2

MATRIX ALGEBRA

In Chapter 1 we used matrices and vectors as simple storage devices. In this chapter matrices and vectors take on a life of their own. We develop the arithmetic of matrices and vectors. Much of what we do is motivated by a desire to extend the ideas of ordinary arithmetic to matrices. Our notational style of writing a matrix in the form $A = [a_{ij}]$ hints that a matrix could be treated like a single number. What if we could manipulate equations with matrix and vector quantities in the same way that we do equations with scalars? We shall see that this is a useful idea. Matrix arithmetic gives us new powers for formulating and solving practical problems. In this chapter we will use it to find effective methods for solving linear and nonlinear systems, solve problems of graph theory and analyze an important modeling tool of applied mathematics called a Markov chain.

2.1 Matrix Addition and Scalar Multiplication

To begin our discussion of arithmetic we consider the matter of equality of matrices. Suppose that A and B represent two matrices. When do we declare them to be equal? The answer is, of course, if they represent the same matrix! Thus we expect that all the usual laws of equalities will hold (e.g., equals may be substituted for equals) and in fact, they do. There are times, however, when we need to prove that two symbolic matrices are equal. For this purpose, we need something a little more precise. So we have the following definition, which includes vectors as a special case of matrices.

Equality of Matrices

Definition 2.1. Two matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ are said to be *equal* if these matrices have the same size, and for each index pair (i, j), $a_{ij} = b_{ij}$, that is, corresponding entries of A and B are equal.

Example 2.1. Which of the following matrices are equal, if any?

(a) $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ (b) $\begin{bmatrix} 0 & 0 \end{bmatrix}$ (c) $\begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix}$ (d) $\begin{bmatrix} 0 & 1 \\ 1-1 & 1+1 \end{bmatrix}$

Solution. The answer is that only (c) and (d) have any chance of being equal, since they are the only matrices in the list with the same size (2×2). As a matter of fact, an entry-by-entry check verifies that they really are equal. $\square$

Matrix Addition and Subtraction

How should we define addition or subtraction of matrices? We take a clue from elementary two- and three-dimensional vectors, such as the type we would encounter in geometry or calculus. There, in order to add two vectors, one condition has to hold: the vectors have to be the same size. If they are the same size, we simply add the vectors coordinate by coordinate to obtain a new vector of the same size. That is precisely what the following definition does.

Matrix Addition and Subtraction

Definition 2.2. Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be $m \times n$ matrices. Then the *sum* of the matrices, denoted by $A + B$, is the $m \times n$ matrix defined by the formula

$$A + B = [a_{ij} + b_{ij}] .$$

The *negative* of the matrix A, denoted by $-A$, is defined by the formula

$$-A = [-a_{ij}] .$$

Finally, the *difference* of A and B, denoted by $A - B$, is defined by the formula

$$A - B = [a_{ij} - b_{ij}] .$$

Notice that matrices must be the same size before we attempt to add them. We say that two such matrices or vectors are *conformable for addition.*

Example 2.2. Let

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

Find $A + B$, $A - B$, and $-A$.

Solution. Here we see that

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

Likewise,

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

The negative of A is even simpler:

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

Scalar Multiplication

The next arithmetic concept we want to explore is that of scalar multiplication. Once again, we take a clue from the elementary vectors, where the idea behind scalar multiplication is simply to "scale" a vector a certain amount by multiplying each of its coordinates by that amount. That is what the following definition says.

Definition 2.3. Let $A = [a_{ij}]$ be an $m \times n$ matrix and c a scalar. Then the *product* of the scalar c with the matrix A, denoted by cA, is defined by the formula

Scalar Multiplication

$$cA = [ca_{ij}] \,.$$

Recall that the default scalars are real numbers, but they could also be complex numbers.

Example 2.3. Let

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

Find cA, $0A$, and $-1A$.

Solution. Here we see that

$$cA = 3 \begin{bmatrix} 3 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 \cdot 3 & 3 \cdot 1 & 3 \cdot 0 \\ 3 \cdot -2 & 3 \cdot 0 & 3 \cdot 1 \end{bmatrix} = \begin{bmatrix} 9 & 3 & 0 \\ -6 & 0 & 3 \end{bmatrix},$$

while

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

and

$$(-1)\, A = (-1) \begin{bmatrix} 3 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -3 & -1 & 0 \\ 2 & 0 & -1 \end{bmatrix} = -A.$$ □

Linear Combinations

Now that we have a notion of scalar multiplication and addition, we can blend these two ideas to yield a very fundamental notion in linear algebra, that of a *linear combination*.

Definition 2.4. A *linear combination* of the matrices $A_1, A_2, \ldots, A_n$ is an expression of the form

Linear Combinations

$$c_1 A_1 + c_2 A_2 + \cdots + c_n A_n$$

where $c_1, c_2, \ldots, c_n$ are scalars and $A_1, A_2, \ldots, A_n$ are matrices all of the same size.

Example 2.4. Given that

$$A_1 = \begin{bmatrix} 2 \\ 6 \\ 4 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 2 \\ 4 \\ 2 \end{bmatrix}, \quad \text{and} \quad A_3 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix},$$

compute the linear combination $-2A_1 + 3A_2 - 2A_3$.

Solution. The solution is that

$$\begin{aligned} -2A_1 + 3A_2 - 2A_3 &= -2 \begin{bmatrix} 2 \\ 6 \\ 4 \end{bmatrix} + 3 \begin{bmatrix} 2 \\ 4 \\ 2 \end{bmatrix} - 2 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \\ &= \begin{bmatrix} -2 \cdot 2 + 3 \cdot 2 - 2 \cdot 1 \\ -2 \cdot 6 + 3 \cdot 4 - 2 \cdot 0 \\ -2 \cdot 4 + 3 \cdot 2 - 2 \cdot (-1) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \end{aligned}$$ □

It seems like too much work to write out objects such as the vector $(0, 0, 0)$ that occurred in the last equation; after all, we know that all the entries are all 0. So we make the following notational convention for convenience. A *zero matrix* is a matrix whose every entry is 0. We shall denote such matrices by the symbol 0.

Zero Matrix

Caution: This convention makes the symbol 0 ambiguous, but the meaning of the symbol will be clear from context, and the convenience gained is worth the potential ambiguity. For example, the equation of the preceding example is stated very simply as $-2A_1 + 3A_2 - 2A_3 = 0$, where we understand from context that 0 has to mean the 3×1 column vector of zeros. If we use boldface for vectors, we will also then use boldface for the vector zero, so some distinction is regained.

Example 2.5. Suppose that a linear combination of matrices satisfies the identity $-2A_1 + 3A_2 - 2A_3 = 0$, as in the preceding example. Use this fact to express A_1 in terms of A_2 and A_3.

Solution. To solve this example, just forget that the quantities A_1, A_2, A_3 are anything special and use ordinary algebra. First, add $-3A_2 + 2A_3$ to both sides to obtain

$$-2A_1 + 3A_2 - 2A_3 - 3A_2 + 2A_3 = -3A_2 + 2A_3,$$

so that

$$-2A_1 = -3A_2 + 2A_3,$$

and multiplying both sides by the scalar $-\frac{1}{2}$ yields the identity

$$A_1 = \frac{-1}{2}(-2A_1) = \frac{-1}{2}(-3A_2 + 2A_3) = \frac{3}{2}A_2 - A_3.$$ □

The linear combination idea has a really useful application to linear systems, namely, it gives us another way to express the solution set of a linear system that clearly identifies the role of free variables. The following example illustrates this point.

Example 2.6. Suppose that a linear system in the unknowns x_1, x_2, x_3, x_4 has general solution $(x_2 + 3x_4, x_2, 2x_2 - x_4, x_4)$, where the variables x_2, x_4 are free. Describe the solution set of this linear system in terms of linear combinations with free variables as coefficients.

Solution. The trick here is to use only the parts of the general solution involving x_2 for one vector and the parts involving x_4 as the other vectors in such a way that these vectors add up to the general solution. In our case we have

$$\begin{bmatrix} x_2 + 3x_4 \\ x_2 \\ 2x_2 - x_4 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_2 \\ x_2 \\ 2x_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 3x_4 \\ 0 \\ -x_4 \\ x_4 \end{bmatrix} = x_2 \begin{bmatrix} 1 \\ 1 \\ 2 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 3 \\ 0 \\ -1 \\ 1 \end{bmatrix}.$$

Now simply define vectors $A_1 = (1, 1, 2, 0)$, $A_2 = (3, 0, -1, 1)$, and we see that since x_2 and x_4 are arbitrary, the solution set is

$$S = \{x_2 A_1 + x_4 A_2 \mid x_2, x_4 \in \mathbb{R}\}.$$

In other words, the solution set to the system is the set of all possible linear combinations of the vectors A_1 and A_2. □

The idea of solution sets as linear combinations is an important one that we will return to in later chapters. You might notice that once we have the general form of a solution vector we can see that there is an easier way to determine the constant vectors A_1 and A_2. Simply set $x_2 = 1$ and the other free variable(s) equal to zero—in this case just x_4—to get the solution vector A_1, and set $x_4 = 1$ and $x_2 = 0$ to get the solution vector A_2.

Laws of Arithmetic

The last example brings up an important point: to what extent can we rely on the ordinary laws of arithmetic and algebra in our calculations with matrices and vectors? For matrix *multiplication* there are some surprises. On the other hand, the laws for addition and scalar multiplication are pretty much what we would expect them to be. Here are the laws with their customary names. These same names can apply to more than one operation. For instance, there is a closure law for addition and one for scalar multiplication as well.

Laws of Matrix Addition and Scalar Multiplication

Let A, B, C be matrices of the same size $m \times n$, 0 the $m \times n$ zero matrix, and c and d scalars.

(1) (Closure Law) $A + B$ is an $m \times n$ matrix.
(2) (Associative Law) $(A + B) + C = A + (B + C)$
(3) (Commutative Law) $A + B = B + A$
(4) (Identity Law) $A + 0 = A$
(5) (Inverse Law) $A + (-A) = 0$
(6) (Closure Law) cA is an $m \times n$ matrix.
(7) (Associative Law) $c(dA) = (cd)A$
(8) (Distributive Law) $(c + d)A = cA + dA$
(9) (Distributive Law) $c(A + B) = cA + cB$
(10) (Monoidal Law) $1A = A$

It is fairly straightforward to prove from definitions that these laws are valid. The verifications all follow a similar pattern, which we illustrate by verifying the commutative law for addition: let $A = [a_{ij}]$ and $B = [b_{ij}]$ be given $m \times n$ matrices. Then we have that

$$\begin{aligned} A + B &= [a_{ij} + b_{ij}] \\ &= [b_{ij} + a_{ij}] \\ &= B + A, \end{aligned}$$

where the first and third equalities come from the definition of matrix addition, and the second equality follows from the fact that for all indices i and j, $a_{ij} + b_{ij} = b_{ij} + a_{ij}$ by the commutative law for addition of scalars.

2.1 Exercises and Problems

Exercise 1. Calculate the following where possible.

(a) $\begin{bmatrix} 1 & 2 & -1 \\ 0 & 2 & 2 \end{bmatrix} - \begin{bmatrix} 3 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ (b) $2\begin{bmatrix} 1 \\ 3 \end{bmatrix} - 5\begin{bmatrix} 2 \\ 2 \end{bmatrix} + 3\begin{bmatrix} 4 \\ 1 \end{bmatrix}$ (c) $2\begin{bmatrix} 1 & 4 \\ 0 & 0 \end{bmatrix} + 3\begin{bmatrix} 0 & 0 \\ 2 & 1 \end{bmatrix}$

(d) $a\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + b\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ (e) $\begin{bmatrix} 1 & 2 & -1 \\ 0 & 0 & 2 \\ 0 & 2 & -2 \end{bmatrix} + 2\begin{bmatrix} 3 & 1 & 0 \\ 5 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ (f) $x\begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} + y\begin{bmatrix} 4 \\ 1 \\ 0 \end{bmatrix}$

Exercise 2. Calculate the following where possible.

(a) $8\begin{bmatrix} 1 & 2 & -1 \\ 1 & 0 & 0 \\ 2 & -1 & 3 \end{bmatrix}$ (b) $-\begin{bmatrix} 2 \\ 3 \end{bmatrix} + 3\begin{bmatrix} 2 \\ -1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 4 & 2 \\ 1 & 0 & 3 \end{bmatrix} + (-4)\begin{bmatrix} 0 & 0 & 1 \\ 2 & 1 & -2 \end{bmatrix}$

(d) $4\begin{bmatrix} 0 & 1 & -1 \\ 2 & 0 & 2 \\ 0 & 2 & 0 \end{bmatrix} - 2\begin{bmatrix} 0 & 2 & 0 \\ -3 & 0 & 1 \\ 1 & -2 & 0 \end{bmatrix}$ (e) $2\begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} + u\begin{bmatrix} -2 \\ 2 \\ 3 \end{bmatrix} + v\begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$

Exercise 3. Let $A = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 1 & 2 \end{bmatrix}$, $B = \begin{bmatrix} 2 & 2 \\ 1 & -2 \end{bmatrix}$, $C = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 1 & 0 \end{bmatrix}$, and compute the following, where possible.
(a) $A + 3B$ (b) $2A - 3C$ (c) $A - C$ (d) $6B + C$ (e) $2C - 3(A - 2C)$

Exercise 4. With A, B, C as in Exercise 3, solve for the unknown matrix X in the equations
(a) $X + 3A = C$ (b) $A - 3X = 3C$ (c) $2X + \begin{bmatrix} 2 & 2 \\ 1 & -2 \end{bmatrix} = B$.

Exercise 5. Write the following vectors as a linear combination of constant vectors with scalar coefficients x, y, or z.
(a) $\begin{bmatrix} x + 2y \\ 2x - z \end{bmatrix}$ (b) $\begin{bmatrix} x - y \\ 2x + 3y \end{bmatrix}$ (c) $\begin{bmatrix} 3x + 2y \\ -z \\ x + y + 5z \end{bmatrix}$ (d) $\begin{bmatrix} x - 3y \\ 4x + z \\ 2y - z \end{bmatrix}$

Exercise 6. Write the following vectors as a linear combination of constant vectors with scalar coefficients x, y, z, or w.
(a) $\begin{bmatrix} 3x + y \\ x + y + z \end{bmatrix}$ (b) $\begin{bmatrix} 3x + 2y - w \\ w - z \\ x + y - 2w \end{bmatrix}$ (c) $\begin{bmatrix} x + 3y \\ 2y - x \end{bmatrix}$ (d) $\begin{bmatrix} x - 2y \\ 4x + z \\ 3w - z \end{bmatrix}$

Exercise 7. Find scalars a, b, c such that

$$\begin{bmatrix} c & b \\ 0 & c \end{bmatrix} = \begin{bmatrix} a - b & c + 2 \\ a + b & a - b \end{bmatrix}.$$

Exercise 8. Find scalars a, b, c, d such that

$$\begin{bmatrix} d & 2a \\ 2d & a \end{bmatrix} = \begin{bmatrix} a - b & b + c \\ a + b & c - b + 1 \end{bmatrix}.$$

Exercise 9. Express the matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ as a linear combination of the four matrices $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$, and $\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$.

Exercise 10. Express the matrix $D = \begin{bmatrix} 3 & 3 \\ 1 & -3 \end{bmatrix}$ as a linear combination of the matrices $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$, and $C = \begin{bmatrix} 0 & 2 \\ 0 & -1 \end{bmatrix}$.

Exercise 11. Verify that the associative law and commutative laws for addition hold for

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

Exercise 12. Verify that both distributive laws for addition hold for $c = 2$, $d = -3$, and A, B, and C as in Exercise 11.

Problem 13. Show by examples that it is false that for arbitrary matrices A and B, and constant c,
(a) $\operatorname{rank}(cA) = \operatorname{rank} A$ (b) $\operatorname{rank}(A + B) \geq \operatorname{rank} A + \operatorname{rank} B$.

Problem 14. Prove that the associative law for addition of matrices holds.

Problem 15. Prove that both distributive laws hold.

*Problem 16. Prove that if A and B are matrices such that $2A - 4B = 0$ and $A + 2B = I$, then $A = \frac{1}{2}I$.

Problem 17. Prove the following assertions for $m \times n$ matrices A and B by using the laws of matrix addition and scalar multiplication. Clearly specify each law that you use.
(a) If $A = -A$, then $A = 0$.
(b) If $cA = 0$ for some scalar c, then either $c = 0$ or $A = 0$.
(c) If $B = cB$ for some scalar $c \neq 1$, then $B = 0$.

2.2 Matrix Multiplication

Matrix multiplication is somewhat more subtle than matrix addition and scalar multiplication. Of course, we could define matrix multiplication to be a coordinatewise operation, just as addition is (there is such a thing, called Hadamard multiplication). But our motivation is not merely to make definitions, but rather to make *useful* definitions for basic problems.

Definition of Multiplication

To motivate the definition, let us consider a single linear equation

$$2x - 3y + 4z = 5.$$

We will find it handy to think of the left-hand side of the equation as a "product" of the coefficient matrix $[2, -3, 4]$ and the column matrix of unknowns $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$. Thus, we have that the product of this row and column is

$$[2, -3, 4] \begin{bmatrix} x \\ y \\ z \end{bmatrix} = [2x - 3y + 4z].$$

Notice that we have made the result of the product into a 1×1 matrix. This introduces us to a permanent abuse of notation that is almost always used in linear algebra: we don't distinguish between the scalar a and the 1×1 matrix $[a]$, though technically perhaps we should. In the same spirit, we make the following definition.

Definition 2.5. The *product* of the $1 \times n$ row $[a_1, a_2, \ldots, a_n]$ with the $n \times 1$ column $\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$ is defined to be the 1×1 matrix $[a_1 b_1 + a_2 b_2 + \cdots + a_n b_n]$.

Row Column Product

It is this row-column product strategy that guides us to the general definition. Notice how the column number of the first matrix had to match the row number of the second, and that this number disappears in the size of the resulting product. This is exactly what happens in general.

Definition 2.6. Let $A = [a_{ij}]$ be an $m \times p$ matrix and $B = [b_{ij}]$ a $p \times n$ matrix. Then the *product* of the matrices A and B, denoted by AB, is the $m \times n$ matrix whose (i, j)th entry, for $1 \leq i \leq m$ and $1 \leq j \leq n$, is the entry of the product of the ith row of A and the jth column of B; more specifically, the (i, j)th entry of AB is

Matrix Product

$$a_{i1} b_{1j} + a_{i2} b_{2j} + \cdots + a_{ip} b_{pj}.$$

Notice that, in contrast to the case of addition, two matrices may be of different sizes when we can multiply them together. If A is $m \times p$ and B is $p \times n$, we say that A and B are *conformable* for multiplication. It is also worth noticing that if A and B are square *and of the same size,* then the products AB and BA are always defined.

Some Illustrative Examples

Let's check our understanding with a few examples.

Example 2.7. Compute, if possible, the products AB of the following pairs of matrices A, B.

(a) $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & -1 \end{bmatrix}, \begin{bmatrix} 4 & -2 \\ 0 & 1 \\ 2 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

(d) $\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 2 \end{bmatrix}$ (e) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & -1 \end{bmatrix}$ (f) $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$

Solution. In part (a) A is 2×3 and B is 3×2. First check conformability for multiplication. Stack these dimensions alongside each other and see that the 3's match; now "cancel" the matching middle 3's to obtain that the dimension of the product is $2 \times \cancel{3}\ \cancel{3} \times 2 = 2 \times 2$. To obtain, for example, the $(1, 2)$th entry of the product matrix, multiply the first row of A and second column of B to obtain

$$[1, 2, 1] \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = [1 \cdot (-2) + 2 \cdot 1 + 1 \cdot 1] = [1] \, .$$

The full product calculation looks like this:

$$\begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & -1 \end{bmatrix} \begin{bmatrix} 4 & -2 \\ 0 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 \cdot 4 + 2 \cdot 0 + 1 \cdot 2 & 1 \cdot (-2) + 2 \cdot 1 + 1 \cdot 1 \\ 2 \cdot 4 + 3 \cdot 0 + (-1) \cdot 2 & 2 \cdot (-2) + 3 \cdot 1 + (-1) \cdot 1 \end{bmatrix}$$
$$= \begin{bmatrix} 6 & 1 \\ 6 & -2 \end{bmatrix}.$$

A size check of part (b) reveals a mismatch between the column number of the first matrix (3) and the row number (2) of the second matrix. Thus these matrices are *not conformable* for multiplication in the specified order. Hence

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

is undefined.

In part (c) a size check shows that the product has size $2 \times \cancel{1} \ \cancel{1} \times 2 = 2 \times 2$. The calculation gives

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} 0 \cdot 1 & 0 \cdot 2 \\ 0 \cdot 1 & 0 \cdot 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

For part (d) the size check shows gives $1 \times \cancel{2} \ \cancel{2} \times 1 = 1 \times 1$. Hence the product exists and is 1×1. The calculation gives

$$\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = [1 \cdot 0 + 2 \cdot 0] = [0].$$

Something very interesting comes out of parts (c) and (d). Notice that AB and BA are *not* the same matrices—never mind that their entries are all 0's—the important point is that these matrices are not even the same size! Thus a very familiar law of arithmetic, the commutativity of multiplication, has just fallen by the wayside.

Matrix Multiplication Not Commutative or Cancellative

Things work well in (e), where the size check gives $2 \times \cancel{2} \ \cancel{2} \times 3 = 2 \times 3$ as the size of the product. As a matter of fact, this is a rather interesting calculation:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & -1 \end{bmatrix} = \begin{bmatrix} 1 \cdot 1 + 0 \cdot 2 & 1 \cdot 2 + 0 \cdot 3 & 1 \cdot 1 + 0 \cdot (-1) \\ 0 \cdot 1 + 1 \cdot 2 & 0 \cdot 2 + 1 \cdot 3 & 0 \cdot 1 + 1 \cdot (-1) \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & -1 \end{bmatrix}.$$

Notice that we end up with the second matrix in the product. This is similar to the arithmetic fact that $1 \cdot x = x$ for a given real number x. So the matrix on the left acted like a multiplicative identity. We'll see that this is no accident.

Finally, for the calculation in (f), notice that

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

There's something very curious here, too. Notice that two nonzero matrices of the same size multiplied together to give a zero matrix. This kind of thing never happens in ordinary arithmetic, where the cancellation law assures that if $a \cdot b = 0$ then $a = 0$ or $b = 0$. □

The calculation in (e) inspires some more notation. The left-hand matrix of this product has a very important property. It acts like a "1" for matrix multiplication. So it deserves its own name. A matrix of the form

Identity Matrix

$$I_n = \begin{bmatrix} 1 & 0 & \dots & & 0 \\ 0 & 1 & 0 & & \\ \vdots & & \ddots & & \\ & & & 1 & 0 \\ 0 & & & 0 & 1 \end{bmatrix} = [\delta_{ij}]$$

is called an $n \times n$ *identity matrix*. The (i, j)th entry of I_n is designated by the Kronecker symbol δ_{ij}, which is 1 if $i = j$ and 0 otherwise. If n is clear from context, we simply write I in place of I_n.

Kronecker Symbol

So we see in the previous example that the left-hand matrix of part (e) is

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2.$$

Linear Systems as a Matrix Product

Let's have another look at a system we examined in Chapter 1. We'll change the names of the variables from x, y, z to x_1, x_2, x_3 in anticipation of a notation that will work with any number of variables.

Example 2.8. Express the following linear system as a matrix product:

$$\begin{aligned} x_1 + x_2 + x_3 &= 4 \\ 2x_1 + 2x_2 + 5x_3 &= 11 \\ 4x_1 + 6x_2 + 8x_3 &= 24 \end{aligned}$$

Solution. Recall how we defined multiplication of a row vector and column vector at the beginning of this section. We use that as our inspiration. Define

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 4 \\ 11 \\ 24 \end{bmatrix}, \quad \text{and } A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 5 \\ 4 & 6 & 8 \end{bmatrix}.$$

Of course, A is just the coefficient matrix of the system and b is the right-hand-side vector, which we have seen several times before. But now these take on a new significance. Notice that if we take the first row of A and multiply it by $\mathbf{x}$ we get the left-hand side of the first equation of our system. Likewise for the second and third rows. Therefore, we may write in the language of matrices that

$$A\mathbf{x} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 5 \\ 4 & 6 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 11 \\ 24 \end{bmatrix} = \mathbf{b}.$$

Thus the system is represented very succinctly as $A\mathbf{x} = \mathbf{b}$. □

Once we understand this example, it is easy to see that the general abstract system that we examined in Section 1.1 can just as easily be abbreviated. Now we have a new way of looking at a system of equations: it is just like a simple first-degree equation in one variable. Of course, the catch is that the symbols $A, \mathbf{x}, \mathbf{b}$ now represent an $m \times n$ matrix, and $n \times 1$ and $m \times 1$ vectors, respectively. In spite of this, the matrix multiplication idea is very appealing. For instance, it might inspire us to ask whether we could somehow solve the system $A\mathbf{x} = \mathbf{b}$ by multiplying both sides of the equation by some kind of matrix "$1/A$" so as to cancel the A and get

$$(1/A)A\mathbf{x} = I\mathbf{x} = \mathbf{x} = (1/A)\mathbf{b}.$$

We'll follow up on this idea in Section 2.5.

Here is another perspective on matrix–vector multiplication that gives a powerful way of thinking about such multiplications.

Example 2.9. Interpret the matrix product of Example 2.8 as a linear combination of column vectors.

Solution. Examine the system of this example and we see that the column $(1, 2, 4)$ appears to be multiplied by x_1. Similarly, the column $(1, 2, 6)$ is multiplied by x_2 and the column $(1, 5, 8)$ by x_3. Hence, if we use the same right-hand-side column $(4, 11, 24)$ as before, we obtain that this column can be expressed as a linear combination of column vectors, namely

$$x_1 \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 2 \\ 6 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 5 \\ 8 \end{bmatrix} = \begin{bmatrix} 4 \\ 11 \\ 24 \end{bmatrix}.$$ □

Matrix-Vector Multiplication

We could write the equation of the previous example very succinctly as follows: let A have columns $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$, so that $A = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3]$, and let $\mathbf{x} = (x_1, x_2, x_3)$. Then

$$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + x_3\mathbf{a}_3.$$

This formula extends to general matrix–vector multiplication. It is extremely useful in interpreting such products, so we will elevate its status to that of a theorem worth remembering.

Theorem 2.1. Let $A = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n]$ be an $m \times n$ matrix with columns $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ and let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Then

$$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n.$$

Laws of Arithmetic

We have already seen that the laws of matrix arithmetic may not be quite the same as the ordinary arithmetic laws that we are used to. Nonetheless, as long as we don't assume a cancellation law or a commutative law for multiplication, things are pretty much what one might expect.

Laws of Matrix Multiplication

> Let A, B, C be matrices of the appropriate sizes so that the following multiplications make sense, I a suitably sized identity matrix, and c and d scalars.
> (1) (Closure Law) The product AB is a matrix.
> (2) (Associative Law) $(AB)C = A(BC)$
> (3) (Identity Law) $AI = A$ and $IB = B$
> (4) (Associative Law for Scalars) $c(AB) = (cA)B = A(cB)$
> (5) (Distributive Law) $(A + B)C = AC + BC$
> (6) (Distributive Law) $A(B + C) = AB + AC$

One can formally verify these laws by working through the definitions. For example, to verify the first half of the identity law, let $A = [a_{ij}]$ be an $m \times n$ matrix, so that $I = [\delta_{ij}]$ has to be I_n in order for the product AI to make sense. Now we see from the formal definition of matrix multiplication that

$$AI = \left[\sum_{k=1}^{n} a_{ik}\delta_{kj}\right] = [a_{ij} \cdot 1] = A.$$

The middle equality follows from the fact that δ_{kj} is 0 unless $k = j$. Thus the sum collapses to a single term. A similar calculation verifies the other laws.

We end our discussion of matrix multiplication with a familiar-looking notation that will prove to be extremely handy in the sequel. This notation applies only to *square* matrices. Let A be a square $n \times n$ matrix and k a nonnegative integer. Then we define the *kth power* of A to be

Exponent Notation

$$A^k = \begin{cases} I_n & \text{if } k = 0, \\ \underbrace{A \cdot A \cdots A}_{k \text{ times}} & \text{if } k > 0. \end{cases}$$

As a simple consequence of this definition we have the standard exponent laws.

Laws of Exponents

> For nonnegative integers i, j and square matrix A:
> (1) $A^{i+j} = A^i \cdot A^j$
> (2) $A^{ij} = (A^i)^j$

Notice that the law $(AB)^i = A^i B^i$ is missing. It won't work with matrices. Why not? The following example illustrates a very useful application of the exponent notation.

Example 2.10. Let $f(x) = 1 - 2x + 3x^2$ be a polynomial function. Use the definition of matrix powers to derive a sensible interpretation of $f(A)$, where A is a square matrix. Evaluate $f\left(\left[\begin{smallmatrix}2 & -1\\ 0 & 1\end{smallmatrix}\right]\right)$ explicitly with this interpretation.

Solution. Let's take a closer look at the polynomial expression

$$f(x) = 1 - 2x + 3x^2 = 1x^0 - 2x + 3x^2.$$

Once we've rewritten the polynomial in this form, we recall that $A^0 = I$ and that other matrix powers make sense since A is square, so the interpretation is easy:

$$f(A) = A^0 - 2A^1 + 3A^2 = I - 2A + 3A^2.$$

In particular, for a 2×2 matrix we take $A = I = \left[\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right]$ and obtain

$$\begin{aligned}
f\left(\begin{bmatrix}2 & -1\\ 0 & 1\end{bmatrix}\right) &= I - 2\begin{bmatrix}2 & -1\\ 0 & 1\end{bmatrix} + 3\begin{bmatrix}2 & -1\\ 0 & 1\end{bmatrix}^2 \\
&= \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} - 2\begin{bmatrix}2 & -1\\ 0 & 1\end{bmatrix} + 3\begin{bmatrix}2 & -1\\ 0 & 1\end{bmatrix}\begin{bmatrix}2 & -1\\ 0 & 1\end{bmatrix} \\
&= \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} - \begin{bmatrix}4 & -2\\ 0 & 2\end{bmatrix} + \begin{bmatrix}12 & -9\\ 0 & 3\end{bmatrix} = \begin{bmatrix}9 & -7\\ 0 & 2\end{bmatrix}.
\end{aligned}$$

□

2.2 Exercises and Problems

Exercise 1. Carry out these calculations or indicate they are impossible, given that $\mathbf{a} = \begin{bmatrix}2\\ 1\end{bmatrix}$, $\mathbf{b} = \begin{bmatrix}3 & 4\end{bmatrix}$, and $C = \begin{bmatrix}2 & 1+\mathrm{i}\\ 0 & -1\end{bmatrix}$.
(a) $\mathbf{b}C\mathbf{a}$ (b) $\mathbf{ab}$ (c) $C\mathbf{b}$ (d) $(\mathbf{a}C)\,\mathbf{b}$ (e) $C\mathbf{a}$ (f) $C(\mathbf{ab})$ (g) $\mathbf{ba}$ (h) $C(\mathbf{a}+\mathbf{b})$

Exercise 2. For each pair of matrices A, B, calculate the product AB or indicate that the product is undefined.

(a) $\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, \begin{bmatrix}3 & -2 & 0\\ -2 & 5 & 8\end{bmatrix}$ (b) $\begin{bmatrix}2 & 1 & 0\\ 0 & 8 & 2\end{bmatrix}, \begin{bmatrix}1 & 1\\ 2 & 2\end{bmatrix}$ (c) $\begin{bmatrix}3 & 1 & 2\\ 1 & 0 & 0\\ 4 & 3 & 2\end{bmatrix}, \begin{bmatrix}-5 & 4 & -2\\ -2 & 3 & 1\\ 1 & 0 & 4\end{bmatrix}$

(d) $\begin{bmatrix}3 & 1\\ 1 & 0\\ 4 & 3\end{bmatrix}, \begin{bmatrix}-5 & 4 & -2\\ -2 & 3 & 1\end{bmatrix}$ (e) $\begin{bmatrix}3\\ 1\\ 4\end{bmatrix}, \begin{bmatrix}-5 & 4\\ -2 & 3\end{bmatrix}$ (f) $\begin{bmatrix}2 & 0\\ 2 & 3\end{bmatrix}, \begin{bmatrix}3\\ 1\end{bmatrix}$

Exercise 3. Express these systems of equations in the notation of matrix multiplication and as a linear combination of vectors as in Example 2.8.

(a) $\begin{aligned} x_1 - 2x_2 + 4x_3 &= 3\\ x_2 - x_3 &= 2\\ -x_1 + 4x_4 &= 1 \end{aligned}$ (b) $\begin{aligned} x - y - 3z &= 3\\ 2x + 2y + 4z &= 10\\ -x + z &= 3 \end{aligned}$ (c) $\begin{aligned} x - 3y + 1 &= 0\\ 2y &= 0\\ -x + 3y &= 0 \end{aligned}$

Exercise 4. Express these systems of equations in the notation of matrix multiplication and as a linear combination of vectors as in Example 2.8.

(a) $\begin{aligned} x_1 + x_3 &= -1 \\ x_2 + x_3 &= 0 \\ x_1 + x_3 &= 1 \end{aligned}$ (b) $\begin{aligned} x - y - 3z &= 1 \\ z &= 0 \\ -x + y &= 3 \end{aligned}$ (c) $\begin{aligned} x - 4y &= 0 \\ 2y &= 0 \\ -x + 3y &= 0 \end{aligned}$

Exercise 5. Let $A = \begin{bmatrix} 2 & -1 & 1 \\ 2 & 3 & -2 \\ 4 & 2 & -2 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 2 \\ -3 \\ 1 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$, and $\mathbf{X} = \begin{bmatrix} x & 0 & 0 \\ 0 & y & 0 \\ 0 & 0 & z \end{bmatrix}$.

Find the coefficient matrix of the linear system $\mathbf{X}A\mathbf{b} + A\mathbf{x} = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$ in the variables x, y, z.

Exercise 6. Let $A = \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix}$ and $X = \begin{bmatrix} x & y \\ z & w \end{bmatrix}$. Find the coefficient matrix of the linear system $AX - XA = I_2$ in the variables x, y, z, w.

Exercise 7. Let $\mathbf{u} = (1, 1, 0)$, $\mathbf{v} = (0, 1, 1)$, and $\mathbf{w} = (1, 3, 1)$. Write each of the following expressions as single matrix product.
(a) $2\mathbf{u} - 4\mathbf{v} - 3\mathbf{w}$ (b) $\mathbf{w} - \mathbf{v} + 2i\mathbf{u}$ (c) $x_1\mathbf{u} - 3x_2\mathbf{v} + x_3\mathbf{w}$

Exercise 8. Express the following matrix products as linear combinations of vectors.

(a) $\begin{bmatrix} 2 & 1 \\ 0 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ -5 \\ 1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 1 \\ 1 & 1 + i \end{bmatrix} \begin{bmatrix} x_1 \\ -x_2 \end{bmatrix}$

Exercise 9. Let $A = \begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix}$, $f(x) = 1 + x + x^2$, $g(x) = 1 - x$, and $h(x) = 1 - x^3$. Verify that $f(A)\, g(A) = h(A)$.

Exercise 10. Let $A = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \frac{5}{2} & -\frac{3}{2} & 0 \end{bmatrix}$. Compute $f(A)$ and $f(B)$, where $f(x) = 2x^3 + 3x - 5$.

Exercise 11. Find all possible products of two matrices from among the following:

$A = \begin{bmatrix} 1 & -2 \\ 1 & 3 \end{bmatrix}$ $B = \begin{bmatrix} 2 & 4 \end{bmatrix}$ $C = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$ $D = \begin{bmatrix} 1 & 3 & 0 \\ -1 & 2 & 1 \end{bmatrix}$

Exercise 12. Find all possible products of three matrices from among the following:

$A = \begin{bmatrix} -1 & 2 \\ 0 & 2 \end{bmatrix}$ $B = \begin{bmatrix} 2 & 1 \\ 1 & 0 \\ 2 & 3 \end{bmatrix}$ $C = \begin{bmatrix} -3 \\ 2 \end{bmatrix}$ $D = \begin{bmatrix} 2 & 3 & -1 \\ 1 & 2 & 1 \end{bmatrix}$ $E = \begin{bmatrix} -2 & 4 \end{bmatrix}$

Exercise 13. A square matrix A is said to be *nilpotent* if there is a positive integer k such that $A^k = 0$. Determine which of the following matrices are nilpotent. (You may assume that if A is $n \times n$ nilpotent, then $A^n = 0$.)

(a) $\begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ (d) $\begin{bmatrix} 2 & 2 & -4 \\ -1 & 0 & 2 \\ 1 & 1 & -2 \end{bmatrix}$ (e) $\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ -1 & 0 & -2 & -1 \end{bmatrix}$

Exercise 14. A square matrix A is *idempotent* if $A^2 = A$. Determine which of the following matrices are idempotent.

(a) $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix}$ (d) $\begin{bmatrix} 0 & 0 & 2 \\ 1 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$ (e) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix}$

Exercise 15. Show by example that a sum of nilpotent matrices need not be nilpotent.

Exercise 16. Show by example that a product of idempotent matrices need not be idempotent.

Exercise 17. Verify that the product $\mathbf{uv}$, where $\mathbf{u} = (1, 0, 2)$ and $\mathbf{v} = \begin{bmatrix} -1 & 1 & 1 \end{bmatrix}$, is a rank-one matrix.

Exercise 18. Verify that the product $\mathbf{uv} + \mathbf{wu}$, where $\mathbf{u} = (1, 0, 2)$, $\mathbf{v} = \begin{bmatrix} -1 & 1 & 1 \end{bmatrix}$, and $\mathbf{w} = (1, 0, 1)$, is a matrix of rank at most two.

Exercise 19. Verify that both associative laws of multiplication hold for

$$c = 4, \qquad A = \begin{bmatrix} 2 & 0 \\ -1 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 2 \\ 0 & 3 \end{bmatrix}, \qquad C = \begin{bmatrix} 1+\mathrm{i} & 1 \\ 1 & 2 \end{bmatrix}.$$

Exercise 20. Verify that both distributive laws of multiplication hold for

$$A = \begin{bmatrix} 2 & 0 \\ -1 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 2 \\ 0 & 3 \end{bmatrix}, \qquad C = \begin{bmatrix} 1+\mathrm{i} & 1 \\ 1 & 2 \end{bmatrix}.$$

Problem 21. Find examples of 2×2 matrices A and B that fulfill each of the following conditions.

(a) $(AB)^2 \neq A^2B^2$ (b) $AB \neq BA$

Problem 22. Find examples of nonzero 2×2 matrices A, B, and C that fulfill each of the following conditions.

(a) $A^2 = 0,\ B^2 = 0$ (b) $(AB)^2 \neq 0$

*Problem 23. Show that if A is a 2×2 matrix such that $AB = BA$ for every 2×2 matrix B, then A is a multiple of I_2.

Problem 24. Prove that the associative law for scalars is valid.

Problem 25. Prove that both distributive laws for matrix multiplication are valid.

Problem 26. Show that if A is a square matrix such that $A^{k+1} = 0$, then

$$(I - A)\left(I + A + A^2 + \cdots + A^k\right) = I.$$

*Problem 27. Show that if two matrices A and B of the same size have the property that $A\mathbf{b} = B\mathbf{b}$ for every column vector $\mathbf{b}$ of the correct size for multiplication, then $A = B$.

Problem 28. Determine the flop count for multiplication of $m \times p$ matrix A by $p \times n$ matrix B. (See page 48.)

2.3 Applications of Matrix Arithmetic

We next examine a few more applications of the matrix multiplication idea that should reinforce the importance of this idea and provide us with some interpretations of matrix multiplication.

Matrix Multiplication as Function

The function idea is basic to mathematics. Recall that a *function* f is a rule of correspondence that assigns to each argument x in a set called its domain, a unique value $y = f(x)$ from a set called its target. Each branch of mathematics has its own special functions; for example, in calculus differentiable functions $f(x)$ are fundamental. Linear algebra also has its special functions. Suppose that $T(\mathbf{u})$ represents a function whose arguments $\mathbf{u}$ and values $\mathbf{v} = T(\mathbf{u})$ are vectors.

We say that the function T is *linear* if T preserves linear combinations, that is, for all vectors $\mathbf{u}, \mathbf{v}$ in the domain of T, and scalars c, d, we have that $c\mathbf{u} + d\mathbf{v}$ is in the domain of T and

Linear Functions

$$T\left(c\mathbf{u} + d\mathbf{v}\right) = cT\left(\mathbf{u}\right) + dT\left(\mathbf{v}\right).$$

Example 2.11. Show that the function T, whose domain is the set of 2×1 vectors and definition is given by

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = x,$$

is a linear function.

Solution. Let (x,y) and (z,w) be two elements in the domain of T and c,d any two scalars. Now compute

$$\begin{aligned} T\left(c\begin{bmatrix} x \\ y \end{bmatrix} + d\begin{bmatrix} z \\ w \end{bmatrix}\right) &= T\left(\begin{bmatrix} cx \\ cy \end{bmatrix} + \begin{bmatrix} dz \\ dw \end{bmatrix}\right) = T\left(\begin{bmatrix} cx+dz \\ cy+dw \end{bmatrix}\right) \\ &= cx + dz = cT\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) + dT\left(\begin{bmatrix} z \\ w \end{bmatrix}\right). \end{aligned}$$

Thus, T satisfies the definition of linear function. □

One can check that the function T just defined can be expressed as a matrix multiplication, namely, $T\left(\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]\right) = \left[\begin{smallmatrix} 1 & 0 \end{smallmatrix}\right]\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$. This example gives yet another reason for defining matrix multiplication in the way that we do. Here is a general definition for these kinds of functions (also known as linear transformations or linear operators).

Matrix Operator

Definition 2.7. Let A be an $m \times n$ matrix. The function T_A that maps $n \times 1$ vectors to $m \times 1$ vectors according to the formula

$$T_A(\mathbf{u}) = A\mathbf{u}$$

is called the *linear function (operator or transformation)* associated with the matrix A or simply a *matrix operator*.

Let's verify that this function T actually is linear. Use the definition of T_A along with the distributive law of multiplication and associative law for scalars to obtain that

$$\begin{aligned} T_A(c\mathbf{u} + d\mathbf{v}) &= A(c\mathbf{u} + d\mathbf{v}) \\ &= A(c\mathbf{u}) + A(d\mathbf{v}) \\ &= c(A\mathbf{u}) + d(A\mathbf{v}) \\ &= cT_A(\mathbf{u}) + dT_A(\mathbf{v}). \end{aligned}$$

Thus multiplication of vectors by a fixed matrix A is a linear function. Notice that this result contains Example 2.11 as a special case.

Function Composition Notation

Recall that the composition of functions f and g is the function $f \circ g$ whose definition is $(f \circ g)(x) = f(g(x))$ for all x in the domain of g.

Example 2.12. Use the associative law of matrix multiplication to show that the composition of matrix multiplication functions corresponds to the matrix product.

Solution. For all vectors $\mathbf{u}$ and for suitably sized matrices A, B, we have by the associative law that $A(B\mathbf{u}) = (AB)\mathbf{u}$. In function terms, this means that $T_A(T_B(\mathbf{u})) = T_{AB}(\mathbf{u})$. Since this is true for all arguments $\mathbf{u}$, it follows that $T_A \circ T_B = T_{AB}$, which is what we were to show. □

We will have more to say about linear functions in Chapters 3 and 6, where they will go by the name of linear operators. Here is an example that gives another slant on why the "linear" in "linear function."

Example 2.13. Describe the action of the matrix operator T_A on the x-axis and y-axis, where $A = \begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix}$.

Solution. A typical element of the x-axis has the form $\mathbf{v} = (x, 0)$. Thus we have that $T(\mathbf{v}) = T((x, 0))$. Now calculate

$$T(\mathbf{v}) = T_A((x, 0)) = A\mathbf{v} = \begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} x \\ 0 \end{bmatrix} = \begin{bmatrix} 2x \\ 4x \end{bmatrix} = x \begin{bmatrix} 2 \\ 4 \end{bmatrix}.$$

Thus the x-axis is mapped to all multiples of the vector $(2, 4)$. Set $t = 2x$, and we see that $x(2, 4) = (t, 2t)$. Hence, these are simply points on the line given by $x = t$, $y = 2t$. Equivalently, this is the line $y = 2x$. Similarly, one checks that the y-axis is mapped to the line $y = 2x$ as well. □

Example 2.14. Let L be set of points (x, y) defined by the equation $y = x+1$ and let $T_A(L) = \{T(((x, y)) \mid (x, y) \in L\}$, where $A = \begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix}$. Describe and sketch these sets in the plane.

Solution. Of course, the set L is just the straight line defined by the linear equation $y = x + 1$. To see what $T_A(L)$ looks like, write a typical element of L in the form $(x, x + 1)$. Now calculate

$$T_A((x, x + 1)) = \begin{bmatrix} 2 & 1 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} x \\ x + 1 \end{bmatrix} = \begin{bmatrix} 3x + 1 \\ 6x + 2 \end{bmatrix}.$$

Next make the substitution $t = 3x + 1$, and we see that a typical element of $T_A(L)$ has the form $(t, 2t)$, where t is any real number. We recognize these points as exactly the points on the line $y = 2x$. Thus, the function T_A maps the line $y = x + 1$ to the line $y = 2x$. Figure 2.1 illustrates this mapping as well as the fact that T_A maps the line segment from $\left(\frac{-1}{3}, \frac{2}{3}\right)$ to $\left(\frac{1}{6}, \frac{7}{6}\right)$ on L to the line segment from $(0, 0)$ to $\left(\frac{3}{2}, 3\right)$ on $T_A(L)$. □

Graphics specialists and game programmers have a special interest in *real-time rendering*, the discipline concerned with algorithms that create synthetic images fast enough that the viewer can interact with a virtual environment. For a comprehensive treatment of this subject, consult the text [2]. A number of fundamental matrix-defined operators are used in real-time rendering, where they are called *transforms*. Here are a few examples of such operators. A *scaling operator* is effected by multiplying each coordinate of a point by a fixed (positive) scale factor. A *shearing operator* is effected by adding a constant shear factor times one coordinate to another coordinate of the point. A *rotation operator* is effected by rotating each point a fixed angle θ in the counterclockwise direction about the origin.

Real-Time Rendering

Scaling and Shearing Graphics Transforms

Example 2.15. Let the scaling operator S on points in two dimensions have scale factors of $\frac{3}{2}$ in the x-direction and $\frac{1}{2}$ in the y-direction. Let the shearing

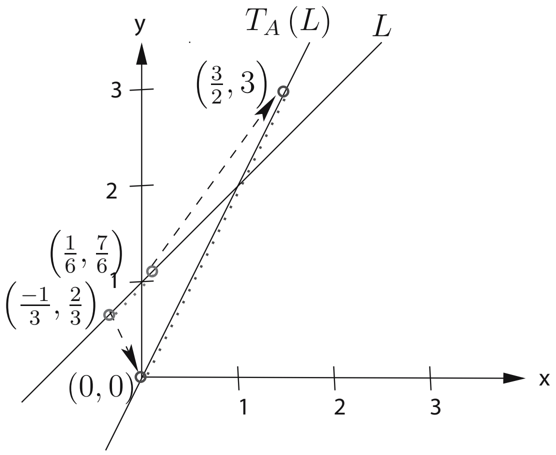


Fig. 2.1. Action of T_A on line L given by $y = x + 1$, points on L, and the segment between them.

operator H on these points have a shear factor of $\frac{1}{2}$ by the y-coordinate on the x-coordinate. Express these operators as matrix operators and graph their action on four unit squares situated diagonally from the origin.

Solution. First consider the scaling operator. The point (x, y) will be transformed into the point $\left(\frac{3}{2}x, \frac{1}{2}y\right)$. Observe that

$$S((x, y)) = \begin{bmatrix} \frac{3}{2}x \\ \frac{1}{2}y \end{bmatrix} = \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = T_A((x, y)),$$

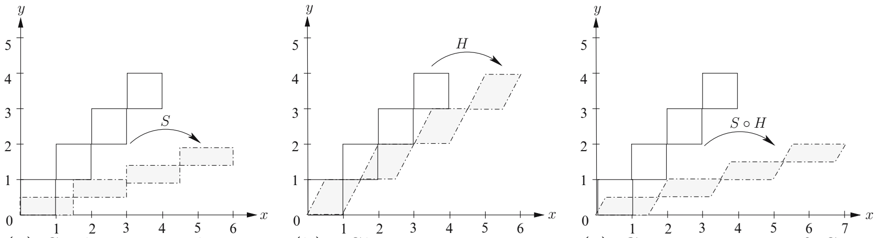
where $A = \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & frac12 \end{bmatrix}$. Similarly, the shearing operator transforms the point (x, y) into the point $\left(x + \frac{1}{2}y, y\right)$. Thus we have

$$H((x, y)) = \begin{bmatrix} x + \frac{1}{2}y \\ y \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = T_B((x, y)),$$

where $B = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{bmatrix}$. The action of these operators on four unit squares is illustrated in Figure 2.2. □

Example 2.16. Express the concatenation $S \circ H$ of the scaling operator S and shearing operator H of Example 2.15 as a matrix operator and graph the action of the concatenation on four unit squares situated diagonally from the origin.

Solution. From Example 2.15 we have that $S = T_A$, where $A = \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$, and $H = T_B$, where $B = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{bmatrix}$. From Example 2.12 we know that function composition corresponds to matrix multiplication, that is,



(a) Scaling in x-direction by $\frac{3}{2}$, y-direction by 1 (b) Shearing in x-direction by y, shear factor $\frac{1}{2}$ (c) Concatenation of S and H

Fig. 2.2. Action of scaling operator, shearing operator, and concatenation.

$$\begin{aligned} S \circ H\left((x, y)\right) &= T_A \circ T_B\left((x, y)\right) = T_{AB}\left((x, y)\right) \\ &= \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\ &= \begin{bmatrix} \frac{3}{2} & \frac{3}{4} \\ 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = T_C\left((x, y)\right), \end{aligned}$$

where $C = AB = \begin{bmatrix} \frac{3}{2} & \frac{3}{4} \\ 0 & \frac{1}{2} \end{bmatrix}$. The action of $S \circ H$ on four unit squares is illustrated in Figure 2.2. $\square$

Example 2.17. Describe the rotation operator (about the origin) for the plane.

Solution. Consult Figure 2.3. Observe that if the point (x, y) is given by $(r\cos\phi, r\sin\phi)$ in polar coordinates, then the rotated point (x', y') has coordinates $(r\cos(\theta+\phi), r\sin(\theta+\phi))$. Now use the double-angle formula for angles and obtain that

$$\begin{aligned} \begin{bmatrix} x' \\ y' \end{bmatrix} &= \begin{bmatrix} r\cos(\theta+\phi) \\ r\sin(\theta+\phi) \end{bmatrix} = \begin{bmatrix} r\cos\theta\cos\phi - r\sin\theta\sin\phi \\ r\sin\theta\cos\phi + r\cos\theta\sin\phi \end{bmatrix} \\ &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} r\cos\phi \\ r\sin\phi \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \end{aligned}$$

Now define the *rotation matrix* $R(\theta)$ by Rotation Matrix

$$R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

It follows that $(x', y') = T_{R(\theta)}\left((x, y)\right)$. $\square$

Discrete Dynamical Systems

Discrete dynamical systems are an extremely useful modeling tool in a wide variety of disciplines. Here is the definition of such a system.

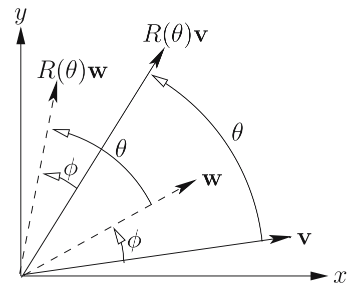


Fig. 2.3. Action of rotation matrix $R(\theta)$ on vectors $\mathbf{v}$ and $\mathbf{w}$.

Discrete Dynamical System

Definition 2.8. A *discrete linear dynamical system* is a sequence of vectors $\mathbf{x}^{(k)}$, $k = 0, 1, \ldots$, called *states*, which is defined by an initial vector $\mathbf{x}^{(0)}$ and by the rule

$$\mathbf{x}^{(k+1)} = A\mathbf{x}^{(k)}, \qquad k = 0, 1, \ldots,$$

where A is a given fixed square matrix, called the *transition matrix* of the system.

A Markov chain is a certain type of discrete dynamical system. Here is an example.

Example 2.18. Suppose two toothpaste companies compete for customers in a fixed market in which each customer uses either Brand A or Brand B. Suppose also that a market analysis shows that the buying habits of the customers fit the following pattern in the quarters that were analyzed: each quarter (three-month period), 30% of A users will switch to B, while the rest stay with A. Moreover, 40% of B users will switch to A in a given quarter, while the remaining B users will stay with B. If we *assume* that this pattern does not vary from quarter to quarter, we have an example of what is called a Markov *chain model.* Express the data of this model in matrix–vector language.

Solution. Notice that if a_0 and b_0 are the fractions of the customers using A and B, respectively, in a given quarter, a_1 and b_1 the fractions of customers using A and B in the next quarter, then our hypotheses say that

$$\begin{aligned} a_1 &= 0.7a_0 + 0.4b_0 \\ b_1 &= 0.3a_0 + 0.6b_0. \end{aligned}$$

We could figure out what happens in the quarter after this by replacing the indices 1 and 0 by 2 and 1, respectively, in the preceding formula. In general, we replace the indices $1, 0$ by $k, k+1$, to obtain

$$\begin{aligned} a_{k+1} &= 0.7a_k + 0.4b_k \\ b_{k+1} &= 0.3a_k + 0.6b_k. \end{aligned}$$

We express this system in matrix form as follows: let

$$\mathbf{x}^{(k)} = \begin{bmatrix} a_k \\ b_k \end{bmatrix} \text{ and } A = \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix}.$$

Then the system may be expressed in the matrix form

$$\mathbf{x}^{(k+1)} = A\mathbf{x}^{(k)}.$$

□

The state vectors $\mathbf{x}^{(k)}$ of the preceding example have the following property: Seach coordinate is nonnegative and all the coordinates sum to 1. Such a vector is called a *probability distribution vector*. Also, the matrix A has the property that each of its columns is a probability distribution vector. Such a square matrix is called a *stochastic* matrix. In these terms we now give a precise definition of a Markov chain.

Probability Distribution Vector

Stochastic Matrix

Definition 2.9. A *Markov chain* is a discrete dynamical system whose initial state $\mathbf{x}^{(0)}$ is a probability distribution vector and whose transition matrix A is stochastic, that is, each column of A is a probability distribution vector.

Markov Chain

Let us return to Example 2.18. The state vectors and transition matrices

$$\mathbf{x}^{(k)} = \begin{bmatrix} a_k \\ b_k \end{bmatrix} \qquad \text{and} \qquad A = \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix}$$

should play an important role. And indeed they do, for in light of our interpretation of a linear system as a matrix product, we see that the two equations of Example 2.18 can be written simply as $\mathbf{x}^{(1)} = Ax^{(0)}$. A little more calculation shows that

$$\mathbf{x}^{(2)} = A\mathbf{x}^{(1)} = A \cdot (A\mathbf{x}^{(0)}) = A^2\mathbf{x}^{(0)}$$

and in general,

$$\mathbf{x}^{(k)} = A\mathbf{x}^{(k-1)} = A^2\mathbf{x}^{(k-2)} = \cdots = A^k\mathbf{x}^{(0)}.$$

In fact, this is true of any discrete dynamical system, and we record this as a *key fact:*

> For any positive integer k and discrete dynamical system with transition matrix A and initial state $\mathbf{x}^{(0)}$, the k-th state is given by
>
> $$\mathbf{x}^{(k)} = A^k\mathbf{x}^{(0)}.$$

Computing DDS States

Now we really have a very good handle on the Markov chain problem. Consider the following instance of our example.

Example 2.19. In the notation of Example 2.18 suppose that initially Brand A has all the customers (i.e., Brand B is just entering the market). What are the market shares 2 quarters later? 20 quarters? Answer the same questions if initially Brand B has all the customers.

Solution. To say that initially Brand A has all the customers is to say that the initial state vector is $\mathbf{x}^{(0)} = (1, 0)$. Now do the arithmetic to find $\mathbf{x}^{(2)}$:

$$\begin{aligned} \begin{bmatrix} a_2 \\ b_2 \end{bmatrix} = \mathbf{x}^{(2)} = A^2\mathbf{x}^{(0)} = & \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix} \left(\begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) \\ = & \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix} \begin{bmatrix} 0.7 \\ 0.3 \end{bmatrix} = \begin{bmatrix} .61 \\ .39 \end{bmatrix}. \end{aligned}$$

Thus, Brand A will have 61% of the market and Brand B will have 39% of the market in the second quarter. We did not try to do the next calculation by hand, but rather used a computer to get the approximate answer:

$$\mathbf{x}^{(20)} = \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix}^{20} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} .57143 \\ .42857 \end{bmatrix}.$$

Thus, after 20 quarters, Brand A's share will have fallen to about 57% of the market and Brand B's share will have risen to about 43%. Now consider what happens if the initial scenario is completely different, i.e., $\mathbf{x}^{(0)} = (0, 1)$. We compute by hand to find that

$$\begin{aligned} \mathbf{x}^{(2)} = & \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix} \left(\begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) \\ = & \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix} \begin{bmatrix} 0.4 \\ 0.6 \end{bmatrix} = \begin{bmatrix} .52 \\ .48 \end{bmatrix}. \end{aligned}$$

Then we use a computer to find that

$$\mathbf{x}^{(20)} = \begin{bmatrix} 0.7 & 0.4 \\ 0.3 & 0.6 \end{bmatrix}^{20} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} .57143 \\ .42857 \end{bmatrix}.$$

Surprise! For $k = 20$ we get the same answer as we did with a completely different initial condition. Coincidence? We will return to this example again in Chapters 3 and 5, where concepts introduced therein will cast new light on this model (no, it isn't a coincidence). Another curious feature of these state vectors: each one is a probability distribution vector. This is no coincidence either (see Problem 18). □

Structured Population Model

Another important type of model is a so-called *structured population model.* In such a model a population of organisms is divided into a finite number of disjoint states, such as age by year or weight by pound, so that the entire population is described by a state vector that represents the population at discrete times that occur at a constant period such as every day or year. A comprehensive development of this concept can be found in Hal Caswell's text [4]. Here is an example.

Example 2.20. A certain insect has three life stages: egg, juvenile, and adult. A population is observed in a certain environment to have the following properties in a two-day time span: 20% of the eggs will not survive, and 60% will move to the juvenile stage. In the same time-span 10% of the juveniles will not survive, and 60% will move to the adult stage, while 80% of the adults will survive. Also, in the same time-span adults will product about 0.25 eggs per adult. Assume that initially, there are 10, 8, and 6 eggs, juveniles, and adults (measured in thousands), respectively. Model this population as a discrete dynamical system and use it to compute the population total in 2, 10, and 100 days.

Solution. We start time at day 0 and the kth stage is day $2k$. Here the time period is two days and a state vector has the form $\mathbf{x}^{(k)} = (a_k, b_k, c_k)$, where a_k is the number of eggs, b_k the number of juveniles, and c_k the number of adults (all in thousands) on day $2k$. We are given that $\mathbf{x}^{(0)} = (10, 8, 6)$. Furthermore, the transition matrix has the form

$$A = \begin{bmatrix} 0.2 & 0 & 0.25 \\ 0.6 & 0.3 & 0 \\ 0 & 0.6 & 0.8 \end{bmatrix}.$$

The first column says that 20% of the eggs will remain eggs over one time period, 60% will progress to juveniles, and the rest do not survive. The second column says that juveniles produce no offspring, 30% will remain juveniles, 60% will become adults, and the rest do not survive. The third column says that .25 eggs results from one adult, no adult becomes a juvenile, and 80% survive. Now do the arithmetic to find the state $\mathbf{x}^{(1)}$ on day 2:

$$\mathbf{x}^{(1)} = \begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix} = A^1\mathbf{x}^{(0)} = \begin{bmatrix} 0.2 & 0 & 0.25 \\ 0.6 & 0.3 & 0 \\ 0 & 0.6 & 0.8 \end{bmatrix} \begin{bmatrix} 10 \\ 8 \\ 6 \end{bmatrix} = \begin{bmatrix} 3.5 \\ 8.4 \\ 9.6 \end{bmatrix}.$$

For the remaining calculations we use a computer (you should check these results with your own calculator or computer) to obtain approximate answers (we use $\approx$ for approximate equality)

$$\mathbf{x}^{(10)} = \begin{bmatrix} a_{10} \\ b_{10} \\ c_{10} \end{bmatrix} = A^{10}\mathbf{x}^{(0)} \approx \begin{bmatrix} 3.33 \\ 2.97 \\ 10.3 \end{bmatrix},$$

$$\mathbf{x}^{(100)} = \begin{bmatrix} a_{100} \\ b_{100} \\ c_{100} \end{bmatrix} = A^{100}\mathbf{x}^{(0)} \approx \begin{bmatrix} 0.284 \\ 0.253 \\ 0.877 \end{bmatrix}.$$

It appears that the population is declining with time. □

Calculating Power of Graph Vertices

Example 2.21. *(Dominance Directed Graphs)* You have incomplete data about four teams who have played each other in matches. Each match produces a winner and a loser, with no score attached. Identify the teams by labels $1, 2, 3,$ and 4. We could describe a match by a pair of numbers (i, j), where team i played and defeated team j (no ties allowed). Here are the given data:

$$\{(1,2),(1,4),(3,1),(2,3),(4,2)\}.$$

Give a reasonable graphical representation of these data.

Solution. We can draw a picture of all the data that we are given by representing each team as a point called a "vertex" and each match by connecting two points with an arrow, called a "directed edge," which points from the winner toward the loser in the match. See Figure 2.4. □

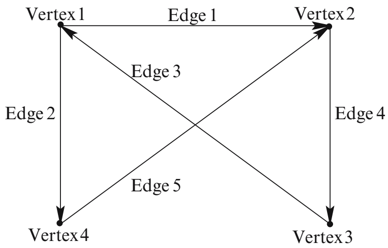


Fig. 2.4. Data from Example 2.21.

Consider the following question relating to Example 2.21. Given this incomplete data about the teams, how would we determine a ranking of each team in some sensible way? In order to answer this question, we are going to introduce some concepts from graph theory that are useful modeling tools for many problems.

The data of Figure 2.4 is an example of a *directed graph*, a modeling tool that can be defined as follows.

Directed Graph

Definition 2.10. A *directed graph* (digraph for short) is a set V whose elements are called *vertices*, together with a set or list (to allow for repeated edges) E of ordered pairs with coordinates in V, whose elements are called *(directed) edges*.

Walk

Another useful idea for us is the following: a *walk* in the digraph G is a sequence of digraph edges $(v_0, v_1), (v_1, v_2), \ldots, (v_{m-1}, v_m)$ that goes from vertex v_0 to vertex v_m. The *length* of the walk is m.

Here is an interpretation of "power" that has proved to be useful in many situations. The *power* of a vertex in a digraph is the number of walks of length 1 or 2 originating at the vertex. In our example, the power of vertex 1 is 4. Why only walks of length 1 or 2? One good reason is that walks of length 3 introduce the possibility of *loops*, i.e., walks that "loop around" to the same point. It isn't very informative to find out that team 1 beat team 2 beat team 3 beat team 1.

The digraph of Example 2.21 has no edges from a vertex to itself (so-called self-loops), and for a pair of distinct vertices, at most one edge connecting the two vertices. In other words, a team doesn't play itself and plays another team at most once. Such a digraph is called a *dominance-directed graph*. Although the notion of power of a point is defined for any digraph, it makes the most sense for dominance-directed graphs.

Dominance Directed Graph

Example 2.22. Find the power of each vertex in the graph of Example 2.21 and use this information to rank the teams.

Solution. In this example we could find the power of all points by inspection of Figure 2.4. Let's do it: simple counting gives that the power of vertex 1 is 4, the power of vertex 3 is 3, and the power of vertices 2 and 4 is 2. Consequently, teams 2 and 4 are tied for last place, team 3 is in second place, and team 1 is first. □

One can imagine situations (like describing the structure of the communications network pictured in Figure 2.5) in which the edges shouldn't really have a direction, since connections are bidirectional. For such situations a more natural tool is the concept of a *graph*, which can be defined as follows: a *graph* is a set V, whose elements are called *vertices*, together with a set or list (to allow for repeated edges) E of *unordered* pairs with coordinates in V, called *edges*.

Graph

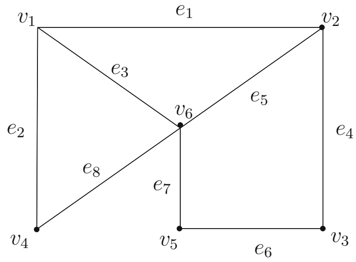


Fig. 2.5. A communications network graph.

Just as with digraphs, we define a *walk* in the graph G as a sequence of digraph edges $(v_0, v_1), (v_1, v_2), \ldots, (v_{m-1}, v_m)$ that goes from vertex v_0 to vertex v_m. The *length* of the walk is m. For example, the graph of Figure 2.5 has vertex set $V = \{v_1, v_2, v_3, v_4, v_5, v_6\}$ and edge set $E = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8\}$, with $e_1 = (v_1, v_2)$, etc, as in the figure. Also, the sequence e_1, e_4, e_6 is a walk from vertex v_1 to v_5 of length 3. As with digraphs, we can define the power of a vertex in any graph as the number of walks of length at most 2 originating at the vertex.

A practical question: how could we write a computer program to compute powers? More generally, how can we compute the total number of walks of a certain length? Here is a key to the answer: all the information about our graph (or digraph) can be stored in its *adjacency matrix*. In general, this is defined to be a square matrix whose rows and columns are indexed by the vertices of the graph and whose (i, j)th entry is the number of edges going from vertex i to vertex j (it is 0 if there are none). Here we understand that a directed edge of a digraph must start at i and end at j, while no such restriction applies to the edges of a graph.

Adjacency Matrix

Just for the record, if we designate the adjacency matrix of the digraph of Figure 2.4 by A and the adjacency matrix of the graph of Figure 2.5 by B, then

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

Notice that we could reconstruct the entire digraph or graph from this matrix. Also notice that in the adjacency matrix for a graph, an edge gets accounted for twice, since it can be thought of as proceeding from one vertex to the other, or from the other to the one.

For a general graph with n vertices and adjacency matrix $A = [a_{ij}]$, we can use this matrix to compute powers of vertices without seeing a picture of the graph. To count up the walks of length 1 emanating from vertex i, simply add up the elements of the ith row of A. Now what about the paths of length 2? Observe that there is an edge from i to k and then from k to j precisely when the product $a_{ik}a_{kj}$ is equal to 1. Otherwise, one of the factors will be 0 and therefore the product is 0. So the number of paths of length 2 from vertex i to vertex j is the familiar sum

$$a_{i1}a_{1j} + a_{i2}a_{2j} + \cdots + a_{in}a_{nj}.$$

This is just the (i, j)th entry of the matrix A^2. A similar argument shows the following fact:

Vertex Power

Theorem 2.2. If A is the adjacency matrix of the graph G, then the (i, j)th entry of A^r gives the number of walks of length r starting at vertex i and ending at vertex j.

Since the power of vertex i is the number of all paths of length 1 or 2 emanating from vertex i, we have the following key fact:

Theorem 2.3. If A is the adjacency matrix of the digraph G, then the power of the ith vertex is the sum of all entries in the ith row of the matrix $A + A^2$.

Example 2.23. Use the preceding facts to calculate the powers of all the vertices in the digraph of Example 2.21.

Solution. Using the matrix A above we calculate that

$$A + A^2 = \begin{bmatrix} 0&1&0&1 \\ 0&0&1&0 \\ 1&0&0&0 \\ 0&1&0&0 \end{bmatrix} + \begin{bmatrix} 0&1&0&1 \\ 0&0&1&0 \\ 1&0&0&0 \\ 0&1&0&0 \end{bmatrix} \begin{bmatrix} 0&1&0&1 \\ 0&0&1&0 \\ 1&0&0&0 \\ 0&1&0&0 \end{bmatrix} = \begin{bmatrix} 0&2&1&1 \\ 1&0&1&0 \\ 1&1&0&1 \\ 0&1&1&0 \end{bmatrix}.$$

An easy way to sum each row is to multiply $A + A^2$ on the right by a column of 1's, but in this case we see immediately that the power of vertex 1 is 4, the power of vertex 3 is 3, and the power of vertices 2 and 4 is 2, which is consistent with what we observed earlier by inspection of the graph. □

Difference Equations

The idea of a difference equation has numerous applications in mathematics and computer science. In the latter field, these equations often go by the name of "recurrence relations." They can be used for a variety of applications ranging from population modeling to analysis of complexity of algorithms. We will introduce them by way of a simple financial model.

Example 2.24. Suppose that you invest in a contractual fund where you must invest in the funds for three years before you can receive any return on your investment (with a positive first-year investment). Thereafter, you are vested in the fund and may remove your money at any time. While you are vested in the fund, annual returns are calculated as follows: money that was in the fund one year ago earns nothing, while money that was in the fund two years ago earns 6% of its value and money that was in the fund three years ago earns 12% of its value. Find an equation that describes your investment's growth.

Solution. Let a_k be the amount of your investment in the kth year. The numbers a_0, a_1, a_2 represent your investments for the first three years (we're counting from 0). Consider the third year amount a_3. According to your contract, your total funds in the third year will be

$$a_3 = a_2 + 0.06a_1 + 0.12a_0.$$

Now it's easy to write out a general formula for a_{k+3} in terms of the preceding three terms, using the same line of thought, namely

$$a_{k+3} = a_{k+2} + 0.06a_{k+1} + 0.12a_k, \quad k = 0, 1, 2, \ldots. \tag{2.1}$$

This is the desired formula. □

Homogeneous Linear Difference Equation

In general, a *homogeneous linear difference equation* (or *recurrence relation*) of order m in the variables $a_0, a_1, \ldots$ is an equation of the form

$$a_{k+m} + c_{m-1}a_{k+m-1} + \cdots + c_1 a_{k+1} + c_0 a_k = 0, \quad k = 0, 1, 2, \ldots.$$

Notice that such an equation cannot determine the numbers $a_0, a_1, \ldots, a_{k-1}$. These values have to be initially specified, just as in our fund example. Notice that in our fund example, we have to bring all terms of equation (2.1) to the left-hand side to obtain the difference equation form

$$a_{k+3} - a_{k+2} - 0.06a_{k+1} - 0.12a_k = 0.$$

Now we see that $c_2 = -1$, $c_1 = -0.06$, and $c_0 = -0.12$.

There are many ways to solve difference equations. We are not going to give a complete solution to this problem at this point; we postpone this issue to Chapter 5, where we introduce eigenvalues and eigenvectors. However, we can now show how to turn a difference equation as given above into a matrix equation. Consider our fund example. The secret is to identify the right vector variables. To this end, define an indexed vector $\mathbf{x}_k$ by the formula

$$\mathbf{x}_k = \begin{bmatrix} a_{k+2} \\ a_{k+1} \\ a_k \end{bmatrix}, \qquad k = 0, 1, 2, \ldots.$$

Thus

$$\mathbf{x}_{k+1} = \begin{bmatrix} a_{k+3} \\ a_{k+2} \\ a_{k+1} \end{bmatrix},$$

from which it is easy to check that since $a_{k+3} = a_{k+2} + 0.06a_{k+1} + 0.12a_k$, we have

$$\mathbf{x}_{k+1} = \begin{bmatrix} 1 & 0.06 & 0.12 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \mathbf{x}_k = A\mathbf{x}_k.$$

This is the matrix form we seek. It appears to have a lot in common with the Markov chains examined earlier in this section, in that we pass from one "state vector" to another by multiplication by a fixed "transition matrix" A.

2.3 Exercises and Problems

Exercise 1. Determine the effect of the matrix operator T_A on the x-axis, y-axis, and the points $(\pm 1, \pm 1)$, where A is one of the following.

(a) $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ (b) $\frac{1}{5}\begin{bmatrix} -3 & -4 \\ -4 & 3 \end{bmatrix}$ (c) $\begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$

Exercise 2. Determine the effect of the matrix operator T_A on the x-axis, y-axis, and the points $(\pm 1, \pm 1)$, where A is one of the following. Plot the images of the squares with corners $(\pm 1, \pm 1)$.

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

Exercise 3. Express the following functions, if linear, as matrix operators.
(a) $T((x_1, x_2)) = (x_1 + x_2, 2x_1, 4x_2 - x_1)$ (b) $T((x_1, x_2)) = (x_1 + x_2, 2x_1x_2)$
(c) $T((x_1, x_2, x_3)) = (2x_3, -x_1)$ (d) $T((x_1, x_2, x_3)) = (x_2 - x_1, x_3, x_2 + x_3)$

Exercise 4. Express the following functions, if linear, as matrix operators.
(a) $T((x_1, x_2, x_3)) = x_1 - x_3 + 2x_2$ (b) $T((x_1, x_2)) = (|x_1|, 2x_2, x_1 + 3x_2)$
(c) $T((x_1, x_2)) = (x_1, 2x_1, -x_1)$ (d) $T((x_1, x_2, x_3)) = (-x_3, x_1, 4x_2)$

Exercise 5. A linear operator on $\mathbb{R}^2$ is defined by first applying a scaling operator with scale factors of 2 in the x-direction and 4 in the y-direction, followed by a counterclockwise rotation about the origin of $\pi/6$ radians. Express this operator and the operator that results from reversing the order of the scaling and rotation as matrix operators.

Exercise 6. A linear operator on $\mathbb{R}^2$ is defined by first applying a shear in the x-direction with a shear factor of 3 followed by a clockwise rotation about the origin of $\pi/4$ radians. Express this operator and the operator that results from reversing the order of the shear and rotation as matrix operators.

Exercise 7. A *fixed-point* of a linear operator T_A is a vector $\mathbf{x}$ such that $T_A(\mathbf{x}) = \mathbf{x}$. Find all fixed points, if any, of the linear operators in Exercise 3.

Exercise 8. Find all fixed points, if any, of the linear operators in Exercise 4.

Exercise 9. Given transition matrices for discrete dynamical systems
(a) $\begin{bmatrix} .1 & .3 & 0 \\ 0 & .4 & 1 \\ .9 & .3 & 0 \end{bmatrix}$ (b) $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ (c) $\begin{bmatrix} .5 & .3 & 0 \\ 0 & .4 & 0 \\ .5 & .3 & 1 \end{bmatrix}$ (d) $\begin{bmatrix} 0 & 0 & 0.9 \\ 0.5 & 0 & 0 \\ 0 & 0.5 & 0.1 \end{bmatrix}$
and initial state vector $\mathbf{x}^{(0)} = \frac{1}{2}(1, 1, 0)$, calculate the first and second state vector for each system and determine whether it is a Markov chain.

Exercise 10. For each of the dynamical systems of Exercise 9, determine by calculation whether the system tends to a limiting steady-state vector. If so, what is it?

Exercise 11. A digraph G has vertex set $V = \{1, 2, 3, 4, 5\}$ and edge set $E = \{(2, 1), (1, 5), (2, 5), (5, 4), (4, 2), (4, 3), (3, 2)\}$. Sketch a picture of the graph G and find its adjacency matrix. Use this to find the power of each vertex of the graph and determine whether this graph is dominance-directed.

Exercise 12. A digraph has the following adjacency matrix:

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

Sketch a picture of this digraph and compute the total number of walks in the digraph of length at most 3.

Exercise 13. Convert these difference equations into matrix–vector form.
(a) $2a_{k+3} + 3a_{k+2} - 4a_{k+1} + 5a_k = 0$ (b) $a_{k+2} - a_{k+1} + 2a_k = 1$

Exercise 14. Convert these difference equations into matrix–vector form.
(a) $2a_{k+3} + 2a_{k+1} - 3a_k = 0$ (b) $a_{k+2} + a_{k+1} - 2a_k = 3$

*Problem 15. Show that if $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is a real 2×2 matrix, then the matrix multiplication function maps a line through the origin onto a line through the origin or a point.

Problem 16. Show how the transition matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ for a Markov chain can be described using only two variables.

*Problem 17. Use the definition of matrix multiplication function to show that if $T_A = T_B$, then $A = B$.

Problem 18. Show that if the state vector $\mathbf{x}^{(k)} = (a_k, b_k, c_k)$ in a Markov chain is a probability distribution vector, then so is $\mathbf{x}^{(k+1)}$.

Problem 19. Suppose that in Example 2.24 you invest $1,000 initially (the zeroth year) and no further amounts. Make a table of the value of your investment for years 0 to 12. Also include a column that calculates the annual interest rate that your investment is earning each year, based on the current and previous year's values. What conclusions do you draw? You will need a computer or calculator for this exercise.

2.4 Special Matrices and Transposes

There are certain types of matrices that are so important that they have acquired names of their own. We introduce some of these in this section, as well as one more matrix operation that has proved to be a very practical tool in matrix analysis, namely the operation of transposing a matrix.

Elementary Matrices and Gaussian Elimination

We are going to show a new way to execute the elementary row operations used in Gaussian elimination. Recall the shorthand we used:

- E_{ij}: The elementary operation of *switching the ith and jth rows* of the matrix.
- $E_i(c)$: The elementary operation of *multiplying the ith row by the nonzero constant c.*

- $E_{ij}(d)$: The elementary operation of *adding d times the jth row to the ith row.*

From now on we will use the very same symbols to represent matrices. The size of the matrix will depend on the context of our discussion, so the notation is ambiguous, but it is still very useful.

An *elementary matrix* of size n is obtained by performing the corresponding elementary row operation on the identity matrix I_n. We denote the resulting matrix by the same symbol as the corresponding row operation.

Elementary Matrix

Example 2.25. Describe the following elementary matrices of size $n = 3$:
(a) $E_{13}(-4)$ (b) $E_{21}(3)$ (c) E_{23} (d) $E_1(\frac{1}{2})$

Solution. We start with

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

For part (a) we add -4 times the 3rd row of I_3 to its first row to obtain

$$E_{13}(-4) = \begin{bmatrix} 1 & 0 & -4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

For part (b) add 3 times the first row of I_3 to its second row to obtain

$$E_{21}(3) = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

For part (c) interchange the second and third rows of I_3 to obtain that

$$E_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Finally, for part (d) we multiply the first row of I_3 by $\frac{1}{2}$ to obtain

$$E_1\left(\frac{1}{2}\right) = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$ □

What good are these matrices? One can see that the following fact is true:

Theorem 2.4. Let $C = BA$ be a product of two matrices and perform an elementary row operation on C. Then the same result is obtained if one performs the same elementary operation on the matrix B and multiplies the result by A on the right.

We won't give a formal proof of this statement, but it isn't hard to see why it is true. For example, suppose one interchanges two rows, say the ith and jth, of $C = BA$ to obtain a new matrix D. How do we get the ith or jth row of C? Answer: multiply the corresponding row of B by the matrix A. Therefore, we would obtain D by interchanging the ith and jth rows of B and multiplying the result by the matrix A, which is exactly what the theorem says. Similar arguments apply to the other elementary operations.

Now take $B = I$, and we see from the definition of elementary matrix and Theorem 2.4 that the following is true.

Corollary 2.1. If an elementary row operation is performed on a matrix A to obtain a matrix A', then $A' = EA$, where E is the elementary matrix corresponding to the elementary row operation performed.

The meaning of this corollary is that we accomplish an elementary row operation by multiplying by the corresponding elementary matrix on the left. Of course, we don't need elementary matrices to accomplish row operations; but they give us another perspective on row operations.

Elementary Operations as Matrix Multiplication

Example 2.26. Express these calculations of Example 1.16 in matrix product form:

$$\begin{bmatrix} 2 & -1 & 1 \\ 4 & 4 & 20 \end{bmatrix} \overrightarrow{E_{12}} \begin{bmatrix} 4 & 4 & 20 \\ 2 & -1 & 1 \end{bmatrix} \overrightarrow{E_1(1/4)} \begin{bmatrix} 1 & 1 & 5 \\ 2 & -1 & 1 \end{bmatrix}$$

$$\overrightarrow{E_{21}(-2)} \begin{bmatrix} 1 & 1 & 5 \\ 0 & -3 & -9 \end{bmatrix} \overrightarrow{E_2(-1/3)} \begin{bmatrix} 1 & 1 & 5 \\ 0 & 1 & 3 \end{bmatrix} \overrightarrow{E_{12}(-1)} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \end{bmatrix}.$$

Solution. One point to observe: the order of elementary operations. We compose the elementary matrices on the left in the same order that the operations are done. Thus we may state the above calculations in the concise form

$$\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \end{bmatrix} = E_{12}(-1)\, E_2(-1/3)\, E_{21}(-2)\, E_1(1/4)\, E_{12} \begin{bmatrix} 2 & -1 & 1 \\ 4 & 4 & 20 \end{bmatrix}. \qquad \square$$

It is important to read the preceding line carefully and understand how it follows from the long form above. This conversion of row operations to matrix multiplication will prove to be very practical in the next section.

Some Matrices with Simple Structure

Certain types of matrices have already turned up in our discussions. For example, the identity matrices are particularly easy to deal with. Another example is the reduced row echelon form. So let us classify some simple matrices and attach names to them. The simplest conceivable matrices are zero matrices. We have already seen that they are important in matrix addition arithmetic. What's next? For square matrices, we have the following definitions, in ascending order of complexity.

Simple Structure Matrices

Definition 2.11. Let $A = [a_{ij}]$ be a square $n \times n$ matrix. Then A is

- *Scalar* if $a_{ij} = 0$ and $a_{ii} = a_{jj}$ for all $i \neq j$. (Equivalently: $A = cI_n$ for some scalar c, which explains the term "scalar.")
- *Diagonal* if $a_{ij} = 0$ for all $i \neq j$. (Equivalently: off-diagonal entries of A are 0.)
- *(Upper) triangular* if $a_{ij} = 0$ for all $i > j$. (Equivalently: subdiagonal entries of A are 0.)
- *(Lower) triangular* if $a_{ij} = 0$ for all $i < j$. (Equivalently: superdiagonal entries of A are 0.)
- *Triangular* if the matrix is upper or lower triangular.
- *Strictly triangular* if it is triangular and the diagonal entries are also zero.
- *Tridiagonal* if $a_{ij} = 0$ when $j > i + 1$ or $j < i - 1$. (Equivalently: entries off the main diagonal, first subdiagonal, and first superdiagonal are zero.)

The index conditions that we use above have simple interpretations. For example, the entry a_{ij} with $i > j$ is located further down than over, since the row number is larger than the column number. Hence, it resides in the "lower triangle" of the matrix. Similarly, the entry a_{ij} with $i < j$ resides in the "upper triangle." Entries a_{ij} with $i = j$ reside along the main diagonal of the matrix. See Figure 2.6 for a picture of these triangular regions of the matrix.

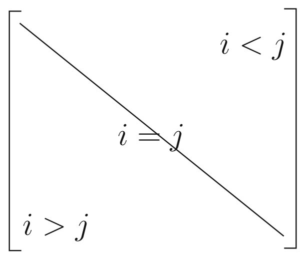


Fig. 2.6: Matrix regions.

Example 2.27. Classify the following matrices (elementary matrices are understood to be 3×3) in the terminology of Definition 2.11.

(a) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$ (b) $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 4 \\ 0 & 0 & 2 \end{bmatrix}$ (d) $\begin{bmatrix} 0 & 0 & 0 \\ 1 & -1 & 0 \\ 3 & 2 & 2 \end{bmatrix}$

(e) $\begin{bmatrix} 0 & 2 & 3 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{bmatrix}$ (f) $E_{21}(3)$ (g) $E_2(-3)$ (h) $\begin{bmatrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 2 \end{bmatrix}$

Solution. Notice that (a) is not scalar, since diagonal entries differ from each other, but it is a diagonal matrix, since the off-diagonal entries are all 0. On the other hand, the matrix of (b) is really just $2I_3$, so this matrix is a scalar matrix. Matrix (c) has all terms below the main diagonal equal to 0, so this matrix is triangular and, specifically, upper triangular. Similarly, matrix (d) is lower triangular. Matrix (e) is clearly upper triangular, but it is also strictly upper triangular since the diagonal terms themselves are 0. Finally, we have

$$E_{21}(3) = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad E_2(-3) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

so that $E_{21}(3)$ is (lower) triangular and $E_2(-3)$ is a diagonal matrix. Matrix (h) comes from Example 1.3, where we saw that an approximation to a certain diffusion problem led to matrices of that form. Moreover, if we want more accurate solutions to the original problem, we would need to solve systems with a similar, but larger, coefficient matrix. This matrix is clearly tridiagonal. In fact, note that the matrices of (a), (b), (f), and (g) also can be classified as tridiagonal. □

Block Matrices

Another type of matrix that occurs frequently enough to be discussed is a *block matrix*. Actually, we already used the idea of blocks when we described the augmented matrix of the system $A\mathbf{x} = \mathbf{b}$ as the matrix $\widetilde{A} = [A \,|\, \mathbf{b}]$. We say that $\widetilde{A}$ has the *block*, or *partitioned*, form $[A, \mathbf{b}]$. What we are really doing is partitioning the matrix $\widetilde{A}$ by inserting a vertical line between elements. There is no reason we couldn't partition by inserting more vertical lines or horizontal lines as well, and this partitioning leads to the blocks. The main point to bear in mind when using the block notation is that the blocks must be correctly sized so that the resulting matrix makes sense. The main virtue of the block form that results from partitioning is that for purposes of matrix addition or multiplication, we can treat the blocks rather like scalars, provided the addition or multiplication that results makes sense. We will use this idea from time to time without fanfare. One could go through a formal description of partitioning and proofs; we won't. Rather, we'll show how this idea can be used by example.

Block Notation

Example 2.28. Use block multiplication to simplify the following multiplication:

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

Solution. The blocking we want to use makes the column numbers of the blocks on the left match the row numbers of the blocks on the right and looks like this:

$$\left[\begin{array}{cc|cc} 1 & 2 & 0 & 0 \\ 3 & 4 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \end{array}\right] \left[\begin{array}{cc|cc} 0 & 0 & 2 & 1 \\ 0 & 0 & -1 & 1 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right].$$

We see that these submatrices are built from zero matrices and these blocks:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 \end{bmatrix}, \qquad C = \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}, \qquad I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Now we can work this product out by interpreting it as

$$\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \begin{bmatrix} 0 & C \\ 0 & I_2 \end{bmatrix} = \begin{bmatrix} A \cdot 0 + 0 \cdot 0 & A \cdot C + 0 \cdot I_2 \\ 0 \cdot 0 + B \cdot 0 & 0 \cdot C + B \cdot I_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 3 \\ 0 & 0 & 2 & 7 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \qquad \square$$

For another (important!) example of block arithmetic, examine Example 2.9 and the discussion following it. There we view a matrix as blocked into its respective columns, and a column vector as blocked into its rows, to obtain

$$A\mathbf{x} = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{a}_1 x_1 + \mathbf{a}_2 x_2 + \mathbf{a}_3 x_3.$$

Transpose of a Matrix

Sometimes we prefer to work with a different form of a given matrix that contains the same information. Transposes are operations that allow us to do that. The idea is simple: interchange rows and columns. It turns out that for complex matrices, there is an analogue that is not quite the same thing as transposing, though it yields the same result when applied to real matrices. This analogue is called the conjugate (Hermitian) transpose. Here are the appropriate definitions.

Definition 2.12. Let $A = [a_{ij}]$ be an $m \times n$ matrix with (possibly) complex entries. Then the *transpose* of A is the $n \times m$ matrix A^T obtained by interchanging the rows and columns of A, so that the (i, j)th entry of A^T is a_{ji}. The *conjugate* of A is the matrix $\overline{A} = [\overline{a_{ij}}]$. Finally, the *conjugate (Hermitian) transpose* of A is the matrix $A^* = \overline{A}^T$.

Transpose and Conjugate Matrices

Notice that in the case of a real matrix (that is, a matrix with real entries) A there is no difference between transpose and conjugate transpose, since in this case $A = \overline{A}$. Consider these examples.

Example 2.29. Compute the transpose and conjugate transpose of the following matrices:

$$\text{(a)} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix}, \quad \text{(b)} \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}, \quad \text{(c)} \begin{bmatrix} 1 & 1 + \mathrm{i} \\ 0 & 2\mathrm{i} \end{bmatrix}.$$

Solution. For matrix (a) we have

$$\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix}^* = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix}^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{bmatrix}.$$

Notice, by the way, how the dimensions of a transpose get switched from the original.

For matrix (b) we have

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

and for matrix (c) we have

$$\begin{bmatrix} 1 & 1+\mathrm{i} \\ 0 & 2\mathrm{i} \end{bmatrix}^* = \begin{bmatrix} 1 & 0 \\ 1-\mathrm{i} & -2\mathrm{i} \end{bmatrix}, \quad \begin{bmatrix} 1 & 1+\mathrm{i} \\ 0 & 2\mathrm{i} \end{bmatrix}^T = \begin{bmatrix} 1 & 0 \\ 1+\mathrm{i} & 2\mathrm{i} \end{bmatrix}.$$

In this case, transpose and conjugate transpose are not the same. □

Even when dealing with vectors alone, the transpose notation is handy. For example, there is a bit of terminology that comes from tensor analysis (a branch of higher linear algebra used in many fields including differential geometry, engineering mechanics, and relativity) that can be expressed very concisely with transposes:

Inner and Outer Products

Definition 2.13. Let $\mathbf{u}$ and $\mathbf{v}$ be column vectors of the same size, say $n \times 1$. Then the *inner product* of $\mathbf{u}$ and $\mathbf{v}$ is the scalar quantity $\mathbf{u}^T\mathbf{v}$, and the *outer product* of $\mathbf{u}$ and $\mathbf{v}$ is the $n \times n$ matrix $\mathbf{u}\mathbf{v}^T$.

Example 2.30. Compute the inner and outer products of the vectors

$$\mathbf{u} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix}.$$

Solution. Here we have the inner product

$$\mathbf{u}^T\mathbf{v} = [2, -1, 1] \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix} = 2 \cdot 3 + (-1)4 + 1 \cdot 1 = 3,$$

while the outer product is

$$\mathbf{u}\mathbf{v}^T = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} [3, 4, 1] = \begin{bmatrix} 2 \cdot 3 & 2 \cdot 4 & 2 \cdot 1 \\ -1 \cdot 3 & -1 \cdot 4 & -1 \cdot 1 \\ 1 \cdot 3 & 1 \cdot 4 & 1 \cdot 1 \end{bmatrix} = \begin{bmatrix} 6 & 8 & 2 \\ -3 & -4 & -1 \\ 3 & 4 & 1 \end{bmatrix}.$$ □

Here are a few basic laws relating transposes to other matrix arithmetic that we have learned. These laws remain correct if transpose is replaced by conjugate transpose, with one exception: $(cA)^* = \bar{c}A^*$.

Laws of Matrix Transpose

Let A and B be matrices of the appropriate sizes so that the following operations make sense, and c a scalar.

(1) $(A + B)^T = A^T + B^T$
(2) $(AB)^T = B^T A^T$
(3) $(cA)^T = cA^T$
(4) $(A^T)^T = A$

These laws are easily verified directly from definition. For example, if $A = [a_{ij}]$ and $B = [b_{ij}]$ are $m \times n$ matrices, then we have that $(A+B)^T$ is the $n \times m$ matrix given by

$$\begin{aligned}(A+B)^T &= [a_{ij}+b_{ij}]^T = [a_{ji}+b_{ji}] \\ &= [a_{ji}] + [b_{ji}] \\ &= A^T + B^T.\end{aligned}$$

The other laws are proved similarly.

We will require explicit formulas for transposes of the elementary matrices in some later calculations. Notice that the matrix $E_{ij}(c)$ is a matrix with 1's on the diagonal and 0's elsewhere, except that the (i,j)th entry is c. Therefore, transposing switches the entry c to the (j,i)th position and leaves all other entries unchanged. Hence $E_{ij}(c)^T = E_{ji}(c)$. With similar calculations we have these facts:

Transposes of Elementary Matrices

- $E_{ij}^T = E_{ij}$
- $E_i(c)^T = E_i(c)$
- $E_{ij}(c)^T = E_{ji}(c)$

Elementary Column Operations

These formulas have an interesting application. Up to this point we have considered only elementary row operations. However, there are situations in which elementary *column* operations on the columns of a matrix are useful. If we want to use such operations, do we have to start over, reinvent elementary column matrices, and so forth? The answer is no and the following example gives an indication of why the transpose idea is useful. This example shows how to do column operations in the language of matrix arithmetic. Here's the basic idea: suppose we want to do an elementary column operation on a matrix A corresponding to elementary row operation E to get a new matrix B from A. To do this, turn the columns of A into rows, do the row operation, and then transpose the result back to get the matrix B that we want. In algebraic terms

$$B = \left(EA^T\right)^T = \left(A^T\right)^T E^T = AE^T.$$

So all we have to do to perform an elementary column operation is multiply by the transpose of the corresponding elementary row matrix on the right. Thus we see that the transposes of elementary row matrices could reasonably be called *elementary column matrices*.

Elementary Column Matrix

Example 2.31. Let A be a given matrix. Suppose that we wish to express the result B of swapping the second and third columns of A, followed by adding -2 times the first column to the second, as a product of matrices. How can this be done? Illustrate the procedure with the matrix

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

Solution. Apply the preceding remark twice to obtain that

$$B = AE_{23}^T E_{21}(-2)^T = AE_{23}E_{12}(-2).$$

Thus we have

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

as a matrix product. □

A very important type of special matrix is one that is invariant under the operation of transposing. These matrices turn up naturally in applied mathematics. They have some very remarkable properties that we will study in Chapters 4, 5, and 6.

Symmetric and Hermitian Matrices

Definition 2.14. The matrix A is said to be *symmetric* if $A^T = A$ and *Hermitian* if $A^* = A$. (Equivalently, $a_{ij} = a_{ji}$ and $a_{ij} = \overline{a_{ji}}$, for all i, j, respectively.)

From the laws of transposing elementary matrices above we see right away that E_{ij} and $E_i(c)$ supply us with examples of symmetric matrices. Here are a few more.

Example 2.32. Are the following matrices symmetric or Hermitian?

$$\text{(a)} \begin{bmatrix} 1 & 1+\mathrm{i} \\ 1-\mathrm{i} & 2 \end{bmatrix}, \quad \text{(b)} \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}, \quad \text{(c)} \begin{bmatrix} 1 & 1+\mathrm{i} \\ 1+\mathrm{i} & 2\mathrm{i} \end{bmatrix}$$

Solution. For matrix (a) we have

$$\begin{bmatrix} 1 & 1+\mathrm{i} \\ 1-\mathrm{i} & 2 \end{bmatrix}^* = \begin{bmatrix} \overline{1} & \overline{1+\mathrm{i}} \\ \overline{1-\mathrm{i}} & \overline{2} \end{bmatrix}^T = \begin{bmatrix} 1 & 1+\mathrm{i} \\ 1-\mathrm{i} & 2 \end{bmatrix}.$$

Hence this matrix is Hermitian. However, it is *not* symmetric since the $(1,2)$th and $(2,1)$th entries differ. Matrix (b) is easily seen to be symmetric by inspection. Matrix (c) is symmetric since the $(1,2)$th and $(2,1)$th entries agree, but it is not conjugate Hermitian since

$$\begin{bmatrix} 1 & 1+\mathrm{i} \\ 1-\mathrm{i} & 2\mathrm{i} \end{bmatrix}^* = \begin{bmatrix} \overline{1} & \overline{1+\mathrm{i}} \\ \overline{1-\mathrm{i}} & \overline{2\mathrm{i}} \end{bmatrix}^T = \begin{bmatrix} 1 & 1+\mathrm{i} \\ 1-\mathrm{i} & -2\mathrm{i} \end{bmatrix},$$

and this last matrix is clearly not equal to matrix (c). □

Example 2.33. Consider the quadratic form (this means a homogeneous second-degree polynomial in the variables)

$$Q(x, y, z) = x^2 + 2y^2 + z^2 + 2xy + yz + 3xz.$$

Express this function in terms of matrix products and transposes.

Solution. Write the quadratic form as

$$x(x+2y+3z)+y(2y+z)+z^2 = \begin{bmatrix} x \; y \; z \end{bmatrix} \begin{bmatrix} x+2y+3z \\ 2y+z \\ z \end{bmatrix}$$
$$= \begin{bmatrix} x \; y \; z \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{x}^T A \mathbf{x},$$

where

$$\mathbf{x} = (x, y, z) \text{ and } A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$ □

Rank of the Matrix Transpose

A basic question is how the rank of a matrix transpose (or Hermitian transpose) is connected to the rank of the matrix. There is a nice answer. We will focus on transposes. First we need the following theorem.

Theorem 2.5. Let A, B be matrices such that the product AB is defined. Then

$$\operatorname{rank} AB \le \operatorname{rank} A.$$

Proof. Let E be a product of elementary matrices such that $EA = R$, where R is the reduced row echelon form of A. If $\operatorname{rank} A = r$, then the first r rows of R have leading entries of 1, while the remaining rows are zero rows. Also, we saw in Chapter 1 that elementary row operations do not change the rank of a matrix, since according to Corollary 1.1 they do not change the reduced row echelon form of a matrix. Therefore,

$$\operatorname{rank} AB = \operatorname{rank} E(AB) = \operatorname{rank}(EA)B = \operatorname{rank} RB.$$

Now the matrix RB has the same number of rows as R, and the first r of these rows may or may not be nonzero, but the remaining rows must be zero rows, since they result from multiplying columns of B by the zero rows of R. If we perform elementary row operations to reduce RB to its reduced row echelon form we will possibly introduce more zero rows than R has. Consequently, $\operatorname{rank} RB \le r = \operatorname{rank} A$, which completes the proof. □

Theorem 2.6. For any matrix A,

Rank Invariant Under Transpose

$$\operatorname{rank} A = \operatorname{rank} A^T.$$

Proof. As in the previous theorem, let E be a product of elementary matrices such that $EA = R$, where R is the reduced row echelon form of A. If

$\operatorname{rank} A = r$, then the first r rows of R have leading entries of 1 whose column numbers form an increasing sequence, while the remaining rows are zero rows. Therefore, $R^T = A^T E^T$ is a matrix whose columns have leading entries of 1 and whose row numbers form an increasing sequence. Use elementary row operations to clear out the nonzero entries below each column with a leading 1 to obtain a matrix whose rank is equal to the number of such leading entries, i.e., equal to r. Thus, $\operatorname{rank} R^T = r$.

From Theorem 2.5 we have that $\operatorname{rank} A^T E^T \leq \operatorname{rank} A^T$. It follows that

$$\operatorname{rank} A = \operatorname{rank} R^T = \operatorname{rank} A^T E^T \leq \operatorname{rank} A^T.$$

If we substitute the matrix A^T for the matrix A in this inequality, we obtain that

$$\operatorname{rank} A^T \leq \operatorname{rank}(A^T)^T = \operatorname{rank} A.$$

It follows from these two inequalities that $\operatorname{rank} A = \operatorname{rank} A^T$, which is what we wanted to show. $\square$

It is instructive to see how a specific example might work out in the preceding proof. For example, R might look like this, where an x designates an arbitrary entry:

$$R = \begin{bmatrix} 1 & 0 & x & 0 & x \\ 0 & 1 & x & 0 & x \\ 0 & 0 & 0 & 1 & x \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

so that R^T would given by

$$R^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ x & x & 0 & 0 \\ 0 & 0 & 1 & 0 \\ x & x & x & 0 \end{bmatrix}.$$

Thus if we use elementary row operations to zero out the entries below a column pivot, all entries to the right and below this pivot are unaffected by these operations. Now start with the leftmost column and proceed to the right, zeroing out all entries under each column pivot. The result is a matrix that looks like

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

Now swap rows to move the zero rows to the bottom if necessary, and we see that the reduced row echelon form of R^T has exactly as many nonzero rows as did R, that is, r nonzero rows.

A first application of this important fact is to give a fuller picture of the rank of a product of matrices than that given by Theorem 2.5:

Corollary 2.2. If the product AB is defined, then

Rank of Matrix Product

$$\operatorname{rank} AB \leq \min\{\operatorname{rank} A, \operatorname{rank} B\}.$$

Proof. We know from Theorem 2.5 that

$$\operatorname{rank} AB \leq \operatorname{rank} A \text{ and } \operatorname{rank} B^T A^T \leq \operatorname{rank} B^T.$$

Since $B^T A^T = (AB)^T$, Theorem 2.6 tells us that

$$\operatorname{rank} B^T A^T = \operatorname{rank} AB \text{ and } \operatorname{rank} B^T = \operatorname{rank} B.$$

Put all this together, and we have

$$\operatorname{rank} AB = \operatorname{rank} B^T A^T \leq \operatorname{rank} B^T = \operatorname{rank} B.$$

It follows that $\operatorname{rank} AB$ is at most the smaller of $\operatorname{rank} A$ and $\operatorname{rank} B$, which is what the corollary asserts. □

2.4 Exercises and Problems

Exercise 1. Convert the following 3×3 elementary operations to matrix form and convert matrices to elementary operation form.

(a) $E_{23}(3)$ (b) E_{13} (c) $E_3(2)$ (d) $E_{23}^T(-1)$

(e) $\begin{bmatrix} 1 & 3 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (f) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -a & 0 & 1 \end{bmatrix}$ (g) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (h) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$

Exercise 2. Convert the following 4×4 elementary operations to matrix form and convert matrices to elementary operation form.

(a) E_{24}^T (b) $E_4(-1)$ (c) $E_3^T(2)$ (d) $E_{14}(-1)$

(e) $\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$ (f) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$ (g) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -3 & 0 & 0 & 1 \end{bmatrix}$

Exercise 3. Describe the effect of multiplying the 3×3 matrix A by each matrix in Exercise 1 on the left.

Exercise 4. Describe the effect of multiplying the 4×4 matrix A by each matrix in Exercise 2 on the right.

Exercise 5. Compute the reduced row echelon form of the following matrices and express each form as a product of elementary matrices and the original matrix.

(a) $\begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 2 & 2 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & -2 \end{bmatrix}$ (3) $\begin{bmatrix} 0 & 1+\mathrm{i} & \mathrm{i} \\ 1 & 0 & -2 \end{bmatrix}$

Exercise 6. Compute the reduced row echelon form of the following matrices and express each form as a product of elementary matrices and the original matrix.

(a) $\begin{bmatrix} 2 & 1 \\ 0 & 1 \\ 0 & 2 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 2 & 2 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 1 \\ 1 & 1+\mathrm{i} \end{bmatrix}$ (d) $\begin{bmatrix} 2 & 2 & 0 & 2 \\ 1 & 1 & -4 & 3 \end{bmatrix}$

Exercise 7. Identify the minimal list of simple structure descriptions that apply to these matrices (e.g., if "upper triangular," omit "triangular.")

(a) $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix}$ (b) $\begin{bmatrix} 2 & 1 & 4 & 2 \\ 0 & 2 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ (c) I_3 (d) $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ (e) $\begin{bmatrix} 2 & 0 \\ 3 & 1 \end{bmatrix}$

Exercise 8. Identify the minimal list of simple structure descriptions that apply to these matrices.

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

Exercise 9. Identify the appropriate blocking and calculate the matrix product AB using block multiplication, where

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

and as many submatrices that form scalar matrices or zero matrices are blocked out as possible.

Exercise 10. Confirm that sizes are correct for block multiplication and calculate the matrix product AB, where $A = \begin{bmatrix} R & 0 \\ S & T \end{bmatrix}$, $B = \begin{bmatrix} U \\ V \end{bmatrix}$, $R = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}$, $S = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix}$, $T = \begin{bmatrix} 1 & -1 \\ 2 & 2 \end{bmatrix}$, $U = \begin{bmatrix} 1 & 0 \\ 1 & 2 \\ 1 & 1 \end{bmatrix}$, and $V = \begin{bmatrix} 3 & 1 \\ 0 & 1 \end{bmatrix}$.

Exercise 11. Express the matrix $\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ 2 & 4 & 2 \end{bmatrix}$ as an outer product of two vectors.

Exercise 12. Express the rank-two matrix $\begin{bmatrix} 1 & -1 & 1 \\ 0 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix}$ as the sum of two outer products of vectors.

Exercise 13. Compute the transpose and conjugate transpose of the following matrices and determine which are symmetric or Hermitian.

(a) $\begin{bmatrix} 1 & -3 & 2 \end{bmatrix}$ (b) $\begin{bmatrix} 2 & 1 \\ 0 & 3 \\ 1 & -4 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & \mathrm{i} \\ -\mathrm{i} & 2 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 1 & 3 \\ 1 & 0 & 0 \\ 3 & 0 & 2 \end{bmatrix}$

Exercise 14. Determine which of the following matrices are symmetric or Hermitian.

(a) $\begin{bmatrix} 1 & -3 & 2 \\ -3 & 0 & 0 \\ 2 & 0 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} \mathrm{i} & 1 \\ -1 & \mathrm{i} \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 1 \\ 0 & 1 & 2 \end{bmatrix}$

Exercise 15. Answer True/False.
(a) $E_{ij}(c)^T = E_{ji}(c)$.
(b) Every elementary matrix is symmetric.
(c) The rank of the matrix A may differ from the rank of A^T.
(d) Every real diagonal matrix is Hermitian.
(e) For matrix A and scalar c, $(cA)^* = cA^*$.

Exercise 16. Answer True/False and give reasons.
(a) For matrices A, B, $(AB)^T = B^T A^T$.
(b) Every diagonal matrix is symmetric.
(c) $\operatorname{rank}(AB) = \min\{\operatorname{rank} A, \operatorname{rank} B\}$.
(d) Every diagonal matrix is Hermitian.
(e) Every tridiagonal matrix is symmetric.

Exercise 17. Express the quadratic form $Q(x, y, z) = 2x^2 + y^2 + z^2 + 2xy + 4yz - 6xz$ in the matrix form $\mathbf{x}^T A\mathbf{x}$, where A has as few nonzero entries as possible.

Exercise 18. Express the quadratic form $Q(x, y, z) = x^2 + y^2 - z^2 + 4yz - 6xz$ in the matrix form $\mathbf{x}^T A\mathbf{x}$, where A is a lower triangular matrix.

Exercise 19. Let $A = \begin{bmatrix} -2 & 1 - 2\mathrm{i} \\ 0 & 3 \end{bmatrix}$ and verify that both A^*A and AA^* are Hermitian.

Exercise 20. A square matrix A is called *normal* if $A^*A = AA^*$. Determine which of the following matrices are normal.

(a) $\begin{bmatrix} 2 & \mathrm{i} \\ 1 & 2 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & \mathrm{i} \\ 1 & 2 + \mathrm{i} \end{bmatrix}$ (d) $\begin{bmatrix} 1 & -\sqrt{3} \\ \sqrt{3} & 1 \end{bmatrix}$

Problem 21. Show that a triangular and symmetric matrix is a diagonal matrix.

*Problem 22. Let A and C be square matrices and suppose that the matrix $M = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$ is in block form. Show that for some matrix D, $M^2 = \begin{bmatrix} A^2 & D \\ 0 & C^2 \end{bmatrix}$.

Problem 23. Show that if $C = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ in block form, then $\operatorname{rank} C = \operatorname{rank} A + \operatorname{rank} B$.

Problem 24. Prove from the definition that $(A^T)^T = A$.

Problem 25. Let A be an $m \times n$ matrix. Show that both A^*A and AA^* are Hermitian.

Problem 26. Use Corollary 2.2 to prove that the outer product of any two vectors is either a rank-one matrix or zero.

Problem 27. Let A be a square real matrix. Show the following.
(a) The matrix $B = \frac{1}{2}\left(A + A^T\right)$ is symmetric.
(b) The matrix $C = \frac{1}{2}\left(A - A^T\right)$ is skew-symmetric (a matrix C is *skew-symmetric* if $C^T = -C$.)
(c) The matrix A can be expressed as the sum of a symmetric matrix and a skew-symmetric matrix.
(d) With B and C as in parts (a) and (b), show that for any vector $\mathbf{x}$ of conformable size, $\mathbf{x}^T A\mathbf{x} = \mathbf{x}^T B\mathbf{x}$.
(e) Express $A = \begin{bmatrix} 2 & 2 & -6 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix}$ as a sum of a symmetric and a skew-symmetric matrix.

Problem 28. Find all 2×2 idempotent upper triangular matrices A (idempotent means $A^2 = A$).

*Problem 29. Let D be a diagonal matrix with distinct entries on the diagonal and B any other matrix of the same size. Show that $DB = BD$ if and only if B is diagonal.

Problem 30. Show that an $n\times n$ strictly upper triangular matrix N is nilpotent. (It might help to see what happens in a 2×2 and a 3×3 case first.)

Problem 31. Use Problem 27 to show that every quadratic form $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ defined by matrix A can be defined by a symmetric matrix $B = (A + A^T)/2$ as well. Apply this result to the matrix of Example 2.33.

*Problem 32. Suppose that $A = B + C$, where B is a symmetric matrix and C is a skew-symmetric matrix. Show that $B = \frac{1}{2}(A + A^T)$ and $C = \frac{1}{2}(A - A^T)$.

2.5 Matrix Inverses

Definitions

We have seen that if we could make sense of "$1/A$," then we could write the solution to the linear system $A\mathbf{x} = \mathbf{b}$ as simply $\mathbf{x} = (1/A)\mathbf{b}$. We are going to tackle this problem now. First, we need a definition of the object that we are trying to uncover. Notice that "inverses" could work only on one side. For example,

$$\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = [1] = \begin{bmatrix} 2 & 3 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix},$$

which suggests that both $\begin{bmatrix} 1 & 2 \end{bmatrix}$ and $\begin{bmatrix} 2 & 3 \end{bmatrix}$ should qualify as left inverses of the matrix $\left[\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right]$], since multiplication on the left by them results in a 1×1 identity matrix. As a matter of fact, right and left inverses are studied and do have applications. But they have some unusual properties such as nonuniqueness. We are going to focus on a type of inverse that is closer to the familiar inverses in fields of numbers, namely, *two-sided* inverses. These make sense only for square matrices, so the nonsquare example above is ruled out.

Definition 2.15. Let A be a square matrix. Then a *(two-sided) inverse* for A is a square matrix B of the same size as A such that $AB = I = BA$. If such a B exists, then the matrix A is said to be *invertible*.

Invertible Matrix

Of course, any nonsquare matrix is noninvertible. Square matrices are classified as either "*singular*," i.e., noninvertible, or "*nonsingular*," i.e., invertible. Since we will mostly be concerned with two-sided inverses, the unqualified term "inverse" will be understood to mean a "two-sided inverse." Notice that this definition is actually symmetric in A and B. In other words, if B is an inverse for A, then A is an inverse for B.

Nonsingular Matrix

Examples of Inverses

Example 2.34. Show that $B = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$ is an inverse for $A = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$.

Solution. All we have to do is check the definition. But remember that there are *two* multiplications to confirm. (We'll show later that this isn't necessary, but right now we are working strictly from the definition.) We have

$$AB = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 2 \cdot 1 - 1 \cdot 1 & 2 \cdot 1 - 1 \cdot 2 \\ -1 \cdot 1 + 1 \cdot 1 & -1 \cdot 1 + 1 \cdot 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

and similarly

$$BA = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 \cdot 2 + 1 \cdot (-1) & 1 \cdot (-1) + 1 \cdot 1 \\ 1 \cdot 2 + 2 \cdot (-1) & 1 \cdot (-1) + 2 \cdot 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I.$$

Therefore the definition for inverse is satisfied, so that A and B work as inverses to each other. □

Example 2.35. Show that the matrix $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ cannot have an inverse.

Solution. How do we get our hands on a "noninverse"? We try an indirect approach. If A had an inverse B, then we could always find a solution to the linear system $A\mathbf{x} = \mathbf{b}$ by multiplying each side on the left by B to obtain that $BA\mathbf{x} = I\mathbf{x} = \mathbf{x} = B\mathbf{b}$, *no matter what right-hand-side vector* $\mathbf{b}$ we used. Yet it is easy to come up with right-hand-side vectors for which the system has no solution. For example, try $\mathbf{b} = (1, 2)$. Since the resulting system is clearly inconsistent, there cannot be an inverse matrix B, which is what we wanted to show. □

The moral of this last example is that it is not enough for every entry of a matrix to be nonzero for the matrix itself to be invertible. Our next example contains a gold mine of invertible matrices, namely any elementary matrix we can construct.

Example 2.36. Find formulas for inverses of all the elementary matrices.

Solution. Recall from Corollary 2.1 that left multiplication by an elementary matrix is the same as performing the corresponding elementary row operation. Furthermore, from the discussion following Theorem 1.2 we see the following:

Elementary Matrix Inverses

- E_{ij}: The elementary operation of switching the ith and jth rows is undone by applying E_{ij}. Hence
$$E_{ij}E_{ij} = E_{ij}E_{ij}I = I,$$
so that E_{ij} works as its own inverse. (This is rather like -1, since $(-1) \cdot (-1) = 1$.)
- $E_i(c)$: The elementary operation of multiplying the ith row by the nonzero constant c, is undone by applying $E_i(1/c)$. Hence
$$E_i(1/c)E_i(c) = E_i(1/c)E_i(c)I = I.$$
- $E_{ij}(d)$: The elementary operation of adding d times the jth row to the ith row is undone by applying $E_{ij}(-d)$. Hence
$$E_{ij}(-d)E_{ij}(d) = E_{ij}(-d)E_{ij}(d)I = I.$$ □

More examples of invertible matrices:

Diagonal Matrix Inverse

Example 2.37. Show that if D is a diagonal matrix with nonzero diagonal entries, then D is invertible.

Solution. Suppose that

$$D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}.$$

For a convenient shorthand, we write $D = \operatorname{diag}\{d_1, d_2, \ldots, d_n\}$. It is easily checked that if $E = \operatorname{diag}\{e_1, e_2, \ldots, e_n\}$, then

Diagonal Matrix Shorthand

$$DE = \operatorname{diag}\{d_1e_1, d_2e_2, \ldots, d_ne_n\} = \operatorname{diag}\{e_1d_1, e_2d_2, \ldots, e_nd_n\} = ED.$$

Therefore, if $d_i \neq 0$, for $i = 1, \ldots, n$, then

$$\operatorname{diag}\{d_1, d_2, \ldots, d_n\} \operatorname{diag}\{1/d_1, 1/d_2, \ldots, 1/d_n\} = \operatorname{diag}\{1, 1, \ldots, 1\} = I_n,$$

which shows that $\operatorname{diag}\{1/d_1, 1/d_2, \ldots, 1/d_n\}$ is an inverse of D. □

Laws of Inverses

Here are some of the basic laws of inverse calculations.

Laws of Matrix Inverses

Let A, B, C be matrices of the appropriate sizes so that the following multiplications make sense, I a suitably sized identity matrix, and c a nonzero scalar. Then

(1) (Uniqueness) If the matrix A is invertible, then it has only one inverse, which is denoted by A^{-1}.
(2) (Double Inverse) If A is invertible, then $\left(A^{-1}\right)^{-1} = A$.
(3) (2/3 Rule) If any two of the three matrices A, B, and AB are invertible, then so is the third, and moreover, $(AB)^{-1} = B^{-1}A^{-1}$.
(4) If A is invertible, then $(cA)^{-1} = (1/c)A^{-1}$.
(5) (Inverse/Transpose) If A is invertible, then $(A^T)^{-1} = (A^{-1})^T$ and $(A^*)^{-1} = (A^{-1})^*$.
(6) (Cancellation) Suppose A is invertible. If $AB = AC$ or $BA = CA$, then $B = C$.

Notes: Observe that the 2/3 rule reverses order when taking the inverse of a product. This should remind you of the operation of transposing a product. A common mistake is to forget to reverse the order. Secondly, notice that the cancellation law restores something that appeared to be lost when we first discussed matrices. Yes, we can cancel a common factor from both sides of an equation, but (1) the factor must be on the same side and (2) the factor must be an invertible matrix.

Verification of Laws: Suppose that both B and C work as inverses to the matrix A. We will show that these matrices must be identical. The associative and identity laws of matrices yield

$$B = BI = B(AC) = (BA)C = IC = C.$$

Henceforth, we shall write A^{-1} for the unique (two-sided) inverse of the square matrix A, provided of course that there is an inverse at all (remember that existence of inverses is not a sure thing).

Matrix Inverse Notation

The double inverse law is a matter of examining the definition of inverse:

$$AA^{-1} = I = A^{-1}A$$

shows that A is an inverse matrix for A^{-1}. Hence, $(A^{-1})^{-1} = A$.

Now suppose that A and B are both invertible and of the same size. Using the laws of matrix arithmetic, we see that

$$AB(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$$

and that

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

In other words, the matrix $B^{-1}A^{-1}$ works as an inverse for the matrix AB, which is what we wanted to show. We leave the remaining cases of the 2/3 rule as an exercise.

Suppose that c is nonzero and perform the calculation

$$(cA)(1/c)A^{-1} = (c/c)AA^{-1} = 1 \cdot I = I.$$

A similar calculation on the other side shows that $(cA)^{-1} = (1/c)A^{-1}$.

Next, apply the transpose operator to the definition of inverse (equation (2.15)) and use the law of transpose products to obtain that

$$(A^{-1})^T A^T = I^T = I = A^T (A^{-1})^T.$$

This shows that the definition of inverse is satisfied for $(A^{-1})^T$ relative to A^T, that is, that $(A^T)^{-1} = (A^{-1})^T$, which is the inverse/transpose law. The same argument works with conjugate transpose in place of transpose.

Finally, if A is invertible and $AB = AC$, then multiply both sides of this equation on the left by A^{-1} to obtain that

$$A^{-1}(AB) = (A^{-1}A)B = B = A^{-1}(AC) = (A^{-1}A)C = C,$$

which is the cancellation that we want. □

Negative Matrix Power

We can now extend the power notation to negative exponents. Let A be an invertible matrix and k a positive integer. Then we write

$$A^{-k} = A^{-1}A^{-1} \cdots A^{-1},$$

where the product is taken over k terms.

The laws of exponents that we saw earlier can now be expressed for arbitrary integers, *provided* that A is invertible. Here is an example of how we can use the various laws of arithmetic and inverses to carry out an inverse calculation.

Example 2.38. Let

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

Show that $(I - A)^3 = 0$ and use this to find A^{-1}.

Solution. First we calculate that

$$(I - A) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & -2 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

and check that

$$\begin{aligned} (I - A)^3 &= \begin{bmatrix} 0 & -2 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -2 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -2 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -2 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

Next we do some symbolic algebra, using the laws of matrix arithmetic:

$$0 = (I - A)^3 = (I - A)(I^2 - 2A + A^2) = I - 3A + 3A^2 - A^3.$$

Subtract all terms involving A from both sides to obtain that

$$3A - 3A^2 + A^3 = A \cdot 3I - 3A^2 + A^3 = A(3I - 3A + A^2) = I.$$

Since $A(3I - 3A + A^2) = (3I - 3A + A^2)A$, we see from definition of inverse that

$$A^{-1} = 3I - 3A + A^2 = \begin{bmatrix} 1 & -2 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}. \qquad \square$$

Notice that in the preceding example we were careful not to leave a "3" behind when we factored out A from $3A$. The reason is that $3+3A+A^2$ makes no sense as a sum, since one term is a scalar and the other two are matrices.

Rank and Inverse Calculation

Although we can calculate a few examples of inverses such as the last example, we really need a general procedure. So let's get right to the heart of the matter. How can we find the inverse of a matrix, or decide that none exists? Actually, we already have done all the hard work necessary to understand computing inverses. The secret is in the notions of reduced row echelon form and rank. (Remember, we use elementary row operations to reduce a matrix to its reduced row echelon form. Once we have done so, the rank of the matrix is simply the number of nonzero rows in the reduced row echelon form.) Let's recall the results of Example 2.24:

$$\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \end{bmatrix} = E_{12}(-1)E_2(-1/3)E_{21}(-2)E_1(1/4)E_{12} \begin{bmatrix} 2 & -1 & 1 \\ 4 & 4 & 20 \end{bmatrix}.$$

Now remove the last column from each of the matrices at the right of each side and we have this result:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = E_{12}(-1)E_2(-1/3)E_{21}(-2)E_1(1/4)E_{12} \begin{bmatrix} 2 & -1 \\ 4 & 4 \end{bmatrix}.$$

This suggests that if $A = \begin{bmatrix} 2 & -1 \\ 4 & 4 \end{bmatrix}$, then

$$A^{-1} = E_{12}(-1)E_2(-1/3)E_{21}(-2)E_1(1/4)E_{12}.$$

To prove this, we argue in the general case as follows: let A be an $n \times n$ matrix and suppose that by a succession of elementary row operations $E_1, E_2, \ldots, E_k$, we reduce A to its reduced row echelon form R, which happens to be I. In the language of matrix multiplication, what we have obtained is

$$I = E_k E_{k-1} \cdots E_1 A.$$

Now let $B = E_k E_{k-1} \cdots E_1$. By repeated application of the 2/3 rule, we see that a product of any number of invertible matrices is invertible. Since each elementary matrix is invertible, it follows that B is. Multiply both sides of the equation $I = BA$ by B^{-1} to obtain that $B^{-1}I = B^{-1} = B^{-1}BA = A$. Therefore, A is the inverse of the matrix B, hence is itself invertible.

Superaugmented Matrix

Here's a practical trick for computing this product of elementary matrices on the fly: form what we term the *superaugmented matrix* $[A \mid I]$. If we perform the elementary operation E on the superaugmented matrix, we have the same result as

$$E[A \mid I] = [EA \mid EI] = [EA \mid E].$$

So the matrix occupied by the I part of the superaugmented matrix is just the product of the elementary matrices that we have used so far. Now continue applying elementary row operations until the part of the matrix originally occupied by A is reduced to the reduced row echelon form of A. We end up with this schematic picture of our calculations:

$$\left[\, A \mid I \,\right] \overrightarrow{E_1, E_2, \ldots, E_k} \left[\, R \mid B \,\right],$$

where R is the reduced row echelon form of A and $B = E_k E_{k-1} \cdots E_1$ is the product of the various elementary matrices we used, composed in the correct order of usage. We can summarize this discussion with the following algorithm:

Inverse Algorithm

Given an $n \times n$ matrix A, to compute A^{-1}:
(1) Form the superaugmented matrix $\widetilde{A} = [A \mid I_n]$.
(2) Reduce the first n columns of $\widetilde{A}$ to reduced row echelon form by performing elementary operations on the matrix $\widetilde{A}$ resulting in the matrix $[R \mid B]$.
(3) If $R = I_n$ then set $A^{-1} = B$, otherwise, A is singular and A^{-1} does not exist.

Example 2.39. Use the inverse algorithm to compute the inverse of Example 2.8,

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

Solution. Notice that this matrix is already upper triangular. Therefore, as in Gaussian elimination, it is a bit more efficient to start with the bottom pivot and clear out entries above in reverse order. So we compute

$$[A \mid I_3] = \begin{bmatrix} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \xrightarrow[E_{23}(-1)]{} \begin{bmatrix} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \xrightarrow[E_{1,2}(-2)]{} \begin{bmatrix} 1 & 0 & 0 & 1 & -2 & 2 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

We conclude that A is indeed invertible and

$$A^{-1} = \begin{bmatrix} 1 & -2 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$ □

There is a simple formula for the inverse of a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Set $D = ad - bc$. It is easy to verify that if $D \neq 0$, then

Two by Two Inverse

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

Example 2.40. Use the 2×2 inverse formula to find the inverse of the matrix $A = \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}$, and verify that the same answer results if we use the inverse algorithm.

Solution. First we apply the inverse algorithm:

$$\begin{bmatrix} 1 & -1 & 1 & 0 \\ 1 & 2 & 0 & 1 \end{bmatrix} \xrightarrow[E_{21}(-1)]{} \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 3 & -1 & 1 \end{bmatrix} \xrightarrow[E3(1/3)]{} \begin{bmatrix} 1 & -1 & 1 & 0 \\ 0 & 1 & -1/3 & 1/3 \end{bmatrix}$$
$$\xrightarrow[E_{12}(1)]{} \begin{bmatrix} 1 & 0 & 2/3 & 1/3 \\ 0 & 1 & -1/3 & 1/3 \end{bmatrix}.$$

Thus we have found that
$$\begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix}^{-1} = \frac{1}{3} \begin{bmatrix} 2 & 1 \\ -1 & 1 \end{bmatrix}.$$
To apply the inverse formula, calculate $D = 1 \cdot 2 - 1 \cdot (-1) = 3$. Swap diagonal entries of A, negate the off-diagonal entries, and divide by D to get the same result as in the preceding equation for the inverse. □

The formula of the preceding example is well worth memorizing, since we will frequently need to find the inverse of a 2×2 matrix. Notice that in order for it to make sense, we have to have D nonzero. The number D is called the *determinant* of the matrix A. We will have more to say about this number in the next section. It is fairly easy to see why A must have $D \neq 0$ in order for its inverse to exist if we look ahead to the next theorem. Notice in the above elementary operation calculations that if $D = 0$ then elementary operations on A lead to a matrix with a row of zeros. Therefore, the rank of A will be smaller than 2. Here is a summary of our current knowledge of the invertibility of a square matrix.

Conditions for Invertibility

Theorem 2.7. The following are equivalent conditions on the square $n \times n$ matrix A:

(1) The matrix A is invertible.
(2) There is a square matrix B such that $BA = I$.
(3) The linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution for every right-hand-side vector $\mathbf{b}$.
(4) The linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution for some right-hand-side vector $\mathbf{b}$.
(5) The linear system $A\mathbf{x} = 0$ has only the trivial solution.
(6) $\operatorname{rank} A = n$.
(7) The reduced row echelon form of A is I_n.
(8) The matrix A is a product of elementary matrices.
(9) There is a square matrix B such that $AB = I$.

Proof. The method of proof is to show that each of conditions (1)–(7) implies the next, and that condition (8) implies (1). This connects (1)–(8) in a circle, so that any one condition will imply any other and therefore all are equivalent to each other. Finally, we show that (9) is equivalent to (1)–(8). Here is our chain of reasoning:

(1) implies (2): Assume A is invertible. Then the choice $B = A^{-1}$ certainly satisfies condition (2).

(2) implies (3): Assume (2) is true. Given a system $A\mathbf{x} = \mathbf{b}$, we can multiply both sides on the left by B to get that $B\mathbf{b} = BA\mathbf{x} = I\mathbf{x} = \mathbf{x}$. So there is only one solution, if any. On the other hand, if the system were inconsistent then we would have $\operatorname{rank} A < n$. By Corollary 2.2, $\operatorname{rank} BA < n$, contradicting the fact that $\operatorname{rank} I_n = n$. Hence, there is a solution, which proves (3).

(3) implies (4): This statement is obvious.

(4) implies (5): Assume (4) is true. Say the unique solution to $A\mathbf{x} = \mathbf{b}$ is $\mathbf{x}_0$. If the system $A\mathbf{x} = 0$ had a nontrivial solution, say $\mathbf{z}$, then we could add $\mathbf{z}$ to $\mathbf{x}_0$ to obtain a different solution $\mathbf{x}_0 + \mathbf{z}$ of the system $A\mathbf{x} = \mathbf{b}$ (check: $A(\mathbf{z} + \mathbf{x}_0) = A\mathbf{z} + A\mathbf{x}_0 = 0 + \mathbf{b} = \mathbf{b}$). This is impossible since (4) is true, so (5) follows.

(5) implies (6): Assume (5) is true. We know from Theorem 1.5 that the consistent system $A\mathbf{x} = 0$ has a unique solution precisely when the rank of A is n. Hence (6) must be true.

(6) implies (7): Assume (6) is true. The reduced row echelon form of A is the same size as A, that is, $n \times n$, and must have a row pivot entry 1 in every row. Also, the pivot entry must be the only nonzero entry in its column. This exactly describes the matrix I_n, so that (7) is true.

(7) implies (8): Assume (7) is true. We know that the matrix A is reduced to its reduced row echelon form by applying a sequence of elementary operations, or what amounts to the same thing, by multiplying the matrix A on the left by elementary matrices $E_1, E_2, \ldots, E_k$, say. Then $E_1 E_2 \cdots E_k A = I$. But we know from Example 2.36 that each elementary matrix is invertible and that their inverses are themselves elementary matrices. By successive multiplications on the left we obtain that $A = E_k^{-1} E_{k-1}^{-1} \cdots E_1^{-1} I$, showing that A is a product of elementary matrices, which is condition (8).

(8) implies (1): Assume (8) is true. Repeated application of the 2/3 rule shows that the product of any number of invertible matrices is itself invertible. Since elementary matrices are invertible, condition (1) must be true.

(9) is equivalent to (1): Assume (1) is true. Then A is invertible and the choice $B = A^{-1}$ certainly satisfies condition (9). Conversely, if (9) is true, then $I^T = I = (AB)^T = B^T A^T$, so that A^T satisfies (2), which is equivalent to (1). However, we already know that if a matrix is invertible, so is its transpose (Law (5) of Matrix Inverses), so $\left(A^T\right)^T = A$ is also invertible, which is condition (1). □

Notice that Theorem 2.7 relieves us of the responsibility of checking that a square one-sided inverse of a square matrix is a two-sided inverse: this is now automatic in view of conditions (2) and (9). Another interesting consequence of this theorem that has been found to be useful is an either/or statement, so it will always have something to say about any square linear system. This type of statement is sometimes called a *Fredholm alternative*. Many theorems go by this name, and we'll state another one in Chapter 5. Notice that a matrix is not invertible if and only if one of the conditions of the theorem fails. Certainly it is true that a square matrix is either invertible or not invertible. That's all the Fredholm alternative really says, but it uses the equivalent conditions (3) and (5) of Theorem 2.7 to say it in a different way:

Corollary 2.3. Given a square linear system $A\mathbf{x} = \mathbf{b}$, either the system has a unique solution for every right-hand-side vector $\mathbf{b}$ or there is a nonzero solution $\mathbf{x} = \mathbf{x}_0$ to the homogeneous system $A\mathbf{x} = \mathbf{0}$. Fredholm Alternative

We conclude this section with an application to the problem of solving nonlinear equations. Although we focus on two equations in two unknowns, the same ideas can be extended to any number of equations in as many unknowns.

Recall from calculus that we could solve the one-variable equation $f(x) = 0$ for a solution point x_1 at which $f(x_1) = 0$ from a given "nearby" point x_0 by setting $dx = x_1 - x_0$, and assuming that the change in f is

$$\begin{aligned}\Delta f &= f(x_1) - f(x_0) = 0 - f(x_0)\\ &\approx df = f'(x_0)\,dx = f'(x_0)(x_1 - x_0).\end{aligned}$$

Now solve for x_1 in the equation $-f(x_0) = f'(x_0)(x_1 - x_0)$ and get the equation

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

Replace 1 by $n+1$ and 0 by n to obtain the famous Newton formula

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{2.2}$$

The idea is to start with x_0, use the formula to get x_1 and if $f(x_1)$ is not close enough to 0, then repeat the calculation with x_1 in place of x_0, and so forth until a satisfactory value of $x = x_n$ is reached. How does this relate to a two-variable problem? We illustrate the basic idea in two variables.

Newton's Method for Systems

Example 2.41. Describe concisely an algorithm analogous to Newton's method in one variable to solve the two-variable problem

$$\begin{aligned}x^2 + \sin(\pi xy) &= 1\\ x + y^2 + e^{x+y} &= 3.\end{aligned}$$

Solution. Our problem can be written as a system of two (nonlinear) equations in two unknowns, namely

$$\begin{aligned}f(x,y) &= x^2 + \sin(\pi xy) - 1 = 0\\ g(x,y) &= x + y^2 + e^{x+y} - 3 = 0.\end{aligned}$$

Now we can pull the same trick with differentials as in the one-variable problem by setting $dx = x_1 - x_0$, $dy = y_1 - y_0$, where $f(x_1, y_1) = 0$, approximating the change in both f and g by total differentials, and recalling the definition of these total differentials in terms of partial derivatives. This leads to the system

$$\begin{aligned}f_x(x_0,y_0)\ dx + f_y(x_0,y_0)\ dy &= -f((x_0,y_0))\\ g_x(x_0,y_0)\ dx + g_y(x_0,y_0)\ dy &= -g((x_0,y_0)).\end{aligned}$$

Next, write everything in vector style, say

$$\mathbf{F}(\mathbf{x}) = \begin{bmatrix} f(\mathbf{x})\\ g(\mathbf{x})\end{bmatrix}, \quad \mathbf{x}^{(0)} = \begin{bmatrix} x_0\\ y_0\end{bmatrix}, \quad \mathbf{x}^{(1)} = \begin{bmatrix} x_1\\ y_1\end{bmatrix}.$$

Now we can write the *vector* differentials in the forms

$$d\mathbf{F} = \begin{bmatrix} df\\ dg\end{bmatrix} \quad \text{and } d\mathbf{x} = \begin{bmatrix} dx\\ dy\end{bmatrix} = \begin{bmatrix} x_1 - x_0\\ y_1 - x_0\end{bmatrix} = \mathbf{x}^{(1)} - \mathbf{x}^{(0)}.$$

The original Newton equations now look like a matrix multiplication involving $d\,\mathbf{x}$, $\mathbf{F}$, and a matrix of derivatives of $\mathbf{F}$, namely the so-called Jacobian matrix

$$J_{\mathbf{F}}\left(x_0, y_0\right) = \begin{bmatrix} f_x\left(\left(x_0, y_0\right)\right) & f_y\left(\left(x_0, y_0\right)\right) \\ g_x\left(\left(x_0, y_0\right)\right) & g_y\left(\left(x_0, y_0\right)\right) \end{bmatrix}.$$

Specifically, we see from the definition of matrix multiplication that the Newton equations are equivalent to the vector equations

$$d\mathbf{F} = J_{\mathbf{F}}(\mathbf{x}_0)\,d\,\mathbf{x} = -\mathbf{F}\left(\mathbf{x}^{(0)}\right).$$

If the Jacobian matrix is invertible, then

$$\mathbf{x}^{(1)} - \mathbf{x}^{(0)} = J_{\mathbf{F}}\left(\mathbf{x}^{(0)}\right)^{-1}\mathbf{F}\left(\mathbf{x}^{(0)}\right),$$

whence by adding $\mathbf{x}_0$ to both sides we see that

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - J_{\mathbf{F}}\left(\mathbf{x}^{(0)}\right)^{-1}\mathbf{F}\left(\mathbf{x}^{(0)}\right).$$

Now replace 1 by $n+1$ and 0 by n to obtain the famous Newton formula in vector form:

Newton's Formula in Vector Form

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} - J_{\mathbf{F}}\left(\mathbf{x}^{(n)}\right)^{-1}\mathbf{F}\left(\mathbf{x}^{(n)}\right).$$

This beautiful analogy to the Newton formula of (2.2)needs the language and algebra of vectors and matrices. One can now calculate the Jacobian for our particular$\mathbf{F}\left(\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]\right)$ and apply this formula. We leave the details as an exercise. □

2.5 Exercises and Problems

Exercise 1. Find the inverse or show that it does not exist.

(a) $\begin{bmatrix} 1 & -2 & 1 \\ 0 & 2 & 0 \\ -1 & 0 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & \mathrm{i} \\ 0 & 4 \end{bmatrix}$ (c) $\begin{bmatrix} 2 & -2 & 1 \\ 0 & 2 & 0 \\ 2 & 0 & 1 \end{bmatrix}$ (d) $\begin{bmatrix} 2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ (e) $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

Exercise 2. Find the inverse or show that it does not exist.

(a) $\begin{bmatrix} 1 & 3 & 0 \\ 0 & 4 & 10 \\ 9 & 3 & 0 \end{bmatrix}$ (b) $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & 1 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix}$ (e) $\begin{bmatrix} \mathrm{i}+1 & 0 \\ 1 & \mathrm{i} \end{bmatrix}$

Exercise 3. Express the following systems in matrix form and solve by inverting the coefficient matrix of the system.

(a) $\begin{aligned} 2x + 3y &= 7 \\ x + 2y &= -2 \end{aligned}$ (b) $\begin{aligned} 3x_1 + 6x_2 - x_3 &= -4 \\ -2x_1 + x_2 + x_3 &= 3 \\ x_3 &= 1 \end{aligned}$ (c) $\begin{aligned} x_1 + x_2 &= -2 \\ 5x_1 + 2x_2 &= 5 \end{aligned}$

Exercise 4. Solve the following systems by matrix inversion.

(a) $\begin{aligned} 2x_1 + 3x_2 &= 7 \\ x_2 + x_3 &= 1 \\ x_2 - x_3 &= 1 \end{aligned}$ (b) $\begin{aligned} x_1 + 6x_2 - x_3 &= 4 \\ x_1 + x_2 &= 0 \\ x_2 &= 1 \end{aligned}$ (c) $\begin{aligned} x_1 - x_2 &= 2 \\ x_1 + 2x_2 &= 11 \end{aligned}$

Exercise 5. Express inverses of the following matrices as products of elementary matrices using the notation of elementary matrices.

(a) $\begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}$ (c) $\begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$ (e) $\begin{bmatrix} -1 & 0 \\ \mathrm{i} & 3 \end{bmatrix}$

Exercise 6. Show that the following matrices are invertible by expressing them as products of elementary matrices.

(a) $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ (d) $\begin{bmatrix} -1 & 0 \\ 3 & 3 \end{bmatrix}$ (e) $\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$

Exercise 7. Find $A^{-1}C$ if $A = \begin{bmatrix} 1 & 2 & -3 \\ 0 & -1 & 1 \\ 2 & 5 & -6 \end{bmatrix}$ and $C = \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & -1 & 1 & 1 \\ 2 & 0 & -6 & 0 \end{bmatrix}$.

Exercise 8. Solve $AX = B$ for X, where $A = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 1 & 0 & -2 \\ 2 & -1 & 1 & 1 \end{bmatrix}$.

Exercise 9. Verify the matrix law $\left(A^T\right)^{-1} = \left(A^{-1}\right)^T$ with $A = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix}$.

Exercise 10. Verify the matrix law $(A^*)^{-1} = \left(A^{-1}\right)^*$ with $A = \begin{bmatrix} 2 & 1 - 2\mathrm{i} \\ 0 & \mathrm{i} \end{bmatrix}$.

Exercise 11. Verify the matrix law $(AB)^{-1} = B^{-1}A^{-1}$ in the case that $A = \begin{bmatrix} 1 & 2 & -3 \\ 1 & 0 & 1 \\ 2 & 4 & -2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 & 2 \\ 0 & -3 & 1 \\ 0 & 0 & 1 \end{bmatrix}$.

Exercise 12. Verify the matrix law $(cA)^{-1} = (1/c)\, A^{-1}$ in the case that $A = \begin{bmatrix} 1 & 2 - \mathrm{i} & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ and $c = 2 + \mathrm{i}$.

Exercise 13. Determine for what values of k the following matrices are invertible and find the inverse in that case.

(a) $\begin{bmatrix} 1 & k \\ 0 & -1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ k & 0 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -6 & 0 \\ 0 & 0 & 0 & k \end{bmatrix}$

Exercise 14. Determine the inverses for the following matrices in terms of the parameter c and conditions on c for which the matrix has an inverse.

(a) $\begin{bmatrix} 1 & 2 \\ c & -1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 2 & c+1 \\ 0 & 1 & 1 \\ 0 & 0 & c \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 0 & c+\mathrm{i} \\ 0 & -1 & 0 \\ 0 & c & c \end{bmatrix}$

Exercise 15. Give a 2×2 example showing that the sum of invertible matrices need not be invertible.

Exercise 16. Give a 2×2 example that the sum of singular matrices need not be singular.

Exercise 17. Problem 26 of Section 2.2 yields a formula for the inverse of the matrix $I-N$, where N is nilpotent, namely, $(I-N)^{-1} = I+N+N^2+\cdots+N^k$. Apply this formula to matrices (a) $\begin{bmatrix} 1 & -1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ and (b) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$.

Exercise 18. If a matrix can be written as $A = D(I-N)$, where D is diagonal with nonzero entries and N is nilpotent, then $A^{-1} = (I-N)^{-1} D^{-1}$. Use this fact and the formulas of Exercise 17 and Example 2.37 to find inverses of the matrices (a) $\begin{bmatrix} 2 & 2 & 4 \\ 0 & 2 & -2 \\ 0 & 0 & 3 \end{bmatrix}$ and (b) $\begin{bmatrix} 2 & 0 \\ \mathrm{i} & 3 \end{bmatrix}$.

Exercise 19. Solve the nonlinear system of equations of Example 2.41 by using nine iterations of the vector Newton formula (2.5), starting with the initial guess $\mathbf{x}^{(0)} = (0,1)$. Evaluate $F\left(\mathbf{x}^{(9)}\right)$.

Exercise 20. Find the minimum value of the function $F(x,y) = \left(x^2+y+1\right)^2 + x^4 + y^4$ by using the Newton method to find critical points of the function $F(x,y)$, i.e., points where $f(x,y) = F_x(x,y) = 0$ and $g(x,y) = F_y(x,y) = 0$.

*Problem 21. Show from the definition that if a square matrix A satisfies the equation $A^3 - 2A + 3I = 0$, then the matrix A must be invertible.

Problem 22. Verify directly from the definition of inverse that the two by two inverse formula gives the inverse of a 2×2 matrix.

Problem 23. Assume that the product of invertible matrices is invertible and deduce that if A and B are invertible matrices of the same size and both B and AB are invertible, then so is A.

*Problem 24. Let A be an invertible matrix. Show that if the product of matrices AB is defined, then $\operatorname{rank}(AB) = \operatorname{rank}(B)$, and if BA is defined, then $\operatorname{rank}(BA) = \operatorname{rank}(B)$.

Problem 25. Prove that if $D = ABC$, where A, C, and D are invertible matrices, then B is invertible.

Problem 26. Given that $C = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ in block form with A and B square, show that C is invertible if and only if A and B are, in which case $C^{-1} = \begin{bmatrix} A^{-1} & 0 \\ 0 & B^{-1} \end{bmatrix}$.

Problem 27. Let T be an upper triangular matrix, say $T = D + M$, where D is diagonal and M is strictly upper triangular.
(a) Show that if D is invertible, then $T = D(I - N)$, where $N = D^{-1}M$ is strictly upper triangular.
(b) Assume that D is invertible and use part (a) and Exercise 26 to obtain a formula for T^{-1} involving D and N.

Problem 28. Show that if the product of matrices BA is defined and A is invertible, then $\text{rank}(BA) = \text{rank}(B)$.

*Problem 29. Given the matrix $M = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$, where the blocks A and C are invertible matrices, find a formula for M^{-1} in terms of A, B, and C.

2.6 Basic Properties of Determinants

What Are They?

Many students have already had some experience with determinants and may have used them to solve square systems of equations. Why have we waited until now to introduce them? In point of fact, they are not really the best tool for solving systems. That distinction goes to Gaussian elimination. Were it not for the *theoretical* usefulness of determinants they might be consigned to a footnote in introductory linear algebra texts as a historical artifact of linear algebra.

To motivate determinants, consider Example 2.40. Something remarkable happened in that example. Not only were we able to find a formula for the inverse of a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, but we were able to compute a single number $D = ad - bc$ that told us whether A was invertible. The condition of noninvertibility, namely that $D = 0$, has a very simple interpretation: this happens exactly when one row of A is a multiple of the other, since the example showed that this is when elementary operations use the first row to zero out the second row. Can we extend this idea? Is there a single number that will tell us whether there are dependencies among the rows of the square matrix A that cause its rank to be smaller than its row size? The answer is yes. This is exactly what determinants were invented for. The concept of determinant is subtle and not intuitive, and researchers had to accumulate a large body of experience before they were able to formulate a "correct" definition for this number. There are alternative definitions of determinants, but the following will suit our purposes. It is sometimes referred to as "expansion down the first column."

Determinant

Definition 2.16. The *determinant* of a square $n \times n$ matrix $A = [a_{ij}]$ is the scalar quantity $\det A$ defined recursively as follows: if $n = 1$ then $\det A = a_{11}$; otherwise, we suppose that determinants are defined for all square matrices of size less than n and specify that

$$\begin{aligned}\det A &= \sum_{k=1}^{n} a_{k1}(-1)^{k+1} M_{k1}(A) \\ &= a_{11}M_{11}(A) - a_{21}M_{21}(A) + \cdots + (-1)^{n+1} a_{n1} M_{n1}(A),\end{aligned}$$

where $M_{ij}(A)$ is the determinant of the $(n-1) \times (n-1)$ matrix obtained from A by deleting the ith row and jth column of A.

Caution: The determinant of a matrix A is a scalar number. It is *not* a matrix quantity.

Example 2.42. Describe the quantities $M_{21}(A)$ and $M_{22}(A)$, where

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

Solution. If we erase the second row and first column of A we obtain something like

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

Now collapse the remaining entries together to obtain the matrix

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

Therefore

$$M_{21}(A) = \det \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}.$$

Similarly, erase the second row and column of A to obtain

$$\begin{bmatrix} 2 & 0 \\ \\ 0 & 2 \end{bmatrix}.$$

Now collapse the remaining entries together to obtain

$$M_{22}(A) = \det \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}. \qquad \square$$

Now how do we calculate these determinants? Part (b) of the next example answers the question.

Example 2.43. Use the definition to compute the determinants of the following matrices.

$$\text{(a) } [-4] \qquad \text{(b) } \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad \text{(c) } \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & -1 \\ 0 & 1 & 2 \end{bmatrix}$$

Solution. (a) From the first part of the definition we see that

$$\det[-4] = -4.$$

For (b) we set $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and use the formula of the definition to obtain that

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = a_{11}M_{11}(A) - a_{21}M_{21}(A) = a \det [d] - c \det [b] = ad - cb.$$

This calculation gives a handy formula for the determinant of a 2×2 matrix. For (c) use the definition to obtain that

$$\begin{aligned} \det \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & -1 \\ 0 & 1 & 2 \end{bmatrix} &= 2 \det \begin{bmatrix} 1 & -1 \\ 1 & 2 \end{bmatrix} - 1 \det \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix} + 0 \det \begin{bmatrix} 1 & 0 \\ 1 & -1 \end{bmatrix} \\ &= 2(1 \cdot 2 - 1 \cdot (-1)) - 1(1 \cdot 2 - 1 \cdot 0) + 0(1 \cdot (-1) - 1 \cdot 0) \\ &= 2 \cdot 3 - 1 \cdot 2 + 0 \cdot (-1) \\ &= 4. \end{aligned}$$

A point worth observing here is that we didn't really have to calculate the determinant of any matrix if it is multiplied by a zero. Hence, the more zeros our matrix has, the easier we expect the determinant calculation to be! $\square$

Another common symbol for $\det A$ is $|A|$, which is also written with respect to the elements of A by suppressing matrix brackets:

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

This notation invites a certain oddity, if not abuse, of language: we sometimes refer to things like the "second row" or "$(2,3)$th element" or the "size" of the determinant. Yet the determinant is only a number and in that sense doesn't really have rows or entries or a size. Rather, it is the underlying matrix whose determinant is being calculated that has these properties. So be careful of this notation; we plan to use it frequently because it's handy, but you should bear in mind that determinants and matrices are *not* the same thing! Another reason that this notation can be tricky is the case of a one-dimensional matrix, say $A = [a_{11}]$. Here it is definitely *not* a good idea to forget the brackets, since we already understand $|a_{11}|$ to be the absolute value of the scalar a_{11}, a nonnegative number. In the 1×1 case use $|[a_{11}]|$ for the determinant, which is just the number a_{11} and may be positive or negative.

Minors and Cofactors

The number $M_{ij}(A)$ is called the (i,j)th *minor* of the matrix A. If we collect the sign term in the definition of determinant together with the minor we obtain the (i,j)th *cofactor* $A_{ij} = (-1)^{i+j} M(A)$ of the matrix A. In the terminology of cofactors,

$$\det A = \sum_{k=1}^{n} a_{k1} A_{k1}.$$

Laws of Determinants

Our primary goal here is to show that determinants have the magical property we promised: a matrix is singular exactly when its determinant is 0. Along the way we will examine some useful properties of determinants. There is a lot of clever algebra that can be done here; we will try to keep matters straightforward (if that's possible with determinants). In order to focus on the main ideas, we place most of the proofs of key facts in the last section for optional reading. Also, a concise summary of the basic determinantal laws is given at the end of this section. Unless otherwise stated, we assume throughout this section that matrices are square, and that $A = [a_{ij}]$ is an $n \times n$ matrix.

For starters, let's observe that it's very easy to calculate the determinant of upper triangular matrices. Let A be such a matrix. Then $a_{k1} = 0$ if $k > 1$, so

$$\det A = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{vmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{vmatrix}$$
$$= \cdots = a_{11} \cdot a_{22} \cdots a_{nn}.$$

Hence we have established our first determinantal law:

D1: If A is an upper triangular matrix, then the determinant of A is the product of all the diagonal elements of A.

Example 2.44. Compute $D = \begin{vmatrix} 4 & 4 & 1 & 1 \\ 0 & -1 & 2 & 3 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 0 & 2 \end{vmatrix}$ and $|I_n| = \det I_n$.

Solution. By D1 we can do this at a glance: $D = 4 \cdot (-1) \cdot 2 \cdot 2 = -16$. Since I_n is diagonal, it is certainly upper triangular. Moreover, the entries down the diagonal of this matrix are 1's, so D1 implies that $|I_n| = 1$. □

Next, suppose that we notice a common factor of the scalar c in a row, say for convenience, the first one. How does this affect the determinantal calculation? In the case of a 1×1 determinant, we could simply factor it out of the original determinant. The general situation is covered by this law:

D2: If B is obtained from A by multiplying one row of A by the scalar c, then $\det B = c \cdot \det A$.

Here is a simple illustration:

Example 2.45. Compute $D = \begin{vmatrix} 5 & 0 & 10 \\ 5 & 5 & 5 \\ 0 & 0 & 2 \end{vmatrix}$.

Solution. Put another way, D2 says that scalars may be factored out of individual rows of a determinant. So use D2 on the first and second rows and then use the definition of determinant to obtain

$$\begin{aligned} \begin{vmatrix} 5 & 0 & 10 \\ 5 & 5 & 5 \\ 0 & 0 & 2 \end{vmatrix} &= 5 \cdot \begin{vmatrix} 1 & 0 & 2 \\ 5 & 5 & 5 \\ 0 & 0 & 2 \end{vmatrix} = 5 \cdot 5 \cdot \begin{vmatrix} 1 & 0 & 2 \\ 1 & 1 & 1 \\ 0 & 0 & 2 \end{vmatrix} = 25 \cdot \left(1 \cdot \begin{vmatrix} 1 & 1 \\ 0 & 2 \end{vmatrix} - 1 \cdot \begin{vmatrix} 0 & 2 \\ 0 & 2 \end{vmatrix} + 0 \cdot \begin{vmatrix} 0 & 2 \\ 1 & 1 \end{vmatrix} \right) \\ &= 50. \end{aligned}$$

One can easily check that this is the same answer we get by working the determinant directly from the definition. □

Next, suppose we interchange two rows of a determinant.

D3: If B is obtained from A by interchanging two rows of A, then $\det B = -\det A$.

Example 2.46. Use D3 to show the following handy fact: if a determinant has a repeated row, then it must be 0.

Solution. Suppose that the ith and jth rows of the matrix A are identical, and B is obtained by switching these two rows of A. Clearly $B = A$. Yet, according to D3, $\det B = -\det A$. It follows that $\det A = -\det A$, i.e., if we add $\det A$ to both sides, $2 \cdot \det A = 0$, so that $\det A = 0$, which is what we wanted to show. □

What happens to a determinant if we add a multiple of one row to another?

D4: If B is obtained from A by adding a multiple of one row of A to another row of A, then $\det B = \det A$.

Example 2.47. Compute $D = \begin{vmatrix} 1 & 4 & 1 & 1 \\ 1 & -1 & 2 & 3 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 1 & 2 \end{vmatrix}$.

Solution. What D4 really says is that any elementary row operation $E_{ij}(c)$ can be applied to the matrix behind a determinant and the determinant will be unchanged. So in this case, add -1 times the first row to the second and $-\frac{1}{2}$ times the third row to the fourth, then apply D1 to obtain

$$\begin{vmatrix} 1 & 4 & 1 & 1 \\ 1 & -1 & 2 & 3 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 1 & 2 \end{vmatrix} = \begin{vmatrix} 1 & 4 & 1 & 1 \\ 0 & -5 & 1 & 2 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 0 & 1/2 \end{vmatrix} = 1 \cdot (-5) \cdot 2 \cdot \frac{1}{2} = -5.$$ □

Example 2.48. Use D3 to show that a matrix with a row of zeros has zero determinant.

Solution. Suppose A has a row of zeros. Add any other row of the matrix A to this zero row to obtain a matrix B with repeated rows. □

We now have enough machinery to establish the most important property of determinants. First of all, we can restate laws D2–D4 in the language of elementary matrices as follows:

Determinant of Elementary Matrices

- D2: $\det(E_i(c)A) = c \cdot \det A$ (remember that for $E_i(c)$ to be an elementary matrix, $c \neq 0$).
- D3: $\det(E_{ij}A) = -\det A$.
- D4: $\det(E_{ij}(s)A) = \det A$.

Apply a sequence of elementary row operations on the $n \times n$ matrix A to reduce it to its reduced row echelon form R, or equivalently, multiply A on the left by elementary matrices $E_1, E_2, \ldots, E_k$ and obtain

$$R = E_1 E_2 \cdots E_k A.$$

Take the determinant of both sides to obtain

$$\det R = \det(E_1 E_2 \cdots E_k A) = \pm(\text{nonzero constant}) \cdot \det A.$$

Therefore, $\det A = 0$ precisely when $\det R = 0$. Now the reduced row echelon form of A is certainly upper triangular. In fact, it is guaranteed to have zeros on the diagonal, and therefore have zero determinant by D1, unless $\operatorname{rank} A = n$, in which case $R = I_n$. According to Theorem 2.7 this happens precisely when A is invertible. Thus:

D5: The matrix A is invertible if and only if $\det A \neq 0$.

Example 2.49. Determine whether the following matrices are invertible without actually finding the inverse.

$$\text{(a)} \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & -1 \\ 0 & 1 & 2 \end{bmatrix} \qquad \text{(b)} \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

Solution. Compute the determinants:

$$\begin{vmatrix} 2 & 1 & 0 \\ 1 & 1 & -1 \\ 0 & 1 & 2 \end{vmatrix} = 2 \begin{vmatrix} 1 & -1 \\ 1 & 2 \end{vmatrix} - 1 \begin{vmatrix} 1 & 0 \\ 1 & 2 \end{vmatrix} = 2 \cdot 3 - 2 = 4,$$

$$\begin{vmatrix} 2 & 1 & 0 \\ 1 & 1 & -1 \\ 0 & -1 & 2 \end{vmatrix} = 2 \begin{vmatrix} 1 & -1 \\ -1 & 2 \end{vmatrix} - 1 \begin{vmatrix} 1 & 0 \\ -1 & 2 \end{vmatrix} = 2 \cdot 1 - 1 \cdot 2 = 0.$$

Hence by D5, matrix (a) is invertible and matrix (b) is not invertible. □

There are two more surprising properties of determinants that we now discuss. Their proofs involve using determinantal properties of elementary matrices (see the next section for details).

D6: Given matrices A, B of the same size,

$$\det AB = \det A \det B.$$

Example 2.50. Verify D6 in the case that $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}$. How do $\det(A + B)$ and $\det A + \det B$ compare in this case?

Solution. We have easily that $\det A = 1$ and $\det B = 2$. Therefore, $\det A + \det B = 1 + 2 = 3$, while $\det A \cdot \det B = 1 \cdot 2 = 2$. On the other hand,

$$AB = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 4 & 3 \end{bmatrix},$$
$$A + B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix},$$

so that $\det AB = 2 \cdot 3 - 4 \cdot 1 = 2 = \det A \cdot \det B$, as expected. On the other hand, we have that $\det(A + B) = 3 \cdot 2 - 1 \cdot 1 = 5 \neq \det A + \det B$. □

This example raises a very important point.

Caution: In general, $\det A + \det B \neq \det(A + B)$, though there are occasional exceptions.

In other words, determinants do not distribute over sums. (It is true, however, that the determinant is additive in *one row at a time*. See the proof of D4 for details.)

Finally, we ask how $\det A^T$ compares to $\det A$. Simple cases suggest that there is no difference in determinant. This is exactly what happens in general.

D7: For all square matrices A, $\det A^T = \det A$.

Example 2.51. Compute $D = \begin{vmatrix} 4 & 0 & 0 & 0 \\ 4 & 1 & 0 & 0 \\ 1 & 2 & -2 & 0 \\ 1 & 0 & 1 & 2 \end{vmatrix}$.

Solution. By D7 and D1 we see immediately that $D = 4 \cdot 1 \cdot (-2) \cdot 2 = -16$. □

D7 is a very useful fact. Let's look at it from this point of view: transposing a matrix interchanges the rows and columns of the matrix. Therefore, everything that we have said about rows of determinants applies equally well to the columns, *including the definition of determinant itself!* Therefore, we could have given the definition of determinant in terms of expanding across the first row instead of down the first column and gotten the same answers. Likewise, we could perform elementary column operations instead of row operations and get the same results as D2–D4. Furthermore, the determinant of a lower triangular matrix is the product of its diagonal elements thanks to D7+D1. By interchanging rows or columns then expanding by first row or column, we see that the same effect is obtained by simply expanding the determinant down any column or across any row. We have to alternate signs starting with the sign $(-1)^{i+j}$ of the first term we use.

Now we can really put it all together and compute determinants to our heart's content with a good deal less effort than the original definition specified. We can use D1–D4 in particular to make a determinant calculation no worse than Gaussian elimination in the amount of work we have to do. We simply reduce a matrix to triangular form by elementary operations, then take the product of the diagonal terms.

Example 2.52. Calculate $D = \begin{vmatrix} 3 & 0 & 6 & 6 \\ 1 & 0 & 2 & 1 \\ 2 & 0 & 0 & 1 \\ -1 & 2 & 0 & 0 \end{vmatrix}$.

Solution. We are going to do this calculation two ways. We may as well use the same elementary operation notation that we have employed in Gaussian elimination. The only difference is that we have equality instead of arrows, provided that we modify the value of the new determinant in accordance with the laws D1–D3. So here is the straightforward method:

$$D = 3 \begin{vmatrix} 1 & 0 & 2 & 2 \\ 1 & 0 & 2 & 1 \\ 2 & 0 & 0 & 1 \\ -1 & 2 & 0 & 0 \end{vmatrix} \underset{\substack{E_{21}(-1) \\ E_{31}(-2) \\ E_{41}(1)}}{=} 3 \begin{vmatrix} 1 & 0 & 2 & 2 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -4 & -3 \\ 0 & 2 & 2 & 2 \end{vmatrix} \underset{E_{24}}{=} -3 \begin{vmatrix} 1 & 0 & 2 & 2 \\ 0 & 2 & 2 & 2 \\ 0 & 0 & -4 & -3 \\ 0 & 0 & 0 & -1 \end{vmatrix} = -24.$$

Here is another approach: let's expand the determinant down the second column, since it is mostly 0's. Remember that the sign in front of the first minor must be $(-1)^{1+2} = -1$. Also, the coefficients of the first three minors are 0, so we need only write down the last one in the second column:

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

Expand down the second column again:

$$D = 2\left(-6\begin{vmatrix} 1 & 1 \\ 2 & 1 \end{vmatrix} + 2\begin{vmatrix} 3 & 6 \\ 2 & 1 \end{vmatrix}\right) = 2(-6 \cdot (-1) + 2 \cdot (-9)) = -24. \qquad \square$$

An Inverse Formula

Let $A = [a_{ij}]$ be an $n \times n$ matrix. We have already seen that we can expand the determinant of A down any column of A (see the discussion following Example 2.51). These expansions lead to cofactor formulas for each column number j:

$$\det A = \sum_{k=1}^{n} a_{kj} A_{kj} = \sum_{k=1}^{n} A_{kj} a_{kj}.$$

This formula resembles a matrix multiplication formula. Consider the slightly altered sum

$$\sum_{k=1}^{n} A_{ki} a_{kj} = A_{1i} a_{1j} + A_{2i} a_{2j} + \cdots + A_{ni} a_{nj}.$$

The key to understanding this expression is to realize that it is exactly what we would get if we replaced the ith column of the matrix A by its jth column and then computed the determinant of the resulting matrix by expansion down the ith column. But such a matrix has two equal columns and therefore has a zero determinant, which we can see by applying Example 2.46 to the transpose of the matrix and using D7. So this sum must be 0 if $i \neq j$. We can combine these two sums by means of the Kronecker delta ($\delta_{ij} = 1$ if $i = j$ and 0 otherwise) in the formula

Kronecker Delta

$$\sum_{k=1}^{n} A_{ki} a_{kj} = \delta_{ij} \det A.$$

In order to exploit this formula we make the following definitions:

Adjoint, Minor, and Cofactor Matrices

Definition 2.17. The *matrix of minors* of the $n \times n$ matrix $A = [a_{ij}]$ is the matrix $M(A) = [M_{ij}(A)]$ of the same size. The *matrix of cofactors* of A is the matrix $A_{\text{cof}} = [A_{ij}]$ of the same size. Finally, the *adjoint matrix* of A is the matrix $\operatorname{adj} A = A_{\text{cof}}^T$.

Example 2.53. Compute the determinant, minors, cofactors, and adjoint matrices for $A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & -1 \\ 0 & 2 & 1 \end{bmatrix}$ and compute $A \operatorname{adj} A$.

Solution. The determinant is easily seen to be 2. Now for the matrix of minors:

$$M(A) = \begin{bmatrix} \begin{vmatrix} 0 & -1 \\ 2 & 1 \end{vmatrix} & \begin{vmatrix} 0 & -1 \\ 0 & 1 \end{vmatrix} & \begin{vmatrix} 0 & 0 \\ 0 & 2 \end{vmatrix} \\ \begin{vmatrix} 2 & 0 \\ 2 & 1 \end{vmatrix} & \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} & \begin{vmatrix} 1 & 2 \\ 0 & 2 \end{vmatrix} \\ \begin{vmatrix} 2 & 0 \\ 0 & -1 \end{vmatrix} & \begin{vmatrix} 1 & 0 \\ 0 & -1 \end{vmatrix} & \begin{vmatrix} 1 & 2 \\ 0 & 0 \end{vmatrix} \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 2 & 1 & 2 \\ -2 & -1 & 0 \end{bmatrix}.$$

To get the matrix of cofactors, simply overlay $M(A)$ with the following "checkerboard" of $+/-$'s $\begin{bmatrix} + & - & + \\ - & + & - \\ + & - & + \end{bmatrix}$ to obtain the matrix $A_{\text{cof}} = \begin{bmatrix} 2 & 0 & 0 \\ -2 & 1 & -2 \\ -2 & 1 & 0 \end{bmatrix}$. Now transpose A_{cof} to obtain

$$\operatorname{adj} A = \begin{bmatrix} 2 & -2 & -2 \\ 0 & 1 & 1 \\ 0 & -2 & 0 \end{bmatrix}.$$

We check that

$$A \operatorname{adj} A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & -1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 & -2 & -2 \\ 0 & 1 & 1 \\ 0 & -2 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} = (\det A) I_3. \qquad \square$$

Of course, the example simply confirms the formula that preceded it since this formula gives the (i, j)th entry of the product $(\operatorname{adj} A)A$. If we were to do determinants by row expansions, we would get a similar formula for the (i, j)th entry of $A \operatorname{adj} A$. We summarize this information in matrix notation as the following determinantal property:

D8: For a square matrix A,

Adjoint Formula

$$A \operatorname{adj} A = (\operatorname{adj} A)A = (\det A)I.$$

What does this have to do with inverses? We already know that A is invertible exactly when $\det A \neq 0$, so the answer is staring at us! Just divide the terms in D8 by $\det A$ to obtain an explicit formula for A^{-1}:

D9: For a square matrix A such that $\det A \neq 0$,

Inverse Formula

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A.$$

Example 2.54. Compute the inverse of the matrix A of Example 2.53 by the inverse formula.

Solution. We already computed the adjoint matrix of A, and the determinant of A is just 2, so we have that

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A = \frac{1}{2} \begin{bmatrix} 2 & -2 & -2 \\ 0 & 1 & 1 \\ 0 & -2 & 0 \end{bmatrix}. \qquad \square$$

Example 2.55. Interpret the inverse formula in the case of the 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$.

Solution. We have $M(A) = \begin{bmatrix} d & c \\ b & a \end{bmatrix}$, $A_{\text{cof}} = \begin{bmatrix} d & -c \\ -b & a \end{bmatrix}$ and $\operatorname{adj} A = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$, so that the inverse formula becomes

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

As you might expect, this is exactly the same as the formula we obtained in Example 2.40. $\square$

Cramer's Rule

Thanks to the inverse formula, we can now find an explicit formula for solving linear systems with a nonsingular coefficient matrix. Here's how we proceed. To solve $A\mathbf{x} = \mathbf{b}$ we multiply both sides on the left by A^{-1} to obtain that $\mathbf{x} = A^{-1}\mathbf{b}$. Now use the inverse formula to obtain

$$\mathbf{x} = A^{-1}\mathbf{b} = \frac{1}{\det A} \operatorname{adj}(A)\mathbf{b}.$$

The explicit formula for the ith coordinate of $\mathbf{x}$ that comes from this fact is

$$x_i = \frac{1}{\det A} \sum_{j=1}^{n} A_{ji} b_j.$$

The summation term is exactly what we would obtain if we started with the determinant of the matrix B_i obtained from A by replacing the ith column of A by $\mathbf{b}$ and then expanding the determinant down the ith column. Therefore, we have arrived at the following rule:

Cramer's Rule

Theorem 2.8. Let A be an invertible $n \times n$ matrix and $\mathbf{b}$ an $n \times 1$ column vector. Denote by B_i the matrix obtained from A by replacing the ith column of A by $\mathbf{b}$. Then the linear system $A\mathbf{x} = \mathbf{b}$ has unique solution $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, where

$$x_i = \frac{\det B_i}{\det A}, \quad i = 1, 2, \ldots, n.$$

Example 2.56. Use Cramer's rule to solve the system

$$\begin{aligned} 2x_1 - x_2 &= 1 \\ 4x_1 + 4x_2 &= 20. \end{aligned}$$

Solution. The coefficient matrix and right-hand-side vectors are

$$A = \begin{bmatrix} 2 & -1 \\ 4 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 1 \\ 20 \end{bmatrix},$$

so that

$$\det A = 8 - (-4) = 12,$$

and therefore

$$x_1 = \frac{\begin{vmatrix} 2 & 1 \\ 4 & 20 \end{vmatrix}}{\begin{vmatrix} 2 & -1 \\ 4 & 4 \end{vmatrix}} = \frac{36}{12} = 3 \quad \text{and} \quad x_2 = \frac{\begin{vmatrix} 1 & -1 \\ 20 & 4 \end{vmatrix}}{\begin{vmatrix} 2 & -1 \\ 4 & 4 \end{vmatrix}} = \frac{24}{12} = 2. \qquad \square$$

Summary of Determinantal Laws

Now that our list of the basic laws of determinants is complete, we record them in a concise summary.

Laws of Determinants

Let A, B be $n \times n$ matrices.

D1: If A is upper triangular, $\det A$ is the product of all the diagonal elements of A.
D2: $\det(E_i(c)A) = c \cdot \det A$.
D3: $\det(E_{ij}A) = -\det A$.
D4: $\det(E_{ij}(s)A) = \det A$.
D5: The matrix A is invertible if and only if $\det A \neq 0$.
D6: $\det AB = \det A \det B$.
D7: $\det A^T = \det A$.
D8: $A \operatorname{adj} A = (\operatorname{adj} A)A = (\det A)I$.
D9: If $\det A \neq 0$, then $A^{-1} = \dfrac{1}{\det A} \operatorname{adj} A$.

2.6 Exercises and Problems

Exercise 1. Compute all cofactors for these matrices.

(a) $\begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 4 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 1-\mathrm{i} \\ 0 & 1 \end{bmatrix}$

Exercise 2. Compute all minors for these matrices.

(a) $\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & -3 & 0 \\ -2 & 1 & 0 \\ 0 & -2 & 0 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & \mathrm{i}+1 \\ \mathrm{i} & 1 \end{bmatrix}$ (d) $\begin{bmatrix} 3 & 1 & -1 \\ 0 & 2 & -2 \\ 0 & 0 & 1 \end{bmatrix}$

Exercise 3. Compute these determinants. Which of the matrices represented are invertible?

(a) $\begin{vmatrix} 2 & -1 \\ 1 & 1 \end{vmatrix}$ (b) $\begin{vmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1+i \end{vmatrix}$ (c) $\begin{vmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 2 & 1 & 1 \end{vmatrix}$ (d) $\begin{vmatrix} 1 & -1 & 4 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 2 & 7 \\ -2 & 3 & 4 & 6 \end{vmatrix}$ (e) $\begin{vmatrix} -1 & -1 \\ 1 & 1-2\mathrm{i} \end{vmatrix}$

Exercise 4. Use determinants to determine which of these matrices are invertible.

(a) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ -2 & 3 & 4 & 6 \end{bmatrix}$ (b) $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 2 & 1 & 1 \\ 0 & 0 & 1 & 3 \\ 1 & 1 & 2 & 7 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 0 & 1 & 3 \end{bmatrix}$ (e) $\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$

Exercise 5. Verify by calculation that determinantal law D7 holds for the following choices of A.

(a) $\begin{bmatrix} -2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & -1 & 1 \\ 1 & 2 & 0 \\ -1 & 0 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 2 & 7 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix}$

Exercise 6. Let $A = B$ and verify by calculation that determinantal law D6 holds for the following choices of A.

(a) $\begin{bmatrix} -2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & -1 & 1 \\ 1 & 2 & 0 \\ -1 & 0 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 3 \\ -1 & 2 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 2 & 7 \end{bmatrix}$

Exercise 7. Use determinants to find conditions on the parameters in these matrices under which the matrices are invertible.

(a) $\begin{bmatrix} a & 1 \\ ab & 1 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 1 & -1 \\ 1 & c & 1 \\ 0 & 0 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}$

Exercise 8. Find conditions on the parameters in these matrices under which the matrices are invertible.

(a) $\begin{bmatrix} a & b & 0 & 0 \\ 0 & a & 0 & 0 \\ 0 & 0 & b & a \\ 0 & 0 & -a & b \end{bmatrix}$ (b) $\begin{bmatrix} \lambda-1 & 0 & 0 \\ 1 & \lambda-2 & 1 \\ 3 & 1 & \lambda-1 \end{bmatrix}$ (c) $\lambda I_2 - \begin{bmatrix} 0 & 1 \\ -c_0 & -c_1 \end{bmatrix}$

Exercise 9. For each of the following matrices calculate the adjoint matrix and the product of the matrix and its adjoint.

(a) $\begin{bmatrix} 2 & 1 & 0 \\ -1 & 1 & 2 \\ 1 & 2 & 2 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 3 \\ -1 & 2 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 2 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 2 & 6 \end{bmatrix}$

Exercise 10. For each of the following matrices calculate the adjoint matrix and the product of the adjoint and the matrix.

(a) $\begin{bmatrix} -1 & 1 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & 2 \end{bmatrix}$ (b) $\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & -1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 1+\mathrm{i} \\ 1-\mathrm{i} & 2 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & -3 \end{bmatrix}$

Exercise 11. Find the inverse of following matrices by adjoints.

(a) $\begin{bmatrix} 1 & 1 \\ 3 & 4 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 0 & 0 \\ 2 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & -2 & 2 \\ -1 & 2 & -1 \\ 1 & -3 & 1 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & \mathrm{i} \\ -2\mathrm{i} & 1 \end{bmatrix}$

Exercise 12. For each of the following matrices, find the inverse by superaugmented matrices and by adjoints.

(a) $\begin{bmatrix} 1 & 0 \\ 2 & 2 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & -1 & 3 \\ 2 & 2 & -4 \\ 1 & 1 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix}$

Exercise 13. Use Cramer's Rule to solve the following systems.

(a) $\begin{aligned} x - 3y &= 2 \\ 2x + y &= 11 \end{aligned}$ (b) $\begin{aligned} 2x_1 + x_2 &= b_1 \\ 2x_1 - x_2 &= b_2 \end{aligned}$ (c) $\begin{aligned} 3x_1 + x_3 &= 2 \\ 2x_1 + 2x_2 &= 1 \\ x_1 + x_2 + x_3 &= 6 \end{aligned}$

Exercise 14. Use Cramer's Rule to solve the following systems.

(a) $\begin{aligned} x + y + z &= 4 \\ 2x + 2y + 5z &= 11 \\ 4x + 6y + 8z &= 24 \end{aligned}$ (b) $\begin{aligned} x_1 - 2x_2 &= 2 \\ 2x_1 - x_2 &= 4 \end{aligned}$ (c) $\begin{aligned} x_1 + x_2 + x_3 &= 2 \\ x_1 + 2x_2 &= 1 \\ x_1 - x_3 &= 2 \end{aligned}$

Problem 15. Verify that $\begin{vmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & e & f \\ 0 & 0 & g & h \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} \begin{vmatrix} e & f \\ g & h \end{vmatrix}.$

Problem 16. Confirm that the determinant of the matrix $A = \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$ is -1. We can now assert without any further calculation that the inverse matrix of A has integer coefficients. Explain why in terms of laws of determinants.

*Problem 17. Let

$$V = \begin{bmatrix} 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \end{bmatrix}.$$

(Such a matrix is called a *Vandermonde* matrix.) Express $\det V$ as a product of factors $(x_j - x_k)$.

Problem 18. Show by example that $\det A^* \neq \det A$ and prove that in general $\det A^* = \overline{\det A}$.

*Problem 19. Use a determinantal law to show that $\det(A)\det\left(A^{-1}\right) = 1$ if A is invertible.

Problem 20. Use the determinantal laws to show that any matrix with a row of zeros has zero determinant.

*Problem 21. If A is a 5×5 matrix, then in terms of $\det(A)$, what can we say about $\det(-2A)$? Explain and express a law about a general matrix cA, c a scalar, that contains your answer.

Problem 22. Let A be a skew-symmetric matrix, that is, $A^T = -A$. Show that if A has odd order n, i.e., A is $n \times n$, then A must be singular.

*Problem 23. Show that if

$$M = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$$

then $\det M = \det A \cdot \det C$.

*Problem 24. Let J_n be the $n \times n$ counteridentity, that is, J_n is a square matrix with ones along the counterdiagonal (the diagonal that starts in the lower left corner and ends in the upper right corner), and zeros elsewhere. Find a formula for $\det J_n$. (Hint: show that $J_n^2 = I_n$, which narrows down $\det J_n$.)

Problem 25. Show that the *companion matrix* of the polynomial $f(x) = c_0 + c_1 x + \cdots + c_{n-1}x^{n-1} + x^n$, which is defined to be

$$C(f) = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ -c_0 & -c_1 & \cdots & -c_{n-2} & -c_{n-1} \end{bmatrix},$$

is invertible if and only if $c_0 \neq 0$.

Prove that if the matrix A is invertible, then $\det(A^T \dot{A}) > 0$.

Problem 26. Suppose that the square matrix A is singular. Prove that if the system $A\mathbf{x} = \mathbf{b}$ is consistent, then $(\operatorname{adj} A)\,\mathbf{b} = \mathbf{0}$.

2.7 *Computational Notes and Projects

LU Factorization

Here is a problem: suppose we want to solve a nonsingular linear system $Ax = b$ repeatedly, with different choices of b. A perfect example of this kind of situation is the heat flow problem Example 1.3, where the right-hand side is determined by the heat source term $f(x)$. Suppose that we need to experiment with different source terms. What happens if we do straight Gaussian elimination or Gauss–Jordan elimination? Each time we carry out a complete calculation on the augmented matrix $\widetilde{A} = [A \mid b]$ we have to resolve the whole system. Yet, the main part of our work is the same: putting the part of $\widetilde{A}$ corresponding to the coefficient matrix A into reduced row echelon form. Changing the right-hand side has no effect on this work. What we want here is a way to somehow record our work on A, so that solving a new system involves very little additional work. This is exactly what the LU factorization is all about.

Definition 2.18. Let A be an $n \times n$ matrix. An LU factorization of A is a pair of $n \times n$ matrices L, U such that

LU Factorization

(1) L is lower triangular.
(2) U is upper triangular.
(3) $A = LU$.

Even if we could find such beasts, what is so wonderful about them? The answer is that *triangular* systems $A\mathbf{x} = \mathbf{b}$ are easy to solve. For example, if A is upper triangular, we learned that the smart thing to do was to use the last equation to solve for the last variable, then the next-to-last equation for the next-to-last variable, etc. This is the secret of Gaussian elimination! But lower triangular systems are just as simple: use the first equation to solve for the first variable, the second equation for the second variable, and so forth. Now suppose we want to solve $A\mathbf{x} = \mathbf{b}$ and we know that $A = LU$. The original system becomes $LU\mathbf{x} = \mathbf{b}$. Introduce an intermediate variable $\mathbf{y} = U\mathbf{x}$. Now perform these steps:

1. (Forward solve) Solve lower triangular system $L\mathbf{y} = \mathbf{b}$ for the variable $\mathbf{y}$.
2. (Back solve) Solve upper triangular system $U\mathbf{x} = \mathbf{y}$ for the variable $\mathbf{x}$.

This does it! Once we have the matrices L, U, we don't have to worry about right-hand sides, except for the small amount of work involved in solving two triangular systems. Notice, by the way, that since A is assumed nonsingular, we have that if $A = LU$, then $\det A = \det L \det U \neq 0$. Therefore, neither triangular matrix L or U can have zeros on its diagonal. Thus, the forward and back solve steps can always be carried out to give a unique solution.

Example 2.57. You are given that

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

Use this fact to solve $Ax = b$, where $\mathbf{b} = [1, 0, 1]^T$ or $\mathbf{b} = [-1, 2, 1]^T$.

Solution. Set $\mathbf{x} = [x_1, x_2, x_3]^T$ and $\mathbf{y} = [y_1, y_2, y_3]^T$. For $\mathbf{b} = [1, 0, 1]^T$, forward solve

$$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

to get $y_1 = 1$, then $y_2 = 0 + 1y_1 = 1$, then $y_3 = 1 - 1y_1 - 2y_2 = -2$. Then back solve

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

to get $x_3 = -2/(-1) = 2$, then $x_2 = 1 + x_3 = 3$, then $x_1 = (1 - 1x_2)/2 = -1$.

For (b) forward solve

$$\begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$$

to get $y_1 = -1$, then $y_2 = 0 + 1y_1 = -1$, then $y_3 = 1 - 1y_1 - 2y_2 = 4$. Then back solve

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

to get $x_3 = 4/(-1) = -4$, then $x_2 = 1 + x_3 = -3$, then $x_1 = (1 - 1x_2)/2 = 2$.

□

Notice how simple the previous example was, given the LU factorization. Now how do we find such a factorization? In general, a nonsingular matrix may not have such a factorization. A good example is the matrix $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ However, if Gaussian elimination can be performed on the matrix A *without row exchanges*, then such a factorization is really a by-product of Gaussian elimination. In this case let $[a_{ij}^{(k)}]$ be the matrix obtained from A after using the kth pivot to clear out entries below it (thus $A = [a_{ij}^{(0)}]$). Remember that in Gaussian elimination we need only two types of elementary operations, namely row exchanges and adding a multiple of one row to another. Furthermore, the only elementary operations of the latter type that we use are of this form: $E_{ij}(-a_{jj}^{(k)}/a_{ij}^{(k)})$, where $[a_{ij}^{(k)}]$ is the matrix obtained from A from the various elementary operations up to this point. The numbers $m_{ij} = -a_{jj}^{(k)}/a_{ij}^{(k)}$, where $i > j$, are sometimes called *multipliers*. In the way of notation, let us call a triangular matrix a *unit* triangular matrix if its diagonal entries are all 1's.

Multipliers

Theorem 2.9. If Gaussian elimination is used without row exchanges on the nonsingular matrix A, resulting in the upper triangular matrix U, and if L is the unit lower triangular matrix whose entries below the diagonal are the negatives of the multipliers m_{ij}, then $A = LU$.

Proof. The proof of this theorem amounts to noticing that the product of all the elementary operations that reduces A to U is a unit lower triangular matrix $\widetilde{L}$ with the multipliers m_{ij} in the appropriate positions. Thus $\widetilde{L}A = U$. To undo these operations, multiply by a matrix L with the negatives of the multipliers in the appropriate positions. This results in

$$L\widetilde{L}A = A = LU$$

as desired. □

The following example shows how one can write an efficient program to implement LU factorization. The idea is this: as we do Gaussian elimination, the U part of the factorization gradually appears in the upper parts of the transformed matrices $A^{(k)}$. Below the diagonal we replace nonzero entries with zeros, column by column. Instead of wasting this space, use it to store the negative of the multipliers in place of the element it zeros out. Of course, this storage part of the matrix should not be changed by subsequent elementary row operations. When we are finished with elimination, the diagonal and upper part of the resulting matrix is just U, and the strictly lower triangular part on the unit lower triangular matrix L is stored in the lower part of the matrix.

Example 2.58. Use the shorthand of the preceding discussion to compute an LU factorization for

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

Solution. Proceed as in Gaussian elimination, but store negative multipliers:

$$\begin{bmatrix} 2 & 1 & 0 \\ -2 & 0 & -1 \\ 2 & 3 & -3 \end{bmatrix} \xrightarrow[E_{31}(-1)]{E_{21}(1)} \begin{bmatrix} 2 & 1 & 0 \\ -1 & 1 & -1 \\ 1 & 2 & -3 \end{bmatrix} \xrightarrow[E_{32}(-2)]{} \begin{bmatrix} 2 & 1 & 0 \\ -1 & 1 & -1 \\ 1 & 2 & -1 \end{bmatrix}.$$

Now we read off the results from the last matrix:

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -1 & 2 & 1 \end{bmatrix} \text{ and } U = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & -1 \end{bmatrix}.$$ □

What can be said if row exchanges are required (for example, we might want to use a partial pivoting strategy)? Take the point of view that we could see our way to the end of Gaussian elimination and store the product P of all row-exchanging elementary operations that we use along the way. A product of such matrices is called a *permutation matrix*; such a matrix is invertible, since

Permutation Matrix

it is a product of invertible matrices. Thus if we apply the correct permutation matrix P to A we obtain a matrix for which Gaussian elimination will succeed without further row exchanges. Consequently, we have a theorem that applies to all nonsingular matrices. Notice that it does not limit the usefulness of LU factorization since the linear system $Ax = b$ is equivalent to the system $PAx = Pb$. The following theorem could be called the "PLU factorization theorem."

Theorem 2.10. If A is a nonsingular matrix, then there exists a permutation matrix P, upper triangular matrix U, and unit lower triangular matrix L such that $PA = LU$.

There are many other useful factorizations of matrices that numerical analysts have studied, e.g., LDU and Cholesky. We will stop at LU, but there is one last point we want to make. The amount of work in finding the LU factorization is the same as Gaussian elimination itself, which is approximately $2n^3/3$ flops (see Section 1.5). The additional work of back and forward solving is about $2n^2$ flops. So the dominant amount of work is done by computing the factorization rather than the back and forward solving stages.

Efficiency of Determinants and Cramer's Rule in Computation

Computational Efficiency of Determinants

The truth of the matter is that Cramer's Rule and adjoints are good only for small matrices and theoretical arguments. For if you evaluate determinants in a straightforward way from the definition, the work in doing so is about $n \cdot n!$ flops for an $n \times n$ system. (Recall that a "flop" in numerical linear algebra is a single addition or subtraction, or multiplication or division.) For example, it is not hard to show that the operation of adding a multiple of one row vector of length n to another requires $2n$ flops. This number $n \cdot n!$ is vast when compared to the number $2n^3/3$ flops required for Gaussian elimination, even with "small" n, say $n = 10$. In this case we have $2 \cdot 10^3/3 \approx 667$, while $10 \cdot 10! = 36,288,000$.

On the other hand, there is a clever way to evaluate determinants that requires much less work than the definition: use elementary row operations together with D2, D6, and the elementary operations that correspond to these rules to reduce the determinant to that of a triangular matrix. This requires about $2n^3/3$ flops. As a matter of fact, it is tantamount to Gaussian elimination. But to use Cramer's Rule, you will have to calculate $n+1$ determinants. So why bother with Cramer's Rule on larger problems when it still will take about n times as much work as Gaussian elimination? A similar remark applies to computing adjoints instead of using Gauss–Jordan elimination on the superaugmented matrix of A.

Proofs of Some of the Laws of Determinants

D2: If B is obtained from A by multiplying one row of A by the scalar c, then $\det B = c \cdot \det A$.

To keep the notation simple, assume that the first row is multiplied by c, the proof being similar for other rows. Suppose we have established this for all determinants of size less than n (this is really another "proof by induction," which is how most of the following determinantal properties are established). For an $n \times n$ determinant we have

$$\begin{aligned}\det B &= \begin{vmatrix} c \cdot a_{11} & c \cdot a_{12} & \cdots & c \cdot a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \\ &= c \cdot a_{11} \begin{vmatrix} a_{22} & a_{23} & \cdots & a_{2n} \\ a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & & \vdots \\ a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix} + \sum_{k=2}^{n} a_{k1}(-1)^{k+1} M_{k1}(B).\end{aligned}$$

But the minors $M_{k1}(B)$ all are smaller and have a common factor of c in the first row. Pull this factor out of every remaining term and we get that

$$\begin{vmatrix} c \cdot a_{11} & c \cdot a_{12} & \cdots & c \cdot a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = c \cdot \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}.$$

Thus we have shown that property D2 holds for all matrices.

D3: If B is obtained from A by interchanging two rows of A, then $\det B = -\det A$.

To keep the notation simple, assume we switch the first and second rows. In the case of a 2×2 determinant, we get the negative of the original determinant (check this for yourself). Suppose we have established that the same is true for all matrices of size less than n. For an $n \times n$ determinant we have

$$\begin{aligned}\det B &= \begin{vmatrix} a_{21} & a_{22} & \cdots & a_{2n} \\ a_{11} & a_{12} & \cdots & a_{1n} \\ a_{31} & a_{32} & \cdots & a_{3n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \\ &= a_{21} M_{11}(B) - a_{11} M_{21}(B) + \sum_{k=3}^{n} a_{k1}(-1)^{k+1} M_{k1}(B) \\ &= a_{21} M_{21}(A) - a_{11} M_{11}(A) + \sum_{k=3}^{n} a_{k1}(-1)^{k+1} M_{k1}(B).\end{aligned}$$

But all the determinants in the summation sign come from a submatrix of A with the first and second rows interchanged. Since they are smaller than n, each is just the negative of the corresponding minor of A. Notice that the first two terms are just the first two terms in the determinantal expansion of A, except that they are out of order and have an extra minus sign. Factor this minus sign out of every term and we have obtained D3. $\square$

D4: If B is obtained from A by adding a multiple of one row of A to another row of A, then $\det B = \det A$.

Actually, it's a little easier to answer a slightly more general question: what happens if we replace a row of a determinant by that row plus some other row vector $\mathbf{r}$ (not necessarily a row of the determinant)? Again, simply for convenience of notation, we assume that the row in question is the first. The same argument works for any other row. Some notation: let B be the matrix that we obtain from the $n \times n$ matrix A by adding the row vector $\mathbf{r} = [r_1, r_2, \ldots, r_n]$ to the first row and C the matrix obtained from A by replacing the first row by $\mathbf{r}$. The answer turns out to be that $|B| = |A| + |C|$. So we can say that the determinant function is "additive in each row." Let's see what happens in the one dimensional case:

$$|B| = |[a_{11} + r_1]| = a_{11} + r_1 = |[a_{11}]| + |[r_1]| = |A| + |C|.$$

Suppose we have established that the same is true for all matrices of size less than n and let A be $n \times n$. Then the minors $M_{k1}(B)$, with $k > 1$, are smaller than n, so the property holds for them. Hence we have

$$\begin{aligned}
\det B &= \begin{vmatrix} a_{11}+r_1 & a_{12}+r_2 & \cdots & a_{1n}+r_n \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \\
&= (a_{11}+r_1)M_{11}(A) + \sum_{k=2}^{n} a_{k1}(-1)^{k+1}M_{k1}(B) \\
&= (a_{11}+r_1)M_{11}(A) + \sum_{k=2}^{n} a_{k1}(-1)^{k+1}(M_{k1}(A) + M_{k1}(C)) \\
&= \sum_{k=1}^{n} a_{k1}(-1)^{k+1}M_{k1}(A) + r_1 M_{11}(C) + \sum_{k=2}^{n} a_{k1}(-1)^{k+1}M_{k1}(C) \\
&= \det A + \det C.
\end{aligned}$$

Now what about adding a multiple of one row to another in a determinant? For notational convenience, suppose we add s times the second row to the first. In the notation of the previous paragraph,

$$\det B = \begin{vmatrix} a_{11}+s\cdot a_{21} & a_{12}+s\cdot a_{22} & \cdots & a_{1n}+s\cdot a_{2n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

and

$$\det C = \begin{vmatrix} s \cdot a_{21} & s \cdot a_{22} & \cdots & s \cdot a_{2n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = s \cdot \begin{vmatrix} a_{21} & a_{22} & \cdots & a_{2n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = 0,$$

where we applied D2 to pull the common factor s from the first row and the result of Example 2.46 to get the determinant with repeated rows to be 0. But $|B| = |A| + |C|$. Hence we have shown D4. □

D6: Given matrices A, B of the same size, $\det AB = \det A \det B$.

The key here is that we now know that determinant calculation is intimately connected with elementary matrices, rank, and the reduced row echelon form. First let's reinterpret D2–D4 still one more time. First of all take $A = I$ in the discussion of the previous paragraph, and we see that

- $\det E_i(c) = c$
- $\det E_{ij} = -1$
- $\det E_{ij}(s) = 1$

Therefore, D2–D4 can be restated (yet again) as

- D2: $\det(E_i(c)A) = \det E_i(c) \cdot \det A$ (here $c \neq 0$.)
- D3: $\det(E_{ij}A) = \det E_{ij} \cdot \det A$
- D4: $\det(E_{ij}(s) = \det E_{ij}(s) \cdot \det A$

In summary: For any elementary matrix E and arbitrary matrix A of the same size, $\det(EA) = \det(E)\det(A)$.

Now let's consider this question: how does $\det(AB)$ relate to $\det(A)$ and $\det(B)$? If A is not invertible, $\operatorname{rank} A < n$ by Theorem 2.7 and so $\operatorname{rank} AB < n$ by Corollary 2.2. Therefore, $\det(AB) = 0 = \det A \cdot \det B$ in this case. Next suppose that A is invertible. Express it as a product of elementary matrices, say $A = E_1 E_2 \cdots E_k$, and use our summary of D1–D3 to disassemble and reassemble the elementary factors:

$$\begin{aligned} \det(AB) &= \det(E_1 E_2 \cdots E_k B) = (\det E_1 \det E_2 \cdots \det E_k) \det B \\ &= \det(E_1 E_2 \cdots E_k) \det B = \det A \cdot \det B. \end{aligned}$$

Thus we have shown that **D6** holds. □

D7: For all square matrices A, $\det A^T = \det A$.

Recall these facts about elementary matrices:

- $\det E_{ij}^T = \det E_{ij}$
- $\det E_i(c)^T = \det E_i(c)$
- $\det E_{ij}(c)^T = \det E_{ji}(c) = 1 = \det E_{ij}(c)$

Therefore, transposing does not affect determinants of elementary matrices. Now for the general case observe that since A and A^T are transposes of each other, one is invertible if and only if the other is by the Transpose/Inverse law. In particular, if both are singular, then $\det A^T = 0 = \det A$. On the other hand, if both are nonsingular, then write A as a product of elementary matrices, say $A = E_1 E_2 \cdots E_k$, and obtain from the product law for transposes that $A^T = E_k^T E_{k-1}^T \ldots E_1^T$, so by D6

$$\begin{aligned}\det A^T &= \det E_k^T \det E_{k-1}^T \cdots \det E_1^T = \det E_k \det E_{k-1} \cdots \det E_1 \\ &= \det E_1 \det E_2 \cdots \det E_k = \det A. \qquad \square\end{aligned}$$

Tensor Product of Matrices

How do we solve a system of equations in which the unknowns can be organized into a matrix X and the linear system in question is of the form

$$AX - XB = C, \tag{2.3}$$

Sylvester Equation

where A, B, C are given matrices? We call this equation the *Sylvester equation.* Such systems occur in a number of physical applications; for example, discretizing certain partial differential equations in order to solve them numerically can lead to such a system. Of course, we could simply expand each system laboriously. This direct approach offers us little insight as to the nature of the resulting system.

We are going to develop a powerful "bookkeeping" method that will rearrange the variables of Sylvester's equation automatically. The first basic idea needed here is that of the tensor product of two matrices, which is defined as follows:

Tensor Product

Definition 2.19. Let $A = [a_{ij}]$ be an $m \times p$ matrix and $B = [b_{ij}]$ an $n \times q$ matrix. Then the *tensor product* of A and B is the $mn \times pq$ matrix $A \otimes B$ that can be expressed in block form as

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1j}B & \cdots & a_{1p}B \\ a_{21}B & a_{22}B & \cdots & a_{2j}B & \cdots & a_{2p}B \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1}B & a_{i2}B & \cdots & a_{ij}B & \cdots & a_{ip}B \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mj}B & \cdots & a_{mp}B \end{bmatrix}.$$

Example 2.59. Let $A = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$. Exhibit $A \otimes B$, $B \otimes A$, and $I_2 \otimes A$ and conclude that $A \otimes B \neq B \otimes A$.

Solution. From the definition,

$$A \otimes B = \begin{bmatrix} 1B & 3B \\ 2B & 1B \end{bmatrix} = \begin{bmatrix} 4 & 12 \\ -1 & -3 \\ 8 & 4 \\ -2 & -1 \end{bmatrix}, \quad B \otimes A = \begin{bmatrix} 4A \\ -1A \end{bmatrix} = \begin{bmatrix} 4 & 12 \\ -8 & -2 \\ -1 & -3 \\ -2 & -1 \end{bmatrix},$$

$$\text{and} \quad I_2 \otimes A = \begin{bmatrix} 1A & 0A \\ 0A & 1A \end{bmatrix} = \begin{bmatrix} 1 & 3 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 2 & 1 \end{bmatrix}. \qquad \square$$

The other ingredient that we need to solve equation (2.3) is an operator that turns matrices into vectors. It is defined as follows.

Definition 2.20. Let A be an $m \times n$ matrix. Then the $mn \times 1$ vector $\operatorname{vec} A$ is obtained from A by stacking the n columns of A vertically, with the first column at the top and the last column of A at the bottom. Vec Operator

Example 2.60. Let $A = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix}$. Compute $\operatorname{vec} A$.

Solution. There are two columns to stack, yielding $\operatorname{vec} A = [1, 2, 3, 1]^T$. $\square$

The vec operator is linear ($\operatorname{vec}(aA + bB) = a \operatorname{vec} A + b \operatorname{vec} B$). We leave the proof, along proofs of the following simple tensor facts, to the reader.

Theorem 2.11. Let A, B, C, D be suitably sized matrices. Then

(1) $(A + B) \otimes C = A \otimes C + B \otimes C$
(2) $A \otimes (B + C) = A \otimes B + A \otimes C$
(3) $(A \otimes B) \otimes C = A \otimes (B \otimes C)$
(4) $(A \otimes B)^T = A^T \otimes B^T$
(5) $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$
(6) $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$

The next theorem lays out the key bookkeeping between tensor products and the vec operator.

Theorem 2.12. If A, X, B are matrices conformable for multiplication, then Bookkeeping Theorem

$$\operatorname{vec}(AXB) = \left(B^T \otimes A\right) \operatorname{vec} X.$$

Corollary 2.4. The following linear systems in the unknown X are equivalent.

(1) $A_1 X B_1 + A_2 X B_2 = C$
(2) $\left(\left(B_1^T \otimes A_1\right) + \left(B_2^T \otimes A_2\right)\right) \operatorname{vec} X = \operatorname{vec} C$

For Sylvester's equation, note that $AX - XB = IAX + (-I)\,XB$.

The following is a very basic application of the tensor product. Suppose we wish to model a two-dimensional heat diffusion process on a flat plate that occupies the unit square in the xy-plane. We proceed as we did in the one-dimensional process described in the introduction. To fix ideas, we assume that the heat source is described by a function $f(x, y), 0 \leq x \leq 1, 0 \leq y \leq 1$, and that the temperature is held at 0 at the boundary of the unit square. Also, the conductivity coefficient is assumed to be the constant k. Cover the square with a uniformly spaced set of grid points $(x_i, y_j), 0 \leq i, j \leq n + 1$, called nodes, and assume that the spacing in each direction is a width $h = 1/(n + 1)$. Also assume that the temperature function at the (i, j)th node is $u_{ij} = u(x_i, y_j)$ and that the source is $f_{ij} = f(x_i, y_j)$. Notice that the values of u on boundary grid points is set at 0. For example, $u_{01} = u_{20} = 0$. By balancing the heat flow in the horizontal and vertical directions, one arrives at a system of linear equations, one for each node, of the form

$$-u_{i-1,j} - u_{i+1,j} + 4u_{ij} - u_{i,j-1} - u_{i,j+1} = \frac{h^2}{k} f_{ij}, \quad i, j = 1, \ldots, n. \tag{2.4}$$

Observe that values of boundary nodes are zero, so these are not unknowns, which is why the indexing of the equations starts at 1 instead of 0. There are exactly as many equations as unknown grid point values. Each equation has a "molecule" associated with it that is obtained by circling the nodes that occur in the equation and connecting these circles. A picture of a few nodes is given in Figure 2.7.

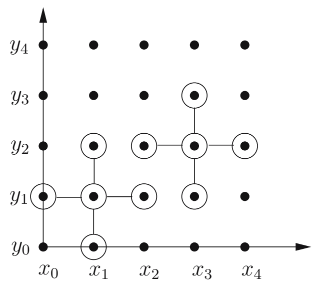


Fig. 2.7. Molecules for $(1, 1)$th and $(3, 2)$th grid points.

Example 2.61. Set up and solve a system of equations for the two-dimensional heat diffusion problem described above.

Solution. Equation (2.4) gives us a system of n^2 equations in the n^2 unknowns u_{ij}, $i, j = 1, 2, \ldots, n$. Rewrite equation (2.4) in the form

$$(-u_{i-1,j} + 2u_{ij} - u_{i+1,j}) + (-u_{i,j-1} + 2u_{ij} - u_{i,j+1}) = \frac{h^2}{k} f_{ij}.$$

Now form the $n \times n$ matrices

$$T_n = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & \ddots & 0 \\ 0 & \ddots & \ddots & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}.$$

Set $U = [u_{ij}]$ and $F = [f_{ij}]$, and the system can be written in matrix form as

$$T_n U + U T_n = T_n U I_n + I_n U T_n = \frac{h^2}{k} F.$$

However, we can't as yet identify a coefficient matrix, which is where Corollary 2.4 comes in handy. Note that both I_n and T_n are symmetric and apply the corollary to obtain that the system has the form

$$(I_n \otimes T_n + T_n \otimes I_n) \operatorname{vec} U = \operatorname{vec} \frac{h^2}{k} F.$$

Now we have a coefficient matrix, and what's more, we have an automatic ordering of the doubly indexed variables u_{ij}, namely

$$u_{1,1}, u_{2,1}, \ldots, u_{n,1}, u_{1,2}, u_{2,2}, \ldots, u_{n,2}, \ldots, u_{1,n}, u_{2,n}, \ldots, u_{n,n}.$$

This is sometimes called the "row ordering," which refers to the rows of the nodes in Figure 2.7, and not the rows of the matrix U. □

Here is one more example of a problem in which tensor notation is an extremely helpful bookkeeper. This is a biological model that gives rise to an inverse theory problem. ("Here's the answer, what's the question?")

Example 2.62. Refer to Example 2.20, where a three-state insect (egg, juvenile, adult) is studied in stages spaced at intervals of two days. One might ask how the entries of the matrix were derived. Clearly, observation plays a role. Let us suppose that we have taken samples of the population at successive stages and recorded our estimates of the population state. Suppose we have estimates of states $\mathbf{x}^{(0)}$ through $\mathbf{x}^{(4)}$. How do we translate these observations into transition matrix entries?

Solution. We postulate that the correct transition matrix has the form

$$A = \begin{bmatrix} P_1 & 0 & F \\ G_1 & P_2 & 0 \\ 0 & G_2 & P_3 \end{bmatrix}.$$

Theoretically, we have the transition equation $\mathbf{x}^{(k+1)} = A\mathbf{x}^{(k)}$ for $k = 0, 1, 2, 3$. Remember that this is an inverse problem, where the "answers," population states $\mathbf{x}^{(k)}$, are given, and the question "What are populations given A?" is unknown. We could simply write out each transition equation and express the results as linear equations in the unknown entries of A. However, this is laborious and not practical for problems involving many states or larger amounts of data.

Here is a better idea: assemble all of the transition equations into a single matrix equation by setting

$$M = \left[\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}\right] = [m_{ij}] \quad \text{and} \quad N = \left[\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \mathbf{x}^{(4)}\right] = [n_{ij}] \,.$$

The entire ensemble of transition equations becomes $AM = N$ with M and N known matrices. Here A is 3×3 and both M, N are 3×4. Next, write the transition equation as $I_3 AM = N$ and invoke the bookkeeping theorem to obtain the system

$$\operatorname{vec}(I_3 AM) = \left(M^T \otimes I_3\right) \operatorname{vec} A = \operatorname{vec} N.$$

This is a system of 12 equations in 9 unknowns. We can simplify it a bit by deleting the third, fourth, and eighth entries of $\operatorname{vec} A$ and the same columns of the coefficient matrix, since we know that the variables a_{31}, a_{12}, and a_{23} are zero. We thus end up with a system of 12 equations in 6 unknowns, which will determine the nonzero entries of A. □

Project Topics

Project: LU Factorization

Write a program module that implements Theorem 2.10 using partial pivoting and implicit row exchanges. This means that space is allocated for the $n \times n$ matrix $A = [a[i, j]]$ and an array of row indices, say indx$[i]$. Initially, indx should consist of the integers $1, 2, \ldots, n$. Whenever two rows need to be exchanged, say the first and third, then the indices indx[1] and indx[3] are exchanged. References to array elements throughout the Gaussian elimination process should be indirect: refer to the $(1, 4)$th entry of A as the element $a\,[\text{indx}[1], 4]$. This method of reference has the same effect as physically exchanging rows, but without the work. It also has the appealing feature that we can design the algorithm as though no row exchanges have taken place provided we replace the direct reference $a[i, j]$ by the indirect reference $a[\text{indx}[i], j]$. The module should return the lower/upper matrix in the format of Example 2.58 as well as the permuted array indx$[i]$. Effectively, this index array tells the user what the permutation matrix P is.

Write an LU system solver module that uses the LU factorization to solve a general linear system. Also write a module that finds the inverse of an $n \times n$ matrix A by first using the LU factorization module, then making repeated use of the LU system solver to solve $A\mathbf{x}^{(i)} = \mathbf{e}_i$, where $\mathbf{e}_i$ is the ith column of the identity. Then we will have

$$A^{-1} = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(n)}].$$

Be sure to document and test your code and report on the results.

Project: Markov Chains

Refer to Example 2.18 and Section 2.3 for background. Three automobile insurance firms compete for a fixed market of customers. Annual premiums are sold to these customers. Label the companies A, B, and C. You work for Company A, and your team of market analysts has done a survey that draws the following conclusions: in each of the past three years, the number of A customers switching to B is 20%, and to C is 30%. The number of B customers switching to A is 20%, and to C is 20%. The number of C customers switching to A is 30%, and to B is 10%. Those who do not switch continue to use their current company's insurance for the next year. Model this market as a Markov chain. Display the transition matrix for the model. Illustrate the workings of the model by showing what it would predict as the market shares three years from now if currently A, B, and C owned equal shares of the market.

The next part of your problem is as follows: your team has tested two advertising campaigns in some smaller test markets and are confident that the first campaign will convince 20% of the B customers who would otherwise stay with B in a given year to switch to A. The second advertising campaign would convince 20% of the C customers who would otherwise stay with C in a given year to switch to A. Both campaigns have about equal costs and would not change other customers' habits. Make a recommendation, based on your experiments with various possible initial state vectors for the market. Will these campaigns actually improve your company's market share? If so, which one do you recommend? Write up your recommendation in the form of a report, with supporting evidence. It's a good idea to hedge on your bets a little by pointing out limitations to your model and claims, so devote a few sentences to those points.

It would be a plus to carry the analysis further (your manager might appreciate that). For instance, you could turn the additional market share from, say B customers, into a variable and plot the long-term gain for your company against this variable. A manager could use this data to decide whether it was worthwhile to attempt gaining more customers from B.

Project: Affine Transforms in Real-Time Rendering

Refer to the examples in Section 2.3 for background. Graphics specialists find it important to distinguish between vector objects and point objects in three-dimensional space. They simultaneously manipulate these two kinds of objects with invertible linear operators, which they term *transforms*. To this end, they use the following clever ruse: identify three-dimensional vectors in the usual way, that is, by their coordinates x_1, x_2, x_3. Do the same with three-dimensional points. To distinguish between the two, embed them in the set of 4×1 vectors $\mathbf{x} = (x_1, x_2, x_3, x_4)$, called *homogeneous vectors*, with the understanding that if $x_4 = 0$, then $\mathbf{x}$ represents a three-dimensional vector object, and if $x_4 \neq 0$, then the vector represents a three-dimensional point whose coordinates are $x_1/x_4, x_2/x_4, x_3/x_4$.

Homogeneous Vector

Transforms (invertible linear operators) have the general form

$$T_M(\mathbf{x}) = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}.$$

Homogeneous and Affine Transforms

If $m_{44} = 1$ and the remaining entries of the last row and column are zero, the transform is called a *homogeneous transform.* If $m_{44} = 1$ and the remaining entries of the last row are zero, the transform is called *affine.* If the transform matrix M takes the block form $M = \begin{bmatrix} I_3 & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}$, the transform T_M is called a *translation* by the vector $\mathbf{t}$. All other operators are called *nonaffine.*

In real-time rendering it is sometimes necessary to invert an affine transform. Computational efficiency is paramount in these calculations (after all, this is real time!). So your objective in this project is to design an algorithm that accomplishes this inversion with a minimum number of flops. Preface discussion of your algorithm with a description of affine transforms. Give a geometrical explanation of what homogeneous and translation transforms do to vectors and points. You might also find it helpful to show that every affine transform is the composition of a homogeneous and a translation transform. Illustrate the algorithm with a few examples. Finally, you might discuss the stability of your algorithm. Could it be a problem? If so, how would you remedy it? See the discussion of roundoff error in Section 1.5.

Project: Modeling with Directed Graphs I

Refer to Example 2.21 and Section 2.3 for background. As a social scientist you have studied the influence factors that relate seven coalition groups. For simplicity, we will label the groups as $1, 2, 3, 4, 5, 6, 7$. Based on empirical studies, you conclude that the influence factors can be well modeled by a dominance-directed graph with each group as a vertex. The meaning of the presence of an edge (i, j) in the graph is that coalition group i can dominate, i.e., swing coalition group j its way on a given political issue. The data you have gathered suggest that the appropriate edge set is the following:

$$\begin{aligned} E = \{&(1,2),(1,3),(1,4),(1,7),(2,4),(2,6),(3,2),(3,5),(3,6), \\ &(4,5),(4,7),(5,1),(5,6),(5,7),(6,1),(6,4),(7,2),(7,6)\}. \end{aligned}$$

Do an analysis of this power structure. This should include a graph. (It might be a good idea to arrange the vertices in a circle and go from there.) It should also include a power rating of each coalition group. Now suppose you were an adviser to one of these coalition groups, and by currying certain favors, this group could gain influence over another coalition group (thereby adding an edge to the graph or reversing an existing edge of the graph). In each case, if you could pick the best group for your client to influence, which would that be? Explain your results in the context of matrix multiplication if you can.

2.7 Exercises and Problems

Exercise 1. Use LU factorization of $A = \begin{bmatrix} 2 & -1 & 1 \\ 2 & 3 & -2 \\ 4 & 2 & -2 \end{bmatrix}$ to solve $A\mathbf{x} = \mathbf{b}$, where
(a) $\mathbf{b} = (6, -8, -4)$ (b) $\mathbf{b} = (2, -1, 2)$ (c) $\mathbf{b} = (1, 2, 4))$ (d) $\mathbf{b} = (1, 1, 1)$.

Exercise 2. Use PLU factorization of $A = \begin{bmatrix} 0 & -1 & 1 \\ 2 & 3 & -2 \\ 4 & 2 & -2 \end{bmatrix}$ to solve $A\mathbf{x} = \mathbf{b}$,
(a) $\mathbf{b} = (3, 1, 4)$ (b) $\mathbf{b} = (2, -1, 3)$ (c) $\mathbf{b} = (1, 2, 0))$ (d) $\mathbf{b} = (1, 0, 0)$.

Exercise 3. Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}$. Calculate the following.
(a) $A \otimes B$ (b)$B \otimes A$ (c)$A^{-1} \otimes B^{-1}$ (d)$(A \otimes B)^{-1}$

Exercise 4. Let $A = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 2 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & -3 \\ 3 & 0 \end{bmatrix}$. Calculate the following.
(a) $A \otimes B$ (b)$B \otimes A$ (c)$A^T \otimes B^T$ (d)$(A \otimes B)^T$

Exercise 5. With A and B as in Exercise 3, $C = \begin{bmatrix} 2 & -1 \\ 1 & 0 \\ 1 & 3 \end{bmatrix}$, and $X = [x_{ij}]$ a 3×2 matrix of unknowns, use tensor products to determine the coefficient matrix of the linear system $AX + XB = C$ in matrix–vector form.

Exercise 6. Use the matrix A and methodology of Example 2.62 with $\mathbf{x}^{(0)} = (1, 2, 3)$, $\mathbf{x}^{(0)} = (0.9, 1.2, 3.6)$, and $\mathbf{x}^{(0)} = (1, 1.1, 3.4)$ to express the resulting system of equations in the six unknown nonzero entries of A in matrix–vector form.

*Problem 7. Show that if A is a nonsingular matrix with a zero $(1, 1)$th entry, then A does not have an LU factorization.

Problem 8. Prove that if A is $n \times n$, then $\det(-A) = (-1)^n \det A$.

Problem 9. Let A and B be invertible matrices of the same size. Use determinantal law D9 to prove that $\operatorname{adj} A^{-1} = (\operatorname{adj} A)^{-1}$ and $\operatorname{adj}(AB) = \operatorname{adj} A \operatorname{adj} B$.

Problem 10. Verify parts 1 and 4 of Theorem 2.11.

Problem 11. Verify parts 5 and 6 of Theorem 2.11.

Problem 12. If heat is transported with a horizontal velocity v as well as diffused in Example 2.61, a new equation results at each node in the form

$$-u_{i-1,j} - u_{i+1,j} + 4u_{ij} - u_{i,j-1} - u_{i,j+1} - \frac{vh}{2k}\left(u_{i+1,j} - u_{i-1,j}\right) = \frac{h^2}{k} f_{ij}$$

for $i, j = 1, \ldots, n$. Vectorize the system and use tensor products to identify the coefficient matrix of this linear system.

*Problem 13. Prove the Bookkeeping Theorem (Theorem 2.12).

Problem 14. Determine the cost of the LU factorization of an invertible $n \times n$ matrix A, ignoring row exchanges.

http://www.springer.com/978-0-387-33194-2

Applied Linear Algebra and Matrix Analysis
Shores, Th.S.
2007, XII, 384 p. 27 illus., Hardcover
ISBN: 978-0-387-33194-2