix} 4 & 3 & 5 \\ 1 & 2 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 3 \end{bmatrix},$$

as it should. Remember: The elementary matrices are arranged *from right to left* in the order in which the operations are done on A. ▲

EXAMPLE 2

Let's revisit Example 6 on p. 47. Let

$$A = \begin{bmatrix} 1 & 1 & 3 & -1 & 0 \\ -1 & 1 & 1 & 1 & 2 \\ 0 & 1 & 2 & 2 & -1 \\ 2 & -1 & 0 & 1 & -6 \end{bmatrix}.$$

To clear out the entries below the first pivot, we must multiply by the product of the two elementary matrices E_1 and E_2:

$$E_2E_1 = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ -2 & & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} = \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ & & 1 & \\ -2 & & & 1 \end{bmatrix};$$

to change the pivot in the second row to 1 and then clear out below, we multiply first by

$$E_3 = \begin{bmatrix} 1 & & & \\ & \frac{1}{2} & & \\ & & 1 & \\ & & & 1 \end{bmatrix}$$

and then by the product

$$E_5E_4 = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & 3 & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & \\ & 1 & & \\ & -1 & 1 & \\ & & & 1 \end{bmatrix} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & -1 & 1 & \\ & 3 & & 1 \end{bmatrix}.$$

We next change the pivot in the third row to 1 and clear out below, multiplying by

$$E_6 = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \frac{1}{2} & \\ & & & 1 \end{bmatrix} \quad \text{and} \quad E_7 = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & -3 & 1 \end{bmatrix}.$$

Now we clear out above the pivots by multiplying by

$$E_8 = \begin{bmatrix} 1 & & 1 & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} \quad \text{and} \quad E_9 = \begin{bmatrix} 1 & -1 & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}.$$

The net result is this: When we multiply the product

$$E_9E_8E_7E_6(E_5E_4)E_3(E_2E_1) = \begin{bmatrix} \frac{1}{4} & -\frac{3}{4} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ -\frac{1}{4} & -\frac{1}{4} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{9}{4} & -\frac{3}{2} & 1 \end{bmatrix}$$

by the original matrix, we do in fact get the reduced echelon form:

$$\begin{bmatrix} \frac{1}{4} & -\frac{3}{4} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ -\frac{1}{4} & -\frac{1}{4} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{9}{4} & -\frac{3}{2} & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 3 & -1 & 0 \\ -1 & 1 & 1 & 1 & 2 \\ 0 & 1 & 2 & 2 & -1 \\ 2 & -1 & 0 & 1 & -6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 & -2 \\ 0 & 1 & 2 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

▲

We now turn to some applications of elementary matrices to concepts we have studied earlier. Recall from Chapter 1 that if we want to find the constraint equations that a vector **b** must satisfy in order for $A\mathbf{x} = \mathbf{b}$ to be consistent, we reduce the augmented matrix $[\,A \mid \mathbf{b}\,]$ to echelon form $[\,U \mid \mathbf{c}\,]$ and set equal to 0 those entries of **c** corresponding to the rows of 0's in U. That is, when A is an $m \times n$ matrix of rank r, the constraint equations are merely the equations $c_{r+1} = \cdots = c_m = 0$. Letting E be the product of the elementary matrices corresponding to the elementary row operations required to put A in echelon form, we have $U = EA$, and so

$$(\dagger) \qquad [\,U \mid \mathbf{c}\,] = [\,EA \mid E\mathbf{b}\,].$$

That is, the constraint equations are the equations

$$\mathbf{E}_{r+1} \cdot \mathbf{b} = 0, \quad \dots, \quad \mathbf{E}_m \cdot \mathbf{b} = 0,$$

where, we recall, $\mathbf{E}_{r+1}, \dots, \mathbf{E}_m$ are the last $m - r$ row vectors of E. Interestingly, we can use the equation (†) to find a simple way to compute E: When we reduce the augmented matrix $[\,A \mid \mathbf{b}\,]$ to echelon form $[\,U \mid \mathbf{c}\,]$, E is the matrix satisfying $E\mathbf{b} = \mathbf{c}$.

EXAMPLE 3

Let's once again consider the matrix

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

from Example 2, and let's find the constraint equations for $A\mathbf{x} = \mathbf{b}$ to be consistent. We start with the augmented matrix

$$[\,A \mid \mathbf{b}\,] = \left[\begin{array}{rrrrr|c} 1 & 1 & 3 & -1 & 0 & b_1 \\ -1 & 1 & 1 & 1 & 2 & b_2 \\ 0 & 1 & 2 & 2 & -1 & b_3 \\ 2 & -1 & 0 & 1 & -6 & b_4 \end{array}\right]$$

and reduce to echelon form

$$[\,U \mid \mathbf{c}\,] = \left[\begin{array}{rrrrr|l} 1 & 1 & 3 & -1 & 0 & b_1 \\ 0 & 2 & 4 & 0 & 2 & b_1 + b_2 \\ 0 & 0 & 0 & 2 & -2 & -\frac{1}{2}b_1 - \frac{1}{2}b_2 + b_3 \\ 0 & 0 & 0 & 0 & 0 & b_1 + 9b_2 - 6b_3 + 4b_4 \end{array}\right].$$

(Note that we have arranged to remove fractions from the entry in the last row.) Now it is easy to see that if

$$E\mathbf{b} = \begin{bmatrix} b_1 \\ b_1 + b_2 \\ -\frac{1}{2}b_1 - \frac{1}{2}b_2 + b_3 \\ b_1 + 9b_2 - 6b_3 + 4b_4 \end{bmatrix}, \quad \text{then} \quad E = \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ -\frac{1}{2} & -\frac{1}{2} & 1 & \\ 1 & 9 & -6 & 4 \end{bmatrix}.$$

The reader should check that, in fact, $EA = U$.

We could continue our Gaussian elimination to reach reduced echelon form:

$$[\,R \mid \mathbf{d}\,] = \left[\begin{array}{rrrrr|l} 1 & 0 & 1 & 0 & -2 & \frac{1}{4}b_1 - \frac{3}{4}b_2 + \frac{1}{2}b_3 \\ 0 & 1 & 2 & 0 & 1 & \frac{1}{2}b_1 + \frac{1}{2}b_2 \\ 0 & 0 & 0 & 1 & -1 & -\frac{1}{4}b_1 - \frac{1}{4}b_2 + \frac{1}{2}b_3 \\ 0 & 0 & 0 & 0 & 0 & b_1 + 9b_2 - 6b_3 + 4b_4 \end{array}\right].$$

From this we see that $R = E'A$, where

$$E' = \begin{bmatrix} \frac{1}{4} & -\frac{3}{4} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ -\frac{1}{4} & -\frac{1}{4} & \frac{1}{2} & 0 \\ 1 & 9 & -6 & 4 \end{bmatrix},$$

which is very close to—but not the same as—the product of elementary matrices we obtained at the end of Example 2. Can you explain why the first three rows must agree here, but not the last? ▲

EXAMPLE 4

If an $m \times n$ matrix A has rank n, then every column is a pivot column, so its reduced echelon form must be $R = \left[\begin{array}{c} I_n \\ \hline \mathrm{O} \end{array}\right]$. If we find a product, E, of elementary matrices so that $EA = R$, then the first m rows of E will give us a left inverse of A. For example, if $A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 2 & 1 \end{bmatrix}$, then we can take

$$E = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & -\frac{3}{2} & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 \\ \frac{3}{2} & -\frac{3}{2} & 1 \end{bmatrix},$$

and so

$$C = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 \end{bmatrix}$$

is a left inverse of A (as the diligent reader should check). ▲

4.1 The *LU* Decomposition

As a final topic in this section, let's reexamine the process of putting a matrix in echelon form by using elementary matrices. The crucial point is that elementary matrices are invertible and their inverses are elementary matrices of the same type (see Exercise 7). If $E = E_k \cdots E_2E_1$ is the product of the elementary matrices we use to reduce A to echelon form, then $U = EA$ and so $A = E^{-1}U$. Suppose that we use *only* lower triangular elementary matrices of type (**iii**): No row interchanges are required, and no rows are multiplied through by a scalar. In this event, all the E_i are lower triangular matrices with 1's on the diagonal, and so E is lower triangular with 1's on the diagonal, and E^{-1} has the same property. In this case, then, we've written $A = LU$, where $L = E^{-1}$ is a lower triangular (square) matrix with 1's on the diagonal. This is called the *LU decomposition* of A.

EXAMPLE 5

Let

$$A = \begin{bmatrix} 1 & 2 & -1 & 1 \\ 2 & 7 & 4 & 2 \\ -1 & 4 & 13 & -1 \end{bmatrix}.$$

We reduce A to echelon form by the following sequence of row operations:

$$\begin{bmatrix} 1 & 2 & -1 & 1 \\ 2 & 7 & 4 & 2 \\ -1 & 4 & 13 & -1 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & -1 & 1 \\ 0 & 3 & 6 & 0 \\ -1 & 4 & 13 & -1 \end{bmatrix}$$
$$\rightsquigarrow \begin{bmatrix} 1 & 2 & -1 & 1 \\ 0 & 3 & 6 & 0 \\ 0 & 6 & 12 & 0 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 2 & -1 & 1 \\ 0 & 3 & 6 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = U.$$

This is accomplished by multiplying by the respective elementary matrices

$$E_1 = \begin{bmatrix} 1 & & \\ -2 & 1 & \\ & & 1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 1 & & \\ & 1 & \\ 1 & & 1 \end{bmatrix}, \quad \text{and} \quad E_3 = \begin{bmatrix} 1 & & \\ & 1 & \\ & -2 & 1 \end{bmatrix}.$$

Thus we have the equation

$$E_3E_2E_1A = U,$$

whence

$$A = (E_3E_2E_1)^{-1}U = E_1^{-1}E_2^{-1}E_3^{-1}U.$$

Note that it is easier to calculate the inverses of the elementary matrices (see Exercise 7) and then calculate their product. In our case,

$$E_1^{-1} = \begin{bmatrix} 1 & & \\ 2 & 1 & \\ & & 1 \end{bmatrix}, \quad E_2^{-1} = \begin{bmatrix} 1 & & \\ & 1 & \\ -1 & & 1 \end{bmatrix}, \quad \text{and} \quad E_3^{-1} = \begin{bmatrix} 1 & & \\ & 1 & \\ & 2 & 1 \end{bmatrix},$$

and so

$$L = E_1^{-1}E_2^{-1}E_3^{-1} = \begin{bmatrix} 1 & & \\ 2 & 1 & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & \\ -1 & & 1 \end{bmatrix} \begin{bmatrix} 1 & & \\ & 1 & \\ & 2 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 1 & & \\ 2 & 1 & \\ -1 & 2 & 1 \end{bmatrix}.$$

In fact, we see that when $i > j$, the ij-entry of L is the *negative* of the multiple of row j that we added to row i during our row operations.

Our LU decomposition, then, is as follows:

$$A = \begin{bmatrix} 1 & 2 & -1 & 1 \\ 2 & 7 & 4 & 2 \\ -1 & 4 & 13 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & -1 & 1 \\ 0 & 3 & 6 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = LU.$$ ▲

EXAMPLE 6

We reiterate that the LU decomposition exists only when no row interchanges are required to reduce the matrix to echelon form. For example, the matrix

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

has no such expression. See Exercise 14. ▲

We shall see in Chapter 3 that, given the LU decomposition of a matrix A, we can read off a great deal of information. But the main reason it is of interest is this: To solve $A\mathbf{x} = \mathbf{b}$ for different vectors $\mathbf{b}$ using computers, it is significantly more cost-effective to use the LU decomposition (see Exercise 13). Notice that $A\mathbf{x} = \mathbf{b}$ if and only if $(LU)\mathbf{x} = L(U\mathbf{x}) = \mathbf{b}$, so first we solve $L\mathbf{y} = \mathbf{b}$ (by "forward substitution") and then we solve $U\mathbf{x} = \mathbf{y}$ (by "back substitution"). Actually, working by hand, it is even easier to determine L^{-1}, which is the product of elementary matrices that puts A in echelon form ($L^{-1}A = U$), so then we find $\mathbf{y} = L^{-1}\mathbf{b}$ and solve $U\mathbf{x} = \mathbf{y}$ as before.

ercises 2.4

*1. For each of the matrices A in Exercise 1.4.3, find a product of elementary matrices $E = \cdots E_2E_1$ so that EA is in echelon form. Use the matrix E you've found to give constraint equations for $A\mathbf{x} = \mathbf{b}$ to be consistent.

*2. For each of the matrices A in Exercise 1.4.3, use the method of Example 3 to find a matrix E so that $EA = U$, where U is in echelon form.

*3. Give the LU decomposition (when it exists) of each of the matrices A in Exercise 1.4.3.

*4. Let $A = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & -1 & 1 & 0 \\ 0 & 1 & -2 & -1 & 1 \\ 2 & 1 & 0 & 0 & 5 \end{bmatrix}$.

a. Give the LU decomposition of A.

b. Give the reduced echelon form of A.

5. Find a left inverse of each of the following matrices A using the method of Example 4.

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

6. Given $A = \begin{bmatrix} 1 & & \\ -1 & 1 & \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 \\ & 3 & 1 \\ & & -1 \end{bmatrix}$, solve $A\mathbf{x} = \mathbf{b}$, where

*a. $\mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$ b. $\mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$ *c. $\mathbf{b} = \begin{bmatrix} 5 \\ -1 \\ 4 \end{bmatrix}$

7. Show that the inverse of every elementary matrix is again an elementary matrix. Indeed, give a simple prescription for determining the inverse of each type of elementary matrix. (See the proof of Theorem 4.1 of Chapter 1.)

8. Prove or give a counterexample: Every invertible matrix can be written as a product of elementary matrices.

9. Use elementary matrices to prove Theorem 4.1 of Chapter 1.

10. *a. Suppose E_1 and E_2 are elementary matrices that correspond to adding multiples of the *same* row to other rows. Show that $E_1E_2 = E_2E_1$ and give a simple description of the product. Explain how to use this observation to compute the LU decomposition more efficiently.

b. In a similar vein, let $i < j$, $i < k$, and $j < \ell$. Let E_1 be an elementary matrix corresponding to adding a multiple of row i to row k, and let E_2 be an elementary matrix corresponding to adding a multiple of row j to row ℓ. Give a simple description of the product E_1E_2, and explain how to use this observation to compute the LU decomposition more efficiently. Does $E_2E_1 = E_1E_2$ this time?

11. Complete the following alternative argument that the matrix obtained by Gaussian elimination must be the inverse matrix of A. It thereby provides another proof of Corollary 3.3. Suppose A is nonsingular.

a. Show that there are finitely many elementary matrices $E_1, E_2, \dots, E_k$ so that $E_kE_{k-1}\cdots E_2E_1A = I$.

b. Let $B = E_kE_{k-1}\cdots E_2E_1$. Apply Proposition 3.4 to show that $A = B^{-1}$ and, thus, that $AB = I$.

12. Assume A and B are two $m \times n$ matrices with the same reduced echelon form. Show that there exists an invertible matrix E so that $EA = B$. Is the converse true?

13. We saw in Exercise 1.4.17 that it takes on the order of $n^3/3$ multiplications to put an $n \times n$ matrix in reduced echelon form (and, hence, to solve a square inhomogeneous system $A\mathbf{x} = \mathbf{b}$). Indeed, in solving that exercise, one shows that it takes on the order of $n^3/3$ multiplications to obtain U (and one obtains L just by bookkeeping). Show now that if one has different vectors $\mathbf{b}$ for which one wishes to solve $A\mathbf{x} = \mathbf{b}$, once one has $A = LU$, it takes on the order of n^2 multiplications to solve for $\mathbf{x}$ each time.

14. a. Show that the matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ has no LU decomposition.

b. Show that for any $m \times n$ matrix A, there is an $m \times m$ permutation matrix P so that PA does have an LU decomposition.

5 The Transpose

The final matrix operation we discuss in this chapter is the *transpose*. When A is an $m \times n$ matrix with entries a_{ij}, the matrix A^{T} (read "A transpose") is the $n \times m$ matrix whose ij-entry is a_{ji}; in other words, the i^{th} *row* of A^{T} is the i^{th} *column* of A. We say a square matrix A is *symmetric* if $A^{\mathsf{T}} = A$ and is *skew-symmetric* if $A^{\mathsf{T}} = -A$.

EXAMPLE 1

Suppose

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 3 & -1 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 3 \\ 2 & -1 \\ 1 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}, \quad \text{and} \quad D = \begin{bmatrix} 1 & 2 & -3 \end{bmatrix}.$$

Then $A^{\mathsf{T}} = B$, $B^{\mathsf{T}} = A$, $C^{\mathsf{T}} = D$, and $D^{\mathsf{T}} = C$. Note, in particular, that the transpose of a column vector, i.e., an $n \times 1$ matrix, is a row vector, i.e., a $1 \times n$ matrix. An example of a symmetric matrix is

$$S = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & -1 \\ 3 & -1 & 7 \end{bmatrix}, \quad \text{since} \quad S^{\mathsf{T}} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & -1 \\ 3 & -1 & 7 \end{bmatrix} = S.$$

▲

The basic properties of the transpose operation are as follows:

Proposition 5.1. *Let A and A' be $m \times n$ matrices, let B be an $n \times p$ matrix, and let c be a scalar. Then*

1. $(A^{\mathsf{T}})^{\mathsf{T}} = A$.
2. $(cA)^{\mathsf{T}} = cA^{\mathsf{T}}$.
3. $(A + A')^{\mathsf{T}} = A^{\mathsf{T}} + A'^{\mathsf{T}}$.
4. $(AB)^{\mathsf{T}} = B^{\mathsf{T}}A^{\mathsf{T}}$.
5. *When A is invertible, then so is A^{T}, and $(A^{\mathsf{T}})^{-1} = (A^{-1})^{\mathsf{T}}$.*

Proof. The first is obvious, inasmuch as we swap rows and columns and then swap again, returning to our original matrix. The second and third are immediate to check. The fourth result is more interesting, and we will use it to derive a crucial result in a moment. To prove **4**, note, first, that AB is an $m \times p$ matrix, so $(AB)^{\mathsf{T}}$ will be a $p \times m$ matrix; $B^{\mathsf{T}}A^{\mathsf{T}}$ is the product of a $p \times n$ matrix and an $n \times m$ matrix and hence will be $p \times m$ as well, so the shapes agree. Now, the ji-entry of AB is the dot product of the j^{th} row vector of A and the i^{th} column vector of B, i.e., the ij-entry of $(AB)^{\mathsf{T}}$ is

$$\left((AB)^{\mathsf{T}}\right)_{ij} = (AB)_{ji} = \mathbf{A}_j \cdot \mathbf{b}_i.$$

On the other hand, the ij-entry of $B^{\mathsf{T}}A^{\mathsf{T}}$ is the dot product of the i^{th} row vector of B^{T} and the j^{th} column vector of A^{T}; but this is, by definition, the dot product of the i^{th} column vector of B and the j^{th} row vector of A. That is,

$$(B^{\mathsf{T}}A^{\mathsf{T}})_{ij} = \mathbf{b}_i \cdot \mathbf{A}_j,$$

and, since dot product is commutative, the two formulas agree. The proof of **5** is left to Exercise 8. □

The transpose matrix will be important to us because of the interplay between dot product and transpose. If $\mathbf{x}$ and $\mathbf{y}$ are vectors in $\mathbb{R}^n$, then by virtue of our very definition of matrix multiplication,

$$\boxed{\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^{\mathsf{T}}\mathbf{y},}$$

provided we agree to think of a 1×1 matrix as a scalar. (On the right-hand side we are multiplying a $1 \times n$ matrix by an $n \times 1$ matrix.) Now we have this highly useful proposition:

Proposition 5.2. *Let A be an $m \times n$ matrix, $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{y} \in \mathbb{R}^m$. Then*

$$A\mathbf{x} \cdot \mathbf{y} = \mathbf{x} \cdot A^{\mathsf{T}}\mathbf{y}.$$

(On the left, we take the dot product of vectors in $\mathbb{R}^m$; on the right, of vectors in $\mathbb{R}^n$.)

Remark. You might remember this: To move the matrix "across the dot product," you must transpose it.

Proof. We just calculate, using the formula for the transpose of a product and, as usual, associativity: □

$$A\mathbf{x} \cdot \mathbf{y} = (A\mathbf{x})^{\mathsf{T}}\mathbf{y} = (\mathbf{x}^{\mathsf{T}}A^{\mathsf{T}})\mathbf{y} = \mathbf{x}^{\mathsf{T}}(A^{\mathsf{T}}\mathbf{y}) = \mathbf{x} \cdot A^{\mathsf{T}}\mathbf{y}.$$

EXAMPLE 2

We return to the economic interpretation of dot product given in the Remark on p. 25. Suppose that m different ingredients are required to manufacture n different products. To manufacture the product vector $\mathbf{x} = (x_1, \ldots, x_n)$ requires the ingredient vector $\mathbf{y} =$

$(y_1, \ldots, y_m)$, and we suppose that $\mathbf{x}$ and $\mathbf{y}$ are related by the equation $\mathbf{y} = A\mathbf{x}$ for some $m \times n$ matrix A. If each unit of ingredient j costs a price p_j, then the cost of producing $\mathbf{x}$ is

$$\sum_{j=1}^{m} p_j y_j = \mathbf{y} \cdot \mathbf{p} = A\mathbf{x} \cdot \mathbf{p} = \mathbf{x} \cdot A^{\mathsf{T}}\mathbf{p} = \sum_{i=1}^{n} q_i x_i,$$

where $\mathbf{q} = A^{\mathsf{T}}\mathbf{p}$. Notice then that q_i is the amount it costs to produce a unit of the i^{th} product. Our fundamental formula, Proposition 5.2, tells us that the total cost of the ingredients should equal the total worth of the products we manufacture. See Exercise 18 for a less abstract (but more fattening) example. ▲

EXAMPLE 3

We just saw that when $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, the matrix product $\mathbf{x}^{\mathsf{T}}\mathbf{y}$ is a 1×1 matrix. However, when we switch the position of the transpose and calculate $\mathbf{x}\mathbf{y}^{\mathsf{T}}$, the result is an $n \times n$ matrix (see Exercise 13). A particularly important application of this has arisen already in Chapter 1. Given a vector $\mathbf{a} \in \mathbb{R}^n$, consider the $n \times n$ matrix $A = \mathbf{a}\mathbf{a}^{\mathsf{T}}$. What does it mean? That is, what is the associated linear transformation μ_A? Well, by the associativity of multiplication, we have $A\mathbf{x} = (\mathbf{a}\mathbf{a}^{\mathsf{T}})\mathbf{x} = \mathbf{a}(\mathbf{a}^{\mathsf{T}}\mathbf{x}) = (\mathbf{a} \cdot \mathbf{x})\mathbf{a}$. When $\mathbf{a}$ is a unit vector, this is the projection of $\mathbf{x}$ onto $\mathbf{a}$. And, in general, we can now write

$$\operatorname{proj}_{\mathbf{a}}\mathbf{x} = \frac{\mathbf{x} \cdot \mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a} = \frac{\mathbf{a} \cdot \mathbf{x}}{\|\mathbf{a}\|^2}\mathbf{a} = \left(\frac{1}{\|\mathbf{a}\|^2}\mathbf{a}\mathbf{a}^{\mathsf{T}}\right)\mathbf{x}.$$

We will see the importance of this formulation in Chapter 4. ▲

Exercises 2.5

1. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, $B = \begin{bmatrix} 2 & 1 \\ 4 & 3 \end{bmatrix}$, $C = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$, and $D = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 2 & 3 \end{bmatrix}$. Calculate each of the following expressions or explain why it is not defined.

a. A^{T} d. $C^{\mathsf{T}} + D$ *g. $C^{\mathsf{T}}A^{\mathsf{T}}$ *j. CC^{T}
*b. $2A - B^{\mathsf{T}}$ *e. $A^{\mathsf{T}}C$ h. BD^{T} *k. $C^{\mathsf{T}}C$
c. C^{T} f. AC^{T} i. $D^{\mathsf{T}}B$ l. $C^{\mathsf{T}}D^{\mathsf{T}}$

2. Let $\mathbf{a} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 0 \\ 3 \\ -1 \end{bmatrix}$. Calculate the following matrices.

*a. $\mathbf{a}\mathbf{a}^{\mathsf{T}}$ c. $\mathbf{b}^{\mathsf{T}}\mathbf{b}$ e. $\mathbf{a}\mathbf{b}^{\mathsf{T}}$ g. $\mathbf{b}^{\mathsf{T}}\mathbf{a}$
*b. $\mathbf{a}^{\mathsf{T}}\mathbf{a}$ d. $\mathbf{b}\mathbf{b}^{\mathsf{T}}$ *f. $\mathbf{a}^{\mathsf{T}}\mathbf{b}$ h. $\mathbf{b}\mathbf{a}^{\mathsf{T}}$

3. Following Example 3, find the standard matrix for the projection $\operatorname{proj}_{\mathbf{a}}$.

a. $\mathbf{a} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ *b. $\mathbf{a} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$ c. $\mathbf{a} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ d. $\mathbf{a} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$

♯**4.** Suppose $\mathbf{a}, \mathbf{b}, \mathbf{c}$, and $\mathbf{d} \in \mathbb{R}^n$. Check that, surprisingly,

$$\begin{bmatrix} | & | \\ \mathbf{a} & \mathbf{b} \\ | & | \end{bmatrix} \begin{bmatrix} — & \mathbf{c}^{\mathsf{T}} & — \\ — & \mathbf{d}^{\mathsf{T}} & — \end{bmatrix} = \mathbf{a}\mathbf{c}^{\mathsf{T}} + \mathbf{b}\mathbf{d}^{\mathsf{T}}.$$

***5.** Suppose A and B are symmetric. Show that AB is symmetric if and only if $AB = BA$.

6. Let A be an arbitrary $m \times n$ matrix. Show that $A^{\mathsf{T}}A$ is symmetric.

7. Explain why the matrix $A^{\mathsf{T}}A$ is a diagonal matrix whenever the column vectors of A are orthogonal to one another.

♯**8.** Suppose A is invertible. Check that $(A^{-1})^{\mathsf{T}}A^{\mathsf{T}} = I$ and $A^{\mathsf{T}}(A^{-1})^{\mathsf{T}} = I$, and deduce that A^{T} is likewise invertible with inverse $(A^{-1})^{\mathsf{T}}$.

9. If P is a permutation matrix (see Exercise 2.1.12 for the definition), show that $P^{\mathsf{T}} = P^{-1}$.

10. Suppose $A = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{bmatrix}$. Check that the vector $\mathbf{y} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ satisfies $A\mathbf{y} = \mathbf{y}$ and $A^{\mathsf{T}}\mathbf{y} = \mathbf{y}$. Show that if $\mathbf{x} \cdot \mathbf{y} = 0$, then $A\mathbf{x} \cdot \mathbf{y} = 0$ as well. Interpret this result geometrically.

11. Let A be an $m \times n$ matrix and let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Prove that if $A\mathbf{x} = \mathbf{0}$ and $\mathbf{y} = A^{\mathsf{T}}\mathbf{b}$ for some $\mathbf{b} \in \mathbb{R}^m$, then $\mathbf{x} \cdot \mathbf{y} = 0$.

♯**12.** Suppose A is a symmetric $n \times n$ matrix. If $\mathbf{x}$ and $\mathbf{y} \in \mathbb{R}^n$ are vectors satisfying the equations $A\mathbf{x} = 2\mathbf{x}$ and $A\mathbf{y} = 3\mathbf{y}$, show that $\mathbf{x}$ and $\mathbf{y}$ are orthogonal. (*Hint:* Consider $A\mathbf{x} \cdot \mathbf{y}$.)

13. Suppose A is an $m \times n$ matrix with rank 1. Prove that there are nonzero vectors $\mathbf{u} \in \mathbb{R}^m$ and $\mathbf{v} \in \mathbb{R}^n$ such that $A = \mathbf{u}\mathbf{v}^{\mathsf{T}}$. (*Hint:* What do the rows of $\mathbf{u}\mathbf{v}^{\mathsf{T}}$ look like?)

14. Given the matrix

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 3 & 1 \\ 0 & 1 & -1 \end{bmatrix} \quad \text{and its inverse matrix} \quad A^{-1} = \begin{bmatrix} 4 & -3 & 1 \\ -1 & 1 & 0 \\ -1 & 1 & -1 \end{bmatrix}.$$

By thinking about rows and columns of these matrices, find the inverse of

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

♯***15.** Suppose A is an $m \times n$ matrix and $\mathbf{x} \in \mathbb{R}^n$ satisfies $(A^{\mathsf{T}}A)\mathbf{x} = \mathbf{0}$. Prove that $A\mathbf{x} = \mathbf{0}$. (*Hint:* What is $\|A\mathbf{x}\|$?)

16. Suppose A is a symmetric matrix satisfying $A^2 = \mathrm{O}$. Show that $A = \mathrm{O}$. Give an example to show that the hypothesis of symmetry is required.

***17.** Let A_θ be the rotation matrix defined on p. 98. Using geometric reasoning, explain why $A_\theta^{-1} = A_\theta^{\mathsf{T}}$.

18. (With thanks to Maida Heatter for approximate and abbreviated recipes) To make 8 dozen *David's cookies* requires 1 lb. semisweet chocolate, 1 lb. butter, 2 c. sugar, 2 eggs, and 4 c. flour. To make 8 dozen *chocolate chip oatmeal cookies* requires 3/4 lb. semisweet chocolate, 1 lb. butter, 3 c. sugar, 2 eggs, 2 1/2 c. flour, and 6 c. oats. With the following approximate prices, what is the cost per dozen for each cookie? Use the approach of Example 2; what are the matrices A and A^{T}?

Item	Cost
1 lb. chocolate	\$4.80
1 lb. butter	3.40
1 c. sugar	0.20
1 dozen eggs	1.40
1 c. flour	0.10
1 c. oats	0.20

♯**19.** We say an $n \times n$ matrix A is *orthogonal* if $A^{\mathsf{T}}A = I_n$.

a. Prove that the column vectors $\mathbf{a}_1, \dots, \mathbf{a}_n$ of an orthogonal matrix A are unit vectors that are orthogonal to one another, i.e.,

$$\mathbf{a}_i \cdot \mathbf{a}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}.$$

b. Fill in the missing columns in the following matrices to make them orthogonal:

$$\begin{bmatrix} \frac{\sqrt{3}}{2} & ? \\ -\frac{1}{2} & ? \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & ? \\ 0 & -1 & ? \\ 0 & 0 & ? \end{bmatrix}, \quad \begin{bmatrix} \frac{1}{3} & ? & \frac{2}{3} \\ \frac{2}{3} & ? & -\frac{2}{3} \\ \frac{2}{3} & ? & \frac{1}{3} \end{bmatrix}$$

c. Show that any 2×2 orthogonal matrix A must be of the form

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$$

for some real number θ. (*Hint:* Use part *a*, rather than the original definition.)

*d. Show that if A is an orthogonal 2×2 matrix, then $\mu_A : \mathbb{R}^2 \to \mathbb{R}^2$ is either a rotation or the composition of a rotation and a reflection.

e. Prove that the row vectors $\mathbf{A}_1, \dots, \mathbf{A}_n$ of an orthogonal matrix A are unit vectors that are orthogonal to one another. (*Hint:* Corollary 3.3.)

20. (Recall the definition of orthogonal matrices from Exercise 19.)

a. Show that if A and B are orthogonal $n \times n$ matrices, then so is AB.

*b. Show that if A is an orthogonal matrix, then so is A^{-1}.

21. Here is an alternative argument that when A is square and $AB = I$, it must be the case that $BA = I$ and so $B = A^{-1}$.

a. Suppose $AB = I$. Prove that A^{T} is nonsingular. (*Hint:* Solve $A^{\mathsf{T}}\mathbf{x} = \mathbf{0}$.)

b. Prove there exists a matrix C so that $A^{\mathsf{T}}C = I$, and hence $C^{\mathsf{T}}A = I$.

c. Use the result of part c of Exercise 2.1.11 to prove that $B = A^{-1}$.

♯**22.** a. Show that the only matrix that is both symmetric and skew-symmetric is O.

b. Given any square matrix A, show that $S = \frac{1}{2}(A + A^{\mathsf{T}})$ is symmetric and $K = \frac{1}{2}(A - A^{\mathsf{T}})$ is skew-symmetric.

c. Deduce that any square matrix A can be written in the form $A = S + K$, where S is symmetric and K is skew-symmetric.