AN INTRODUCTION
TO
MATRIX ALGEBRA

Craig W. Rasmussen
Department of Mathematics
Naval Postgraduate School

AN INTRODUCTION
TO
MATRIX ALGEBRA

Craig W. Rasmussen

Department of Mathematics
Naval Postgraduate School

PREFACE

These notes are intended to serve as the text for MA1042:Matrix Algebra, and for its 6-week refresher equivalent, MAR142. The intent of MA1042 and MAR142 is to equip students with the fundamental mechanical skills for follow-on courses in linear algebra that require a basic working knowledge of matrix algebra and complex arithmetic. In the Department of Mathematics, the principal target course is MA3042, Linear Algebra. Those whose plans lead to MA3046, Computational Linear Algebra, are urged to take MA1043 rather than MA1042, although these notes might serve as the nucleus of the material covered in that course. Other disciplines make heavy use of these techniques, as well; one must venture into the arts and humanities to find fields in which linear algebra does not come into play.

The notes contain two sections that are somewhat beyond what is normally covered in MA1042. The first is section 2.5, which gives a cursory introduction to vector spaces and linear transformations. Optional reading for students in MA1042, this provides a glimpse of the subject matter of MA3042. The second is section 3.5, the topic of which is the cross product. This is of potential interest to students of multivariable (three, to be precise) calculus.

The course opens with an introduction to systems of linear equations and their solution by substitution. The limitations of this method quickly become apparent, and we begin the development of alternatives by discussing the simplicity of the substitution method on certain highly structured systems. This leads us directly to Gaussian elimination. At this point we also observe, and exploit, the ease with which linear equations can be represented by matrices of coefficients, although matrices at this point are viewed as notational conveniences that relieve some of the overhead of solving sytems rather than as mathematical objects in their own right. We take a brief look at the geometry of linear systems, enabling us to describe geometrically the conditions under which a solution exists and to introduce informally the idea of an ill-conditioned system of equations.

In Chapter 2, we take a first look at matrices as algebraic objects, and describe the fundamental operations on these objects. We describe briefly the elementary properties of linear transformations, and close with an introduction to elementary matrices and inverses.

Chapter 3 revisits systems of linear equations from the point of view of matrices, and shows that Gaussian elimination can be viewed as a special case of matrix factorization. This leads naturally to the LU-decomposition of a square matrix. In keeping with the spirit of the course, we make simplifying assumptions that do not apply in the general case. It is here that we introduce the determinant, which is initially applied to the solution of systems of linear equations via Cramer's rule.

Chapter 4 introduces eigenvalues and eigenvectors, which open the door to application of matrix algebra to a multitude of problems that arise throughout the pure and applied sciences. As in Chapter 3, we do not consider the topic of eigenvalues in its full generality; the general view is provided in subsequent courses.

Chapter 5, an introduction to complex numbers, is independent of the preceding chapters, with the exception of a section that is concerned with complex eigenvalues and eigenvectors. We cover the fundamental operations on complex numbers, and briefly discuss the geometry of the complex plane. The shift from rectangular to polar coordinates leads directly to DeMoivre's Theorem and, with the assistance of infinite series, to Euler's formula and the exponential representation of complex numbers.

A small problem set follows each chapter, and an appendix contains solutions to most. Subsequent revisions will include additional exercises. It will be greatly appreciated if any who use these notes will continue to report errors, omissions, and suggestions for improvement directly to the author, who will endeavor to follow up.

MA/Ra, 24 April 2002

COURSE OBJECTIVES

MA1042 MATRIX ALGEBRA

Upon completion of the course, the student should be able to perform the following operations.

1. Solve rectangular systems of linear equations using Gaussian elimination.
2. Write a system of linear equations in matrix form, and describe the nature of the solutions (infinitely many, exactly one, or no solutions).
3. Perform algebraic operations on matrices: addition/subtraction, multiplication, multiplication by a constant, raising to powers, transposition.
4. Describe the elementary algebraic properties of matrix operations: commutativity, associativity, etc.
5. Perform algebraic operations on partitioned matrices in terms of the blocks.
6. Calculate the dot product of two n-tuples.
7. Explain matrix multiplication in three distinct ways: focusing on individual elements, on combinations of columns, and on combinations of rows.
8. Formulate the definition of the inverse of a matrix, and describe its basic properties.
9. Find the inverse of a square matrix using Gaussian elimination.
10. Find the LU-decomposition of a matrix using Gaussian elimination and recording the multipliers.
11. Find the rank of a matrix.
12. Compute the determinant of a square matrix using cofactor expansion.
13. Describe the basic properties of determinants, and the effect on the determinant of each of the elementary row operations.

14. Use Cramer's rule to solve small systems of linear equations.
15. Compute simple eigenvalues and associated eigenvectors of small matrices, including those with complex eigenvalues.
16. Represent complex numbers in rectangular, exponential, and trigonometric form, and be able to convert from any form to any other.
17. Perform arithmetic operations (addition, subtraction, multiplication, division, exponentiation) on complex numbers.
18. Determine the magnitude, argument, real part, imaginary part, and conjugate of a complex number.
19. State and apply DeMoivre's theorem.
20. State and apply Euler's identity.
21. State and apply the Fundamental Theorem of algebra.

Contents

Chapter 1

Systems of Linear Equations

Systems of linear equations arise naturally in an enormous number of disciplines, ranging from the physical sciences and engineering to management science, operations research, and, of course, mathematics. The emphasis here is on working with such systems, rather than on constructing them, so this preliminary edition of the text omits a discussion of the myriad ways in which systems of linear equations can arise.

1.1 Linear Systems and Solution by Substitution

In the Cartesian plane, the equation of a line can be put in the form $ax + by = c$, where x and y are real variables, a, b, and c are real constants, and at least one of a and b is nonzero. This is an example of the following more general idea.

Definition: A *linear equation* in the variables $x_1, x_2, \ldots, x_n$ is an equation that can be put in the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b, \tag{1.1}$$

where b and the coefficients a_i $(1 \leq i \leq n)$ are constants.

Note that, for the equation (1.1) to be linear, every variable x_i occurs to the first power only, and no term in the sum involves products of variables. Thus there are no occurrences of transcendental functions (trigonometric or exponential functions, for example), and no occurrences of the taking of roots. For example, the equations (a) $2x = 3$ and (b) $x - 2y = 0$ are linear, while the equations (c) $x \sin x + \sqrt{y} = 2$ and (d) $x^2 - 2xy = e^z$ are not. In passing, it should be pointed out that a *linear* *inequality* is similarly defined, but with one of $<, \leq, >$, or $\geq$ taking the place of equality in the preceding definition.

A *solution* to the linear equation (1.1) is an ordered set $(s_1, s_2, \ldots, s_n)$ of numbers with the property that $a_1 s_1 + a_2 s_2 + \cdots + a_n s_n = b$ holds. The *solution set* of (1.1) is the set containing all such solutions. To *solve* (1.1) is to find the solution set.

A *system* of linear equations, also called a *linear system*, is a collection of $m > 1$ linear equations in the n variables $x_1, x_2, \ldots, x_n$ that we want to solve simultaneously, i.e., we look for a set $\{x_1, x_2, \ldots, x_n\}$ that simultaneously solves each equation in the set. For example, the system

$$\begin{cases} x + 2y = 0 \\ 2x + y = 3 \end{cases} \tag{1.2}$$

has the solution $x = 2,\ y = -1$. We say that two systems of linear equations are *equivalent* if they have identical solution sets. For example, the system

$$\begin{cases} x + 2y = 0 \\ -4x - 2y = -6 \end{cases}$$

is equivalent to 1.2.

A linear equation may be solved for any one of its variables in terms of the remaining variables, so long as the variable in question has a nonzero coefficient. For example, suppose we want to find all solutions to the equation, $2x - 4y = 6$. We could begin by solving for x, obtaining $x = 3 + 2y$. We now see that to any choice of y there corresponds a value of x, so the number of solutions is infinite. To describe the *general* solution, which is the set of all solutions, we allow y to take on any real value by introducing a parameter, say t, to replace y, obtaining

$$x = 3 + 2t,\ y = t,\ t \in \mathbf{R}.$$

Note that the decision to solve for x was arbitrary; solving for y would proceed along the same lines, and we would have

$$y = \frac{1}{2}t - \frac{3}{2},\ x = t,\ t \in \mathbf{R}.$$

There are several strategies at our disposal for solving linear systems. The most elementary is *substitution*, which involves the sequential elimination of variables from the system, the result at the $(k+1)$st step being an equivalent system in which the kth equation contains only the variables $x_k, \ldots, x_n$. For example, suppose we want to solve (1.2) by substitution. Let's say that we choose initially to solve the first equation for x, obtaining $x = -2y$. We then replace x with $-2y$ in the second equation, obtaining a new system,

$$\left\{ \begin{array}{rcrcr} x & + & 2y & = & 0 \\ & - & 3y & = & 3. \end{array} \right. \tag{1.3}$$

Finally, we solve (1.3) by *back-substitution*: from $-3y = 3$ we obtain $y = -1$. Substituting $y = -1$ in $x + 2y = 0$ we obtain $x = 2$, which agrees with our previous solution to (1.2). In general, back-substitution solves for the variables in reverse order relative to that in which they appear.

In order to facilitate this approach, a few observations are in order. First, since a solution to a system of equations must simultaneously solve each equation in the system, it follows that the order in which the equations are written is irrelevant to the final outcome. For example, the system of equations

$$\left\{ \begin{array}{rcrcr} 2x & + & y & = & 3 \\ x & + & 2y & = & 0 \end{array} \right. \tag{1.4}$$

is equivalent to (1.2). Second, if we multiply an equation by a nonzero constant, we do not change the solution set of that equation. Consequently, if we multiply one equation in a system by a nonzero constant, we obtain an equivalent system. For example, (1.4) and

$$\left\{ \begin{array}{rcrcr} 4x & + & 2y & = & 6 \\ x & + & 2y & = & 0 \end{array} \right. \tag{1.5}$$

are equivalent. Finally, if E_i and E_j are distinct equations in a system, then if we add a multiple of E_i to E_j, the resulting system is equivalent to its predecessor. For example, by adding twice the first equation in (1.5) to the second, we obtain

$$\left\{ \begin{array}{rcrcr} 4x & + & 2y & = & 6 \\ 9x & + & 6y & = & 12, \end{array} \right.$$

which is easily shown to be equivalent to (1.5).

1.2 Gaussian Elimination

We can exploit the observations made above to replace a system of equations with an equivalent system that is easier to work with. For example, consider the system,

$$\left\{\begin{array}{rcrcrcr} 2x_1 & + & x_2 & + & x_3 & = & 3 \\ 4x_1 & + & 2x_2 & - & x_3 & = & 0 \\ 2x_1 & - & x_2 & + & 4x_3 & = & 5. \end{array}\right. \tag{1.6}$$

We add (-2) times the first equation in (1.6) to the second, and (-1) times the first to the third, obtaining

$$\left\{\begin{array}{rcrcrcr} 2x_1 & + & x_2 & + & x_3 & = & 3 \\ & & & - & 3x_3 & = & -6 \\ & - & 2x_2 & + & 3x_3 & = & 2. \end{array}\right. \tag{1.7}$$

We then swap the second and third equations in (1.7), obtaining

$$\left\{\begin{array}{rcrcrcr} 2x_1 & + & x_2 & + & x_3 & = & 3 \\ & - & 2x_2 & + & 3x_3 & = & 2 \\ & & & - & 3x_3 & = & -6, \end{array}\right. \tag{1.8}$$

and we're ready for back-substitution. The third equation gives us $x_3 = 2$. Substituting this result into the second gives us $x_2 = 2$. Finally, we substitute both into the first equation to obtain $x_1 = -1/2$.

This process is *Gaussian elimination*. When applied to a square system (number of equations equals number of unknowns), the result is a *triangular system of equations*, which lends itself to back-substitution. Generally, we stop at this stage. We could, however, continue the process by eliminating x_3 from the second equation and both x_2 and x_3 from the first. To begin, we add 1/2 times the second equation in (1.8) to the first, obtaining

$$\left\{\begin{array}{rcrcrcr} 2x_1 & & & + & \frac{5}{2}x_3 & = & 4 \\ & - & 2x_2 & + & 3x_3 & = & 2 \\ & & & - & 3x_3 & = & -6. \end{array}\right. \tag{1.9}$$

We then add 5/6 times the third equation in (1.9) to the first, and add the third equation to the second, with the resulting system being

$$\left\{\begin{array}{rcrcrcr} 2x_1 & & & & & = & -1 \\ & - & 2x_2 & & & = & -4 \\ & & & - & 3x_3 & = & -6. \end{array}\right.$$

This is a *diagonal system* of equations, even easier to solve then the triangular system. The extended elimination process is called

Gauss-Jordan elimination. There are situations in which Gauss-Jordan elimination is desirable by virtue of the fact that it completely uncouples the equations in the system. Nevertheless, in most cases Gaussian elimination with back-substitution is the preferred method because of its simplicity. If we make the simplifying assumption that at no stage in the elimination are we confronted with the need to divide by zero, we can summarize the process of Gaussian elimination on square systems as follows. Assume that we have n equations in n unknowns.

1. We begin by adding suitable multiples of equation 1 to equations $2, \ldots, n$, introducing 0 as the coefficient on x_1 in each equation.

2. Having introduced 0's as the first k coefficients in each of equations $k+1, \ldots, n$ $(1 \leq k < n-1)$, add suitable multiples of equation $k+1$ to each of equations $k+2, \ldots, n$ to introduce zeros as the coefficient on x_{k+1} in equations $k+2, \ldots, n$.

3. When the elimination is complete, we begin the back-substitution process by solving equation n for x_n.

4. Having solved equations $n, \ldots, n-k$ $(0 \leq k < n-1)$ for the variables $x_n, \ldots, x_{n-k}$, we substitute the values obtained into equation $n-k-1$ and solve for x_{n-k-1}.

1.3 The Geometry of Linear Systems

Each of the systems of equations thus far has had a unique solution. As it turns out, this is not always the case. The underlying geometric situation is most easily described in the case of 2×2 systems, i.e., systems of two equations in two unknowns. The following three systems illustrate the three possibilities.

$$\left\{ \begin{array}{rcrcr} x & + & y & = & 0 \\ x & + & y & = & 1, \end{array} \right. \tag{1.10}$$

$$\left\{ \begin{array}{rcrcr} x & + & y & = & 0 \\ 2x & + & y & = & 1, \end{array} \right. \tag{1.11}$$

$$\left\{ \begin{array}{rcrcr} x & + & y & = & 0 \\ 2x & + & 2y & = & 0. \end{array} \right. \tag{1.12}$$

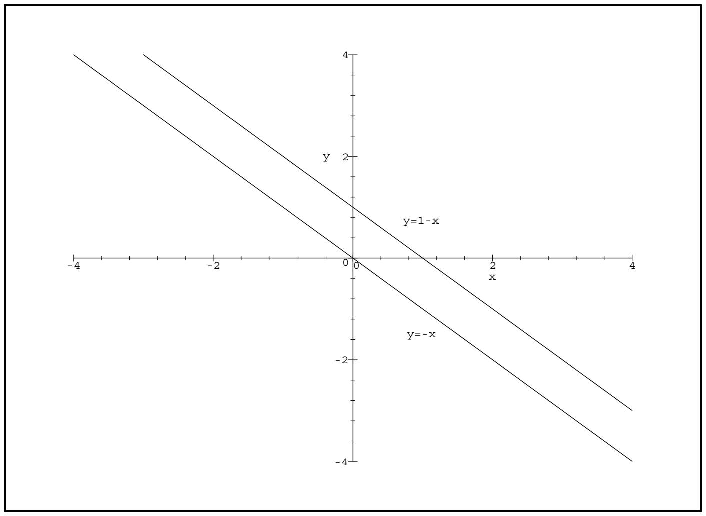

Figure 1.1: Graphical representation of an inconsistent system.

In each system we have the equations for two lines in the plane. From now on, for convenience, we refer to the plane as $\mathbf{R}^2$. A solution to such a system describes the intersection(s) of the two lines. System (1.10) is *inconsistent*, as depicted in Figure 1.1. The first equation refers to the line $y = -x$, while the second refers to the line $y = 1 - x$. These are parallel lines with different y-intercepts, and so have no intersection. System (1.11) has a unique solution $(x, y) = (1, -1)$, the point of intersection of the lines $y = -x$ and $y = 1 - 2x$. Figure 1.2 illustrates this situation. System (1.12) has infinitely many solutions, since the second equation is simply a multiple of the first; both describe the line $y = -x$.

In three dimensions (we subsequently refer to Euclidean 3-space as $\mathbf{R}^3$, and generalize this notation in an obvious way), a linear equation $ax + by + cz = d$ describes a *plane*. A system of k equations in three unknowns therefore describes k planes in $\mathbf{R}^3$. The intersection of k planes in $\mathbf{R}^3$ is either empty (no solution), a point (unique solution), a line (infinitely many solutions), or a plane (infinitely many solutions). The situation in more than three dimensions is more difficult to visualize, but is perfectly analogous.

The geometry of linear systems can have unexpected side effects when solutions are computed. This can be illustrated in the 2×2 case, for

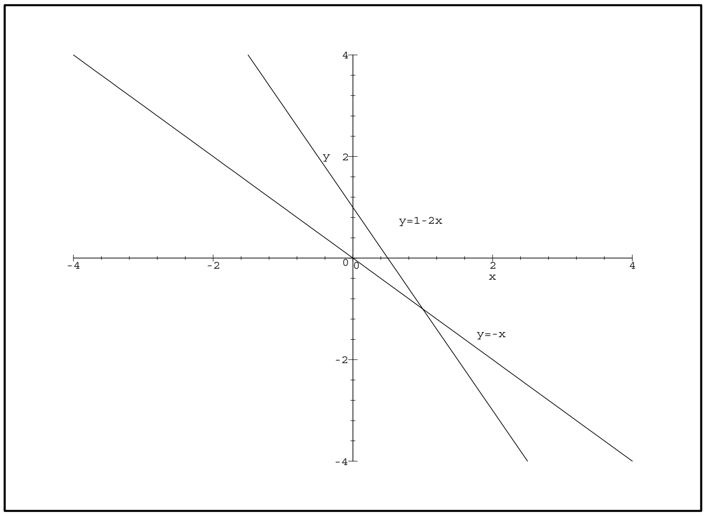

Figure 1.2: Graphical representation of a system with exactly one solution.

simplicity. Suppose the lines described by the equations are *nearly* parallel. Then a small perturbation of one line is likely to result in a large perturbation in the point of intersection. For example, the lines $y = 1.4142x$ and $y = \sqrt{2}x$ are very nearly parallel, but intersect only at $(0, 0)$. Thus the system,

$$\left\{ \begin{array}{rcrcr} 1.4142x & - & y & = & 0 \\ \sqrt{2}x & - & y & = & 0, \end{array} \right. \tag{1.13}$$

has a unique solution given by $(x, y) = (0, 0)$. Now suppose we perturb the right-hand side just a bit, obtaining the system

$$\left\{ \begin{array}{rcrcr} 1.4142x & - & y & = & 0 \\ \sqrt{2}x & - & y & = & 0.001. \end{array} \right.$$

Geometrically, all that has happened is that the line corresponding to the second equation has been shifted very slightly. Its slope is unchanged, but its y-intercept has changed. This system, too, has a unique solution approximated by $(x, y) = (73.7, 104.3)$. The system (1.13) is said to be *ill-conditioned*, meaning that a small perturbation in the coefficients can result in a large perturbation in the solution.

Machine computation itself can result in unexpected errors, independent of the geometry of the system. While such numerical issues are beyond the scope of this course, we illustrate one sort of difficulty that can arise with the following example. Consider the system,

$$\left\{ \begin{array}{rcrcr} 0.0001x & + & y & = & 1 \\ x & + & y & = & 2, \end{array} \right. \tag{1.14}$$

The exact solution (approximated to four decimal places) is $(x, y) = (1.0001, 0.9999)$. If we solve by Gaussian elimination, and round to three significant digits, we find the "solution" $(x, y) = (0, 1)$. On the other hand, if we interchange the equations in (1.14) before elimination and again round to three significant digits, we find the approximate solution $(x, y) = (1, 1)$, which is much closer to the actual solution.

1.4 Matrices and Linear Systems

A *matrix* is a rectangular array of numbers. If a matrix A has m rows and n columns, we say that A is an $m \times n$ matrix. We associate matrices with systems of linear equations in a useful way. Consider the linear system

$$\left\{ \begin{array}{rcrcrcr} x_1 & - & x_2 & + & x_3 & = & 6 \\ 4x_1 & + & 2x_2 & - & x_3 & = & 0 \\ 5x_1 & + & x_2 & + & x_3 & = & 12. \end{array} \right. \tag{1.15}$$

The relative positions of x_1, x_2, and x_3 are identical within each of the three equations. As long as we agree on this, we might just as well suppress the variables. We can represent the system (1.15) by the corresponding *augmented matrix*

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

The three types of operations that we performed on a system of equations to find an equivalent, simpler, system become *elementary row operations* that we perform on the augmented matrix. These operations are:

1. Interchange two rows.

2. Multiply any row by a nonzero constant.

3. Add a constant multiple of any row to another.

The following example illustrates the two approaches.

Linear System	Augmented Matrix
$\begin{cases} x_1 - x_2 + x_3 = 6 \\ 4x_1 + 2x_2 - x_3 = 0 \\ 5x_1 + x_2 + x_3 = 12. \end{cases}$	$\begin{bmatrix} 1 & -1 & 1 & 6 \\ 4 & 2 & -1 & 0 \\ 5 & 1 & 1 & 12 \end{bmatrix}.$
Add (-4) times first equation to second $\begin{cases} x_1 - x_2 + x_3 = 6 \\ 6x_2 - 5x_3 = -24 \\ 5x_1 + x_2 + x_3 = 12. \end{cases}$	Add (-4) times first row to second. $\begin{bmatrix} 1 & -1 & 1 & 6 \\ 0 & 6 & -5 & -24 \\ 5 & 1 & 1 & 12 \end{bmatrix}.$
Add (-5) times first equation to third $\begin{cases} x_1 - x_2 + x_3 = 6 \\ 6x_2 - 5x_3 = -24 \\ 6x_2 - 4x_3 = -18. \end{cases}$	Add (-5) times first row to third. $\begin{bmatrix} 1 & -1 & 1 & 6 \\ 0 & 6 & -5 & -24 \\ 0 & 6 & -4 & -18 \end{bmatrix}.$
Add (-1) times second equation to third. $\begin{cases} x_1 - x_2 + x_3 = 6 \\ 6x_2 - 5x_3 = -24 \\ x_3 = 6. \end{cases}$	Add (-1) times second row to third. $\begin{bmatrix} 1 & -1 & 1 & 6 \\ 0 & 6 & -5 & -24 \\ 0 & 0 & 1 & 6 \end{bmatrix}.$

The solution can now be calculated by back-substitution: The third equation gives $x_3 = 6$. Substitution into the second gives $x_2 = 1$. Substitution of both into the first gives $x_1 = 1$.

If a matrix B can be obtained from a matrix A by a sequence of elementary row operations, then A and B are said to be *row-equivalent*. If the row operations are those of Gaussian elimination, then the final form for the matrix is called *row-echelon* form, which is defined by the following three properties:

1. Any row containing only zeros follows all rows containing nonzeros.
2. If $i < j$, the column containing the first nonzero entry in row i precedes the column containing the first nonzero entry in row j. The first nonzero entry in any row is the *pivot*, or *leading entry*.

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

Figure 1.3: A, B, and C are row-equivalent.

3. All entries below any pivot are zeros.

Using Gaussian elimination, any matrix can be transformed to a row-equivalent matrix that is in row-echelon form. There is some flexibility, though, in choosing multipliers, in selecting row interchanges, etc. Consequently, the row-echelon form achieved is not unique. For example, matrices A and B in Figure 1.3 are row-equivalent. It is sometimes desirable to find a *canonical form*, one that is uniquely associated with A. We can accomplish this by performing Gauss-Jordan elimination and then taking one additional step. We *normalize* the pivot in each row, by multiplying that row by the reciprocal of the pivot. The result is a matrix in *reduced row-echelon* form, which means that

1. the matrix is in row-echelon form,
2. the pivot in any row containing nonzeros is a 1, and
3. all entries above the pivot in any row containing nonzeros are zeros. (Combined with (1), this means that a pivot is the only nonzero in its column.)

For any matrix A, the reduced row-echelon form is unique. For example, the reduced row-echelon form associated with A and B in Figure 1.3 is C.

1.5 Pivot and Free Variables

Consider the following matrix A, which can be interpreted as the row-echelon form of the augmented matrix for a linear system, also shown. If we attempt to apply back-substitution to the system, we have some difficulty due to the fact that every equation in the system involves at least two variables.

$$A = \begin{bmatrix} 2 & 3 & -1 & 5 & 2 \\ 0 & 3 & 2 & -1 & 2 \\ 0 & 0 & -2 & -8 & 4 \end{bmatrix} \qquad \left\{ \begin{array}{rrrrcr} 2x_1 & +3x_2 & -x_3 & +5x_4 & = & 2 \\ & 3x_2 & +2x_3 & -x_4 & = & 2 \\ & & -2x_3 & -8x_4 & = & 4 \end{array} \right.$$

Variables x_1, x_2, and x_3 correspond to the pivot entries in A. For this reason, they are called *pivot variables*. They are the *dependent* variables in the system[1], with their values determined by the right-hand side *and by* x_4. We refer to x_4 as a *free*, or *independent*, variable. By moving the free variable to the right-hand side, we obtain a system that can be solved by back-substitution:

$$\left\{\begin{array}{rcrcrcrcr} 2x_1 & + & 3x_2 & - & x_3 & = & 2 & - & 5x_4 \\ & & 3x_2 & + & 2x_3 & = & 2 & + & x_4 \\ & & & - & 2x_3 & = & 4 & + & 8x_4 \end{array}\right.$$

Since x_4 remains independent, we assign to x_4 a parameter, t. As t is allowed to vary, we obtain an infinite number of solutions, i.e., our solution is $(x_1, x_2, x_3, x_4) = (-3 - 9t, 2 + 3t, -2 - 4t, t)$, where t is any real number. For example, both $(-3, 2, -2, 0)$ and $(6, -1, 2, -1)$ are solutions. It is possible to have multiple free variables, as in the following example. As above, we interpret A to be the augmented matrix for the indicated system.

$$A = \begin{bmatrix} 3 & -2 & 1 & 3 & 1 & 14 \\ 0 & 0 & 2 & 5 & -3 & 2 \\ 0 & 0 & 0 & 0 & -3 & -6 \end{bmatrix} \quad \left\{\begin{array}{rcr} 3x_1 - 2x_2 + x_3 + 3x_4 + x_5 & = & 14 \\ 2x_3 + 5x_4 - 3x_5 & = & 2 \\ -3x_5 & = & -6 \end{array}\right.$$

The pivot variables are $x_1,\ x_3$, and x_5; x_2 and x_4 are free. At this point, we parameterize the free variables by letting $x_2 = s,\ x_4 = t$, and move them to the right-hand side, obtaining

$$\left\{\begin{array}{rcl} 3x_1 + x_3 + x_5 & = & 14 + 2s - 3t \\ 2x_3 - 3x_5 & = & 2 - 5t \\ -3x_5 & = & -6 \end{array}\right.$$

We now apply back-substitution. The third equation is easy: $x_5 = 2$. Substituting $x_5 = 2$ in the second equation, we have $2x_3 - 6 = 2 - 5t$, or $x_3 = 4 - \frac{5}{2}t$. The first equation becomes $3x_1 = 14 + 2s - 3t - (4 - \frac{5}{2}t) - 2$, which simplifies as $x_1 = \frac{8}{3} + \frac{2}{3}s - \frac{1}{6}t$. So we have an infinite number of solutions described by

$$(x_1, x_2, x_3, x_4, x_5) = (\frac{8}{3} + \frac{2}{3}s - \frac{1}{6}t, s, 4 - \frac{5}{2}t, t, 2).$$

[1]Since a linear equation can be solved for any of its variables whose coefficients are nonzero, it is frequently possible to choose a variable other than the pivot to designate as the dependent variable in any row. However, this is implicitly a column interchange, and can lead to headaches when doing hand computation. It is strongly recommended that students stick to the convention of interpreting the pivot variable in any row as dependent only upon the variables to its right.

The parameters s and t represent arbitrary real numbers; since we may replace either with any real number, we have infinitely many solutions.

1.6 Homogeneous and Nonhomogeneous Systems; the General Solution

Each of the systems of equations that we have examined so far has had the property that the right-hand side has contained nonzeros. Such a system is called *nonhomogeneous*; a system in which the right-hand side contains only zeros is called *homogeneous*. A homogeneous system of m equations in n unknowns has the form

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

Such a system always has at least one solution in which $x_1 = x_2 = \cdots = x_n = 0$. This is the *trivial* solution. A solution containing at least one nonzero is *nontrivial*. The *general solution* to a system of equations describes all solutions to the system. We postpone a thorough discussion of the general solution until we have developed a more compact way of representing such systems, but illustrate the idea with the following example.

Consider the system of equations,

$$\begin{cases} x_1 + x_2 + x_3 = 0 \\ \quad -x_2 - x_3 + 2x_4 = 0 \\ \quad -x_4 = 0 \end{cases}$$

The pivots are x_1, x_2, and x_4, with x_3 free. Let $x_3 = s$. From the third equation, we have $x_4 = 0$. The second equation gives $x_2 = -s$, and the first gives $x_1 = -x_2 - x_3 = 0$. The general solution to the homogeneous system is then $(x_1, x_2, x_3, x_4) = (0, -s, s, 0)$. Setting $s = 1$, we have $(x_1, x_2, x_3, x_4) = (0, -1, 1, 0)$, a *basic* solution. For any fixed s, we obtain a basic solution, and by letting s remain a free parameter we obtain the general solution. We return to the discussion of homogeneous and nonhomogeneous systems, and of their solutions, in a later section.

1.7 Summary

In this introductory chapter, we have reviewed the definitions of *linear equation* and *system of linear equations*, and have seen the fundamental solution techniques for solving systems of linear equations: substitution and Gaussian elimination. The goal of each approach is to arrive at an equivalent system of equations that can be solved by *back-substitution*. As a bookkeeping convenience, we have introduced *matrices*. The fundamental matrices associated with a system of linear equations are the *coefficient matrix*, which represents the coefficients of the variables, and the *augmented matrix*, which consists of the coefficient matrix with an additional column appended to the right side; this new column contains the constants from the right-hand side of the system. By using *elementary row operations*, we apply Gaussian elimination directly to the augmented matrix. We distinguish between *pivot*, or *dependent* variables, and *free*, or *independent* variables: a system of equations with no free variables has at most one solution, while a system of equations with at least one free variable can have either no solution or infinitely many solutions. We have taken a very brief look at the geometry of linear systems and at some of the difficulties that can arise when we attempt to solve these systems. We have not yet looked at matrices as algebraic objects in their own right, but shall do so shortly.

1.8 Exercises for Chapter 1

1. Which of the following equations is linear? Explain.

   (a) $x - 2xy = 6$

   (b) $11x - 26y + z - \cos s = e^t$

   (c) $8x + 6y - 47z = 5$

   (d) $5x^2 - 3 = 0$

2. Solve each of the following systems, if possible, by substitution.

   (a) $$\begin{cases} & & 3x_2 & = & 6 \\ 3x_1 & + & 3x_2 & = & 1 \end{cases}$$

   (b) $$\begin{cases} x_1 & + & 2x_2 & + & x_3 & = & 2 \\ 3x_1 & + & 8x_2 & + & x_3 & = & 12 \\ -x_1 & + & 2x_2 & - & 5x_3 & = & 2 \end{cases}$$

   (c) $$\begin{cases} x_1 & + & 2x_2 & + & x_3 & = & 2 \\ 3x_1 & + & 8x_2 & + & x_3 & = & 12 \\ -x_1 & + & 2x_2 & - & 3x_3 & = & 2 \end{cases}$$

3. Solve each of the solvable systems from (2), by applying Gaussian elimination.

4. Perform Gaussian elimination on the augmented matrix for each of the following systems. If any solution involves free variables, introduce parameters as needed. Describe the general solution to each system.

   (a) $$\begin{cases} x_1 & + & 2x_2 & + & x_3 & = & 1 \\ & & x_2 & + & x_3 & = & 1 \end{cases}$$

   (b) $$\begin{cases} x_1 & + & x_2 & + & x_3 & + & x_4 & + & x_5 & = & 5 \\ x_1 & & & & & & & + & x_5 & = & 3 \\ x_1 & - & x_2 & & & & & & & = & 3 \end{cases}$$

   (c) $$\begin{cases} -x_1 & + & 2x_2 & - & x_3 & = & -4 \\ 4x_1 & - & 6x_2 & - & x_3 & = & 7 \\ 3x_1 & + & 4x_2 & + & 2x_3 & = & 15 \end{cases}$$

5. For each of the following systems of two equations in two unknowns, perform the following steps:

(a) Sketch the system.

(b) Solve the system.

(c) Comment on the stability of the solution of either system to small changes in the coefficients.

$$(i)\left\{\begin{array}{rcl} x+y & = & 3 \\ 1.01x+y & = & 3.01 \end{array}\right. \qquad (ii)\left\{\begin{array}{rcl} x+y & = & 3 \\ 1.01x+y & = & 3.03 \end{array}\right.$$

Chapter 2

The Algebra of Matrices

Matrices were introduced in the preceding chapter as a notational convenience, but it turns out that they are interesting and useful algebraic objects in their own right. The goal of this chapter is to introduce the fundamental operations of the algebra of matrices.

2.1 Matrix Operations

It turns out that a number of operations can be defined on matrices, enabling us to treat matrices as algebraic objects. Initially, we look at multiplication by a scalar, addition, subtraction, and matrix multiplication. While division is not defined, we shall see later on that under certain circumstances we can define a multiplicative inverse for a matrix. As a notational convenience, if the i, j-entry of A is a_{ij}, we use A and (a_{ij}) interchangeably.

The simplest of matrix operations is multiplication of a matrix by a scalar. If $A = (a_{ij})$ is an $m \times n$ matrix, and if k is a scalar, then the product kA is defined by $kA = (ka_{ij})$.

EXAMPLE: Given $A = \begin{bmatrix} 2 & 0 & 1 \\ -2 & 4 & 0 \end{bmatrix}$, then
$2A = (2)\begin{bmatrix} 2 & 0 & 1 \\ -2 & 4 & 0 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 2 \\ -4 & 8 & 0 \end{bmatrix}$.

Addition of matrices is defined componentwise, i.e., if $A = (a_{ij})$ and $B = (b_{ij})$ are both $m \times n$ matrices, then their sum is $A + B = C = (c_{ij})$, where $c_{ij} = a_{ij} + b_{ij}$. The difference $A - B$ is defined by

$A - B = A + (-B) = A + (-1)B$. Thus subtraction, like addition, is performed componentwise.

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

The first matrix operation that we consider that is defined in a way that might not seem intuitively obvious is matrix multiplication. If $A = (a_{ij})$ is an $m \times n$ matrix, and if $B = (b_{ij})$ is an $n \times p$ matrix, then their product is the $m \times p$ matrix $AB = C = (c_{ij})$, where $c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$.

EXAMPLE: With $A = \begin{bmatrix} 2 & 0 & -1 \\ 1 & 1 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -1 \\ -2 & 1 \\ 0 & 5 \end{bmatrix}$, we have

$$AB = \begin{bmatrix} 2(1) + 0(-2) + (-1)(0) & 2(-1) + 0(1) + (-1)(5) \\ 1(1) + 1(-2) & 1(-1) + 1(1) \end{bmatrix} = \begin{bmatrix} 2 & -7 \\ -1 & 0 \end{bmatrix}.$$

In this example, the product BA is also defined and is found to be

$$BA = \begin{bmatrix} 1(2) + (-1)(1) & 1(0) + (-1)(1) & 1(-1) + (-1)(0) \\ (-2)(2) + 1(1) & (-2)(0) + 1(1) & (-2)(-1) + 1(0) \\ 0(2) + 5(1) & 0(0) + 5(1) & 0(-1) + 5(0) \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 \\ -3 & 1 & 2 \\ 5 & 5 & 0 \end{bmatrix}.$$

If A is an $m \times n$ matrix, then the *transpose* of A, denoted by A^T, is the $n \times m$ matrix whose columns are precisely the rows of A, i.e., if $A = (a_{ij})$, then $A^T = (b_{ij})$, where $b_{ij} = a_{ji}$.

EXAMPLE: Given $A = \begin{bmatrix} 2 & -1 & 3 \\ 0 & -3 & 3 \end{bmatrix}$, then $A^T = \begin{bmatrix} 2 & 0 \\ -1 & -3 \\ 3 & 3 \end{bmatrix}$.

The operations described above extend to transposes in a natural way, with the possible exception of multiplication of transposes. This should not be too surprising, since if A is $m \times n$ and B is $n \times p$, then while the product AB is defined, the product $A^T B^T$ is defined only if $m = p$. So it is not generally the case that $(AB)^T = A^T B^T$. On the other hand, B^T is $p \times n$, and A^T is $n \times m$, so the product $B^T A^T$ is defined. Could it be that $(AB)^T = B^T A^T$?

EXAMPLE: Suppose we have

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

$$AB = \begin{bmatrix} 5 & 8 \\ 13 & 6 \end{bmatrix}, \ A^T = \begin{bmatrix} 2 & 1 \\ 1 & 4 \\ 3 & 0 \end{bmatrix}, \text{ and } B^T = \begin{bmatrix} 1 & 3 & 0 \\ 2 & 1 & 1 \end{bmatrix}.$$

We find that

$$(AB)^T = \begin{bmatrix} 5 & 13 \\ 8 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 0 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 4 \\ 3 & 0 \end{bmatrix} = B^T A^T.$$

This is always the case. If we denote by $(AB)^T_{ij}$ the i,j-entry of $(AB)^T$ and by $(AB)_{ij}$ the i,j-entry of the product AB, then we have

$$\begin{aligned} (AB)^T_{ij} &= (AB)_{ji} \\ &= \sum_{k=1}^{n} a_{jk} b_{ki} \\ &= \sum_{k=1}^{n} a^t_{kj} b^t_{ik} \\ &= \sum_{k=1}^{n} b^t_{ik} a^t_{kj} \\ &= (B^T A^T)_{ij}, \end{aligned}$$

and it becomes clear that the general case is that $(AB)^T = B^T A^T$. As suggested previously, the other operations behave as expected, i.e., $(A + B)^T = A^T + B^T$ and $(A - B)^T = A^T - B^T$. Naturally, the transpose of the transpose of A is A, i.e., $(A^T)^T = A$.

2.2 Vectors in The Plane

This brief discussion will concern objects in the plane, but it all generalizes upward, to objects in 3-space and in higher-dimensional spaces. A vector in the plane can be viewed as a directed line segment, i.e., a line segment with an orientation that enables us to distinguish its initial point from its terminal point. Given a vector, say v, we are free to move it about in the plane; it is still v as long as we do not rotate it and so alter its orientation. For convenience, we define the standard position for a vector to be such that its initial point is the origin. Thus if, say, x is the vector with initial point $(0,0)$ and terminal point $(1,2)$, and y is the vector with initial point $(2,3)$ and terminal point $(3,5)$, we simply view y as a copy of x that has been displaced 2 units to the right and 3 units upward. But if x and y are twins, there must be some way to refer to either one, independent of location. The simplifying key lies in the fact that x has its initial point at the origin. We can then associate x with its terminal point, knowing precisely where its initial point lies. For reasons that will eventually become clear, we do not denote x by the ordered pair $(1,2)$, but instead by the 2×1 matrix $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. This is a *column vector*. The transpose of a column vector is a row vector[1], and vice versa. The notation that we adopt for the transpose of a column vector is the usual n-tuple representation, i.e., $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}^T = (x_1, x_2, \ldots, x_n)$. We will frequently make use of this as a notational convenience, principally to save space.

2.3 Vectors in $\mathbf{R}^n$ and the Scalar Product

Motivated by the preceding remarks, we may define a *vector* to be an $n \times 1$ matrix[2], where $n > 0$ is some fixed positive integer. Note that this allows us to view the individual columns (and the transposes of individual rows) in an $m \times n$ matrix as vectors. This turns out to be a remarkably useful point of view. As in the case of the coefficient and augmented matrices associated with a system of linear equations, we assume here that the individual entries, or *components*, in a vector are real numbers. The space $\mathbf{R}^n$ is the set of all such $n \times 1$ vectors. In general, the components of a matrix or of a vector are called *scalars*; for our immediate purposes a scalar is simply a real number. We will relax this later on, allowing

[1] Some authors deny the existence of row vectors, but it can be useful to allow such things to exist. Always check local rules!

[2] This is not the general definition, but is sufficient for our present purposes.

complex scalars. With experience, one is able to get by on context alone, but in order to readily distinguish vectors from scalars, in this text we label a vector with a **boldface** lowercase letter[3], e.g., $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. The sum of two vectors $\mathbf{x}$ and $\mathbf{y}$ that have the same number n of components is simply their matrix sum: $\mathbf{x} + \mathbf{y} = \mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$, where $z_i = x_i + y_i$. Thus $\begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \\ 5 \end{bmatrix}$; the difference is defined similarly, with $z_i = x_i - y_i$ in the expression above. If $\mathbf{x}$ and $\mathbf{y}$ have different numbers of components, then their sum and difference are of course undefined. The product of a scalar k and vector $\mathbf{v}$ is, in keeping with the definition of a scalar multiple of any matrix, the vector $k\mathbf{v}$, whose i^{th} component is k times the i^{th} component of $\mathbf{v}$. For example, if $\mathbf{x} = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$, then for any scalar k, $k\mathbf{x} = \begin{bmatrix} 0 \\ 2k \\ k \end{bmatrix}$.

Note that scalar multiples, sums, and differences of vectors are themselves vectors. This is not the case with the product of vectors that is of interest to us here. The product that we describe is the scalar product, or dot product. As the following definition suggests, the reason that this product is called a scalar product is that the result is a scalar rather than a vector. The other name refers to the notation that we see in the definition.

DEFINITION: If $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ are vectors, then their *dot product* is given by

$$\mathbf{x} \cdot \mathbf{y} = \sum_{k=1}^{n} x_k y_k = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

[3]This is difficult to manage on a blackboard, so other conventions are used in the classroom.

EXAMPLE: If $\mathbf{x} = \begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix}$, then we have

$$\mathbf{x} \cdot \mathbf{y} = 3 \cdot (-1) + (-2) \cdot 2 + 1 \cdot 5 = -2.$$

The dot product has a number of uses. In the Cartesian plane, for example, if we consider the vector (the geometric object, that is) with initial point at the origin and terminal point at (x_1, x_2), we can associate with this vector the algebraic vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$. By the Pythagorean theorem we know that the distance from the origin to (x_1, x_2) is given by $d(x_1, x_2) = \sqrt{x_1^2 + x_2^2}$. But this is just the square root of the dot product of $\mathbf{x}$ with itself! In linear algebra, this is called the *norm*[4] of $\mathbf{x}$, denoted $\|\mathbf{x}\|$. As we've just seen, we have $\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$. This generalizes; it matters not whether $\mathbf{x}$ has two components or two hundred. For another application with a geometric flavor, we consider the problem of computing the angle between two vectors. If $\mathbf{x}$ and $\mathbf{y}$ are two vectors with the same number, say n, of components, we can view these vectors as the algebraic representations of the points in Euclidean n-space whose coordinates are the components of $\mathbf{x}$ and $\mathbf{y}$, respectively. In $\mathbf{R}^2$, as a consequence of the Law of Cosines, it turns out that the angle θ between vectors $\mathbf{x}$ and $\mathbf{y}$ is given by

$$\cos\theta = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}.$$

This, too, generalizes to Euclidean n-space. Especially useful is the observation that two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbf{R}^2$ or $\mathbf{R}^3$ are perpendicular if and only if the angle between them is $\theta = \pi/2$ if and only if $\cos\theta = 0$ if and only if $\mathbf{x} \cdot \mathbf{y} = 0$. So we have a handy test for perpendicularity. This carries over into $\mathbf{R}^n$ for $n > 3$, although we no longer refer to such vectors as perpendicular. Instead, we refer to such vectors as being *orthogonal*.

EXAMPLE: If $\mathbf{x} = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, we find that the cosine of the angle θ between them is given by $\theta = \arccos \frac{5}{\sqrt{13}\sqrt{2}} = \arccos \frac{5}{\sqrt{26}}$ radians.

Finally, if we consider our definition of matrix multiplication, then it is easy to see that the i, j-entry of the product AB can be described as the dot product of the transpose of row i from A with column j from B.

[4] This are many vector norms in use, but this is the only one we'll mention here.

2.3.1 Properties of Matrix Operations

Matrix operations enjoy some, but not all, of the properties held by the analogous operations defined on real numbers. For example, it is easy to show that matrix addition is both commutative and associative, i.e., if A, B, and C are $m \times n$ matrices, then $A + B = B + A$ and $A + (B + C) = (A + B) + C$. Since subtraction of real numbers has neither the commutative nor the associative property, subtraction of matrices has neither property. Matrix multiplication distributes over addition (and over subtraction, since subtraction is defined in terms of addition): if the dimensions are such that the operations are defined, then $A(B + C) = AB + AC$ and $(A + B)C = AC + BC$. Matrix multiplication is associative: if the operations are defined, then $A(BC) = (AB)C$. Matrix multiplication, though, is *not* commutative, i.e., $AB \neq BA$ in general. If one considers that for A and B $m \times n$ and $n \times m$, respectively, AB is $m \times m$ and BA is $n \times n$, this might not be too surprising. But even if A and B are square matrices, commutativity can fail.

EXAMPLE: If $A = \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}$, then $AB = \begin{bmatrix} 5 & -2 \\ 4 & -2 \end{bmatrix}$, but $BA = \begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix}$.

There are a few matrices that arise so often that they deserve special attention. The $m \times n$ *zero matrix* is the $m \times n$ matrix $\mathbf{0}$, whose entries are all zeros. This is the *additive identity*: if A is any other $m \times n$ matrix, then $A + \mathbf{0} = \mathbf{0} + A = A$. For any matrix A, we may construct the *additive inverse* $-A$, where $-A = (-1)A$. It then follows that $A + (-A) = -A + A = \mathbf{0}$. The *identity matrix of order* k, denoted I_k, is a $k \times k$ matrix whose *main diagonal* entries (the entries in row i, column i for each $1 \leq i \leq k$) are 1's and all other entries are zeros. This is the *multiplicative identity*: if A is $m \times n$, then $I_m A = A I_n = A$.

EXAMPLE: The 3×3 identity matrix is $I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Given $A = \begin{bmatrix} 1 & 2 & -5 \\ 0 & 2 & -2 \\ -4 & 0 & 1 \end{bmatrix}$ it is easily verified that $AI_3 = I_3 A = A$.

Generally, the subscript notation is unnecessary for the identity matrix and is consequently omitted. We refer to the identity matrix as I, and let context determine its dimensions. So if, say, A is $m \times n$, then any reference to AI must refer to I_n, while any reference to IA must refer to I_m.

An *upper triangular* (resp. *lower triangular*) matrix is an $n \times n$ matrix in which all entries below (resp. above) the main diagonal are zeros. Note that an upper (lower) triangular matrix may have either zeros or nonzeros on and above (below) the main diagonal. If A and B are both $n \times n$ upper triangular matrices, then it is not difficult to show that the product AB is also upper triangular: if $AB = C = (c_{ij})$, then consider an entry c_{ij} for $i > j$. This is given by

$$\begin{aligned} c_{ij} &= \sum_{k=1}^{n} a_{ik}b_{kj} \\ &= \sum_{k=1}^{j} a_{ik}b_{kj} + \sum_{k=j+1}^{n} a_{ik}b_{kj} \\ &= \sum_{k=1}^{j} 0 \cdot b_{kj} + \sum_{k=j+1}^{n} a_{ik} \cdot 0 \\ &= 0. \end{aligned}$$

The analogous result holds for lower triangular matrices.

A *unit triangular* matrix is a triangular matrix having only 1's on the main diagonal. With only a little extra work, the preceding argument can be modified to show that the product of unit upper (lower) triangular matrices is unit upper (lower) triangular. This will prove useful soon, when we look at a particular matrix factorization.

Finally, some matrices enjoy the property that they have *multiplicative* inverses. Our principal concern is the inversion of square matrices. (Properly rectangular matrices can have distinct left and right inverses, but inversion of square matrices is for us the more valuable theoretical tool.)

DEFINITION: Given a square matrix A, if there exists a matrix B (necessarily square) such that $AB = BA = I$, then we say that B is the *inverse* of A, denoted by $B = A^{-1}$. Note that the inverse of a square matrix might not exist. A square matrix that possesses no inverse is called *singular*; in Chapter 3 we will find a nice theoretical characterization of singular matrices.

While actually computing such inverses presents intractable difficulties in the general case, it can be useful to obtain inverses of small matrices.

This is especially true in academic settings, where the utility of obtaining inverses of small matrices outweighs any drawbacks associated with computation of inverses in general.

2.3.2 Block-Partitioned Matrices

Suppose A is an $m \times n$ matrix. A *submatrix* of A is induced by any subset of the rows together with any subset of the columns. For example, if $A = \begin{bmatrix} 1 & 2 & -3 & 2 & 3 \\ 0 & -2 & 1 & 2 & 1 \\ 2 & 2 & 0 & 5 & 1 \\ 4 & 3 & -4 & -4 & 0 \end{bmatrix}$, the submatrix induced by rows 2 and 4 and columns 2 and 3 is $B = \begin{bmatrix} -2 & 1 \\ 3 & -4 \end{bmatrix}$. A *block* in a matrix A is a submatrix induced by contiguous rows and contiguous columns. For example, we could think of the matrix A, above, as the 2×2 *block-partitioned* matrix $A' = \begin{bmatrix} C & D \\ E & F \end{bmatrix}$, where $C = \begin{bmatrix} 1 & 2 \\ 0 & -2 \\ 2 & 2 \end{bmatrix}$, $D = \begin{bmatrix} -3 & 2 & 3 \\ 1 & 2 & 1 \\ 0 & 5 & 1 \end{bmatrix}$, $E = \begin{bmatrix} 4 & 3 \end{bmatrix}$, and $F = \begin{bmatrix} -4 & -4 & 0 \end{bmatrix}$.

It is not hard to see that if A and B are block-structured matrices with the appropriate numbers of blocks in the appropriate configurations, then the product AB can be described in terms of the blocks.

EXAMPLE: Suppose $A = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}$ and $B = \begin{bmatrix} B_1 & B_2 & B_3 \\ B_4 & B_5 & B_6 \end{bmatrix}$, with the dimensions of A_1, A_2, A_3 and A_4, respectively, given by $j \times k$, $j \times (n-k)$, $(m-j) \times k$, and $(m-j) \times (n-k)$ and with the dimensions of $B_1, \ldots, B_6$, respectively, given by $k \times l$, $k \times m$, $k \times (p-l-m)$, $(n-k) \times l$, $(n-k) \times m$, and $(n-k) \times (p-l-m)$. Then we may view the product as

$$\begin{aligned} AB &= \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} \begin{bmatrix} B_1 & B_2 & B_3 \\ B_4 & B_5 & B_6 \end{bmatrix} \\ &= \begin{bmatrix} A_1B_1 + A_2B_4 & A_1B_2 + A_2B_5 & A_1B_3 + A_2B_6 \\ A_3B_1 + A_4B_4 & A_3B_2 + A_4B_5 & A_3B_3 + A_4B_6 \end{bmatrix}. \end{aligned}$$

This view will prove useful soon, when we develop a method for computing (multiplicative) inverses of small matrices by hand, and is also useful in more advanced topics, both theoretical and applied.

2.3.3 Alternate Views of Matrix Multiplication

We can apply the results of the preceding section immediately. Matrix multiplication has been defined in terms of dot products, i.e., if A has n columns and B has n rows, then $AB = C = (c_{ij})$, then $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$, the dot product of row i of A with column j of B. Here are a couple of additional views.

A *linear combination* of $n \times 1$ vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is a sum of the form $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k$, where $c_1, c_2, \ldots, c_k$ are scalars. The set of all such combinations is called the *span* of the set $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ First, suppose that we think of A as a row of columns. That is, $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$, and suppose $\mathbf{x}$ is a vector with n components $x_1, \ldots, x_n$. Then if we expand the product $A\mathbf{x}$, we have

$$\begin{aligned} A\mathbf{x} &= \begin{bmatrix} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n \\ & \vdots & & \vdots & & \vdots & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n \end{bmatrix} \\ &= \begin{bmatrix} x_1\mathbf{a}_1 & + & x_2\mathbf{a}_2 & + & \cdots & + & x_n\mathbf{a}_n \end{bmatrix}. \end{aligned} \tag{2.1}$$

So $A\mathbf{x}$ is a linear combination of the columns of A.

Now consider the product AB, where A is $m \times n$ and B is $n \times p$. Then

$$AB = A\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_p \end{bmatrix} = \begin{bmatrix} A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_p \end{bmatrix}.$$

That is, each column of AB is a linear combination of the columns of A. We can take a similar approach in terms of rows. First consider the result of multiplying an $n \times p$ matrix B from the left by a the transpose of an $n \times 1$ vector $\mathbf{y}$. Expanding B in terms of its rows, we have

$$\mathbf{y}^T B = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix} = \begin{bmatrix} y_1\mathbf{b}_1 + y_2\mathbf{b}_2 + \cdots + y_n\mathbf{b}_n \end{bmatrix},$$

that is, $\mathbf{y}^T B$ is a linear combination of the rows of B. If we now expand A as $A = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_m^T \end{bmatrix}$, we have $AB = \begin{bmatrix} \mathbf{a}_1^T B \\ \mathbf{a}_2^T B \\ \vdots \\ \mathbf{a}_m^T B \end{bmatrix}$, i.e., the i^{th} row of AB is a linear combination of the rows of B.

2.4 Linear Independence, Rank

For a moment, suppose that we are considering vectors in the plane. If $\mathbf{x} = k\mathbf{y}$, with k a scalar, then it is not hard to see that $\mathbf{x} - k\mathbf{y} = \mathbf{0}$. It is also obvious (perhaps a sketch is necessary?) that $\mathbf{x}$ and $\mathbf{y}$ are collinear, which is to say that they lie on a common line. Since we are doomed to consider vectors in environments other than the plane, we need a way to generalize this observation. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be vectors. If the equation,

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \mathbf{0} \tag{2.2}$$

has only the trivial solution $c_1 = c_2 = \cdots = c_k = 0$, then the set $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is said to be *linearly independent*, else S is linearly dependent. Why linearly *dependent*? In the plane, it is clear. More generally, suppose (2.2) has a nontrivial solution in which, say, $c_i \neq 0$. Then we can write $\mathbf{v}_i$ as $\mathbf{v}_i = \frac{-1}{c_i}\sum\limits_{j \neq i} c_j\mathbf{v}_j$, so $\mathbf{v}_i$ can be expressed as a linear combination of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_k$, and we then say that $\mathbf{v}_i$ is dependent upon them. So in any linearly dependent set we can find at least one vector that can be described as a linear combination of the others.

When we use Gaussian elimination to transform a matrix A to row-echelon form, then the nonzero rows that remain are linearly independent, as are the columns containing the pivots. The number of linearly independent rows (or of linearly independent columns, since it turns out that these numbers match) is called the *rank* of A. This gives us a useful test for linear independence of a set $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ of vectors from $\mathbf{R}^n$: Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be the columns of a $n \times k$ matrix A. If the homogeneous equation $A\mathbf{x} = \mathbf{0}$ has nontrivial solutions, it follows from 2.1 that the vectors in S are linearly dependent. Moreover, if $\mathbf{x}$ is a nontrivial solution to $A\mathbf{x} = \mathbf{0}$, then the entries in $\mathbf{x}$ tell us how to express some element of S as a linear combination of the remaining elements of S.

EXAMPLE: Let $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ be the set of columns of $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & 1 \end{bmatrix}$. Performing Gaussian elimination, we find a row-echelon form B for A, i.e., $B = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -4 & -5 \end{bmatrix}$. The free variable is x_3. Letting $x_3 = s$ and solving $-4x_2 = 5s$, we have $x_2 = \frac{-5}{4}s$; solving $x_1 + 2x_2 = -3s$, we have $x_1 = \frac{-1}{2}s$.

So we have $\frac{-1}{2}\mathbf{a}_1 + \frac{-5}{4}\mathbf{a}_2 + \mathbf{a}_3 = \mathbf{0}$. This can be solved for any of a_1, a_2, a_3, e.g., $\mathbf{a}_3 = \frac{1}{2}\mathbf{a}_1 + \frac{5}{4}\mathbf{a}_2$, and the linear dependence of the three vectors is made explicit.

2.5 Vector Spaces and Linear Transformations

Let $f : D \to R$ be a function. If f has the property that $f(rx_1 + x_2) = rf(x_1) + f(x_2)$ for all x_1, x_2 in D, then f is said to be a *linear* function. We typically call functions $f : \mathbf{R}^n \to \mathbf{R}^m$ *linear transformations*, or *linear operators*. It turns out that if $f : \mathbf{R}^n \to \mathbf{R}^m$ is a linear transformation then there exists an $m \times n$ matrix $A = A_f$ such that $f(\mathbf{x}) = A\mathbf{x}$ for all vectors x in $\mathbf{R}^n$. So the study of linear transformations and the study of matrices are intimately connected. Before one can fully understand either, it is necessary to develop additional machinery. We offer here only thumbnail descriptions; for a full development, see any standard text on linear algebra.

A *vector space* V is a set that is closed under scalar multiplication and addition and that satisfies the following properties:

1. Commutativity of Addition: for all vectors $\mathbf{x}, \mathbf{y} \in V$,

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}.$$

2. Associativity of Addition: for all vectors $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$,

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

3. Additive identity: there exists $\mathbf{0} \in V$ such that $\mathbf{x} + \mathbf{0} = \mathbf{0} + \mathbf{x} = \mathbf{x}$ for all vectors $\mathbf{x} \in V$.

4. Additive inverse: for any $\mathbf{x} \in V$, there exists $-\mathbf{x} \in V$ such that

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

5. Distributivity: for any scalars r, s and vectors $\mathbf{u}, \mathbf{v} \in V$,
   (a) $r(\mathbf{u} + \mathbf{v}) = r\mathbf{u} + r\mathbf{v}$.
   (b) $(r + s)\mathbf{u} = r\mathbf{u} + s\mathbf{u}$.

6. Associativity of Scalar Multiplication: for any scalars r, s and vector $\mathbf{v} \in V$, $(rs)\mathbf{u} = r(s\mathbf{u})$.

7. Identity of Scalar Multiplication: for any vector $\mathbf{u} \in V$, $1\mathbf{u} = \mathbf{u}$.

Examples abound. For instance, the real numbers, $\mathbf{R}^n$ for any integer n, the set of all $n \times n$ matrices are all vector spaces. If V is a vector space, then any subset S of V that is itself a vector space is called a *subspace* of V. To illustrate, we need look no further than $\mathbf{R}^2$; a line through the origin in $\mathbf{R}^2$ is a subspace of $\mathbf{R}^2$.

Of particular importance in the study of any linear transformation $T : D \to R$ are the following sets:

1. The *kernel* of T, given by $\ker(T) = \{\mathbf{x} \in D | T(\mathbf{x}) = \mathbf{0}\}$.

2. The *range* of T, given by $\mathrm{R(T)} = \{\mathbf{y} \in \mathrm{R} | \mathbf{y} = \mathrm{T}(\mathbf{x}) \text{ for some } \mathbf{x} \in \mathrm{D}\}$.

If $T : \mathbf{R}^n \to \mathbf{R}^m$, then $T(\mathbf{x}) = A\mathbf{x}$ for an $m \times n$ matrix $A = A_T$, and the kernel of T is simply the solution set of the homogeneous equation $A\mathbf{x} = \mathbf{0}$, which we already know how to find. It is not hard to show that this is a subspace of $\mathbf{R}^n$. In the lore of linear algebra and matrix theory, this subspace is called the *nullspace* of A. The range of T is a subspace of $\mathbf{R}^m$, typically called the *range*, or *column space*, of A. We already know from §2.2.2 that if $\mathbf{y} = A\mathbf{x}$ then $\mathbf{y}$ is a linear combination of the columns of A; the column space of A is therefore the span of the set $\{\mathbf{a}_1, \mathbf{a}_2 \ldots, \mathbf{a}_n\}$ of columns of A.

We are now in a position to compactly describe the general solution to a system of equations. Rewriting the system in matrix form as $A\mathbf{x} = \mathbf{y}$, it is straightforward to see that if this is the homogeneous equation (i.e., $\mathbf{y} = \mathbf{0}$), then the general solution is simply the nullspace of A. So consider the nonhomogeneous case. Suppose that we have in hand a particular vector $\mathbf{x_0}$ that satisfies this equation. If $\mathbf{z}$ is an element of the nullspace of A, then

$$\begin{aligned} A(\mathbf{x_0} + \mathbf{z}) &= A\mathbf{x_0} + A\mathbf{z} \\ &= \mathbf{y} + \mathbf{0} \\ &= \mathbf{y}, \end{aligned}$$

and it follows that the general solution of our system is the collection of all vectors of the form $\mathbf{x} = \mathbf{x_0} + \mathbf{z}$, where $A\mathbf{x_0} = \mathbf{y}$ and $A\mathbf{z} = \mathbf{0}$.

2.6 Elementary Matrices and Matrix Inverses

If a single elementary row operation e is applied to the $n \times n$ identity matrix, the result $E = e(I)$ is called an *elementary matrix*. In general, we obtain $E_i(c)$ by multiplying row i of the identity matrix by c, E_{ij} by interchanging rows i and j of the identity matrix, and $E_{ij}(c)$ by adding c times row i to row j in the identity matrix.

EXAMPLE: If we multiply row 2 of I_3 by (-2), we obtain the matrix

$$E_2(-2) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

By interchanging rows 2 and 3 in I_3, we obtain the matrix

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

If we add twice row 1 to row 2 in I_3, we obtain

$$E_{12}(2) = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Theorem 1 *Let e be an elementary row operation with corresponding $m \times m$ elementary matrix $E = e(I)$. Then for any $m \times n$ matrix A, $e(A) = EA$, i.e., the elementary row operation can be performed on A by multiplying A from the left by E.*

EXAMPLE: Let

$$A = \begin{bmatrix} 1 & 2 & 4 & 1 \\ 2 & -2 & 3 & 3 \\ 1 & 0 & 2 & 1 \end{bmatrix}, \; E_1 = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}, E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{bmatrix}, \text{ and } E_4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

$$\text{Then } E_1 A = \begin{bmatrix} 1 & 2 & 4 & 1 \\ 0 & -6 & -5 & 1 \\ 1 & 0 & 2 & 1 \end{bmatrix}, \; E_2(E_1 A) = \begin{bmatrix} 1 & 2 & 4 & 1 \\ 0 & -6 & -5 & 1 \\ 0 & -2 & -2 & 0 \end{bmatrix},$$

$$E_3(E_2E_1A) = \begin{bmatrix} 1 & 2 & 4 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & -2 & -2 & 0 \end{bmatrix}, \text{ and } E_4(E_3E_2E_1A) = \begin{bmatrix} 1 & 2 & 4 & 1 \\ 0 & -2 & -2 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

Intuitively, any elementary row operation is easily reversed. The effect of multiplying row i by $k \neq 0$ can be reversed by multiplying row i by $1/k$. The effect of interchanging rows i and j can be reversed by repeating the interchange. The effect of adding k·row i to row j can be reversed by adding $-k$·row i to row j. For each of the three operations, we have defined an inverse operation. One might conjecture at this point that if we multiply an elementary matrix by the elementary matrix corresponding to the inverse operation we will obtain I as a result, and this is indeed the case.

EXAMPLE: If we multiply E_1, above, by $B = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, we have

$$E_1B = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3.$$

So $B = E_1^{-1}$.

Summarizing what we can say about inverses of elementary matrices, we have the following

Theorem 2 *Let E be an elementary matrix. Then E is invertible, and the inverse is given by*

$$E^{-1} = \begin{cases} E_{ij}; & E = E_{ij} \\ E_i(\frac{1}{c}); & E = E_i(c) \\ E_{ij}(-c); & E = E_{ij}(c) \end{cases}$$

The elementary matrices are not the only matrices that have inverses. For example, let $A = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}$. Then it easy to verify (do so!) that $A^{-1} = \begin{bmatrix} 2 & -3 \\ -1 & 2 \end{bmatrix}$. Do all matrices have inverses? As it turns out, the answer is no. For example, let $A = \begin{bmatrix} 2 & 4 \\ 1 & 2 \end{bmatrix}$. We will soon show that A has no inverse.

The reason for this is suggested by the following example. Suppose $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then if $ad - bc \neq 0$, let $B = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. Computing the product AB, we have

$$AB = \frac{1}{ad-bc}\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad-bc}\begin{bmatrix} ad-bc & 0 \\ 0 & ad-bc \end{bmatrix} = I.$$

It is easily verified that $BA = I$, so we have $B = A^{-1}$.

This is a good moment for a word of warning: inverses of matrices are theoretically valuable, but the explicit computation of inverses is rarely of much practical use. The computation of inverses is both expensive and difficult to accomplish accurately. What do we mean by "theoretically valuable?" Inverses allow us to state concisely certain results that would be quite difficult to state in other terms. Indeed, they allow us to *derive* results that might be difficult to state otherwise. For example, suppose that we want the solution to $A\mathbf{x} = \mathbf{b}$. If A is invertible, this is easy! The solution is $\mathbf{x} = A^{-1}\mathbf{b}$. In view of the warning just given, though, the derivation might be of little practical use. In practice, this approach to solving matrix equations is not used. We summarize in the following theorems two important properties of inverse matrices.

Theorem 3 *Let A be an $n \times n$ matrix. If A^{-1} exists, then A^{-1} is unique.*

Proof: Suppose A is $n \times n$, with inverses B and C. Then

$$\begin{aligned} B &= BI \\ &= B(AC) \\ &= (BA)C \\ &= IC \\ &= C \end{aligned}$$

□

Theorem 4 *Let A, B be $n \times n$ matrices. If A^{-1} and B^{-1} exist, then $(AB)^{-1}$ exists, and is given by*

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

Proof: By the uniqueness of inverses, it suffices to show that

$$\begin{aligned}(AB)(B^{-1}A^{-1}) &= A(B^{-1}B)A^{-1} \\ &= AIA^{-1} \\ &= AA^{-1} \\ &= I\end{aligned}$$

□

It follows inductively that

$$(A_1A_2\cdots A_k)^{-1} = A_k^{-1}A_{k-1}^{-1}\cdots A_1^{-1}.$$

We postpone to a later section (§3.2) a practical technique for computing inverses of small matrices by hand.

2.7 Exercises for Chapter 2

1. Let $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$, and $\mathbf{z} = \begin{bmatrix} 2 \\ -2 \end{bmatrix}$. Compute the following:

   (a) $\mathbf{x} + \mathbf{y}$, $\mathbf{y} + \mathbf{z}$, and $\mathbf{x} + \mathbf{z}$.

   (b) $\mathbf{x} - \mathbf{y}$ and $\mathbf{y} - \mathbf{x}$.

   (c) $2\mathbf{x} - 3\mathbf{y} + \mathbf{z}$.

   (d) $\mathbf{x} \cdot \mathbf{y}$, $\mathbf{x} \cdot \mathbf{z}$, and $\mathbf{y} \cdot \mathbf{z}$

2. Repeat (1), this time using $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$, and $\mathbf{z} = \begin{bmatrix} 2 \\ -2 \\ 3 \end{bmatrix}$.

3. For each pair of vectors from (1), compute the angle θ between those vectors.

4. Repeat (3), using the vectors from (2).

5. Let $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$, $B = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}$, and $C = \begin{bmatrix} 3 & 2 \\ -2 & 1 \end{bmatrix}$. Perform the following:

   (a) Compute the sum, difference, and product of each pair of matrices.

   (b) Compute $2A - 3B + 4C$.

   (c) Find the transpose of each matrix.

   (d) For each pair, verify that the transpose of the product is equal to the product of the transposes, but taken in reverse order. (e.g., that $(AB)^T = B^T A^T$)

6. Let $A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix}$, and $B = \begin{bmatrix} 1 & -2 & -2 \\ 0 & 1 & -1 \\ 3 & 2 & 1 \end{bmatrix}$. Repeat the operations from the preceding exercise for A and B.

7. Let $A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 1 & -1 \end{bmatrix}$, and $B = \begin{bmatrix} 1 & -2 \\ 0 & 1 \\ 3 & 2 \end{bmatrix}$. Find each of the following, or explain why the indicated operation is not defined.

   (a) $A + B$

   (b) $A + B^T$

   (c) AB

   (d) $A^T B$

8. Let A and B be as in problem #6. Are the columns of A linearly independent? The columns of B?

9. Let A and B be as in problem #7. Are the columns of A linearly independent? The columns of B? What can you say about the linear (in-)dependence of the *rows* of these matrices?

10. For each pair A, B, verify that $B = A^{-1}$.

    (a) $A = \begin{bmatrix} 1 & 2 \\ 3 & 7 \end{bmatrix}$, $B = \begin{bmatrix} 7 & -2 \\ -3 & 1 \end{bmatrix}$.

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

11. (a) Find the rank of each matrix from problems 6 and 7.

    (b) Find the rank of A, where

$$A = \begin{bmatrix} 3 & 2 & -4 & 1 & 5 \\ 6 & 4 & -7 & 3 & 1 \\ -3 & -2 & 6 & 1 & 2 \\ 9 & 6 & -11 & 4 & 6 \end{bmatrix}.$$

    What can you say about the linear independence of the columns in A? The rows?

Chapter 3

Systems of Linear Equations, Revisited

Here we take a second look at systems of linear equations and methods for their solution. We begin by bringing elementary matrices back into the picture. This enables us to invent a useful way of factoring a matrix that will facilitate the solution of systems of the form $A\mathbf{x} = \mathbf{b}_i$, where $\{\mathbf{b}_i\}$ is a sequence of vectors.

3.1 Gaussian Elimination as Matrix Factorization

Recall the example of §2.6, which is repeated here. We had

$$A = \begin{bmatrix} 1 & 2 & 4 & 1 \\ 2 & -2 & 3 & 3 \\ 1 & 0 & 2 & 1 \end{bmatrix},\ E_1 = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},\ E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix},$$

$$E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{bmatrix}, \quad \text{and } E_4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

$$\text{Then } E_1 A = \begin{bmatrix} 1 & 2 & 4 & 1 \\ 0 & -6 & -5 & 1 \\ 1 & 0 & 2 & 1 \end{bmatrix},\ E_2(E_1 A) = \begin{bmatrix} 1 & 2 & 4 & 1 \\ 0 & -6 & -5 & 1 \\ 0 & -2 & -2 & 0 \end{bmatrix},$$

$$E_3(E_2 E_1 A) = \begin{bmatrix} 1 & 2 & 4 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & -2 & -2 & 0 \end{bmatrix},\ \text{and } E_4(E_3 E_2 E_1 A) = \begin{bmatrix} 1 & 2 & 4 & 1 \\ 0 & -2 & -2 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

By the associativity of matrix multiplication, we have

$$(E_4E_3E_2E_1)A = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 1 \\ 1 & 1 & -3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 4 & 1 \\ 2 & -2 & 3 & 3 \\ 1 & 0 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 4 & 1 \\ 0 & -2 & -2 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

Note that we have the row-echelon matrix on the right-hand side as the product of two matrices. What good can this possibly do us? The key lies in the fact that we have multiplied the original matrix A from the left by a *product of elementary matrices*; we know that each of these is invertible and that each inverse is easily computed. Writing $A = E_1^{-1}E_2^{-1}E_3^{-1}E_4^{-1}B$, we have a *factored form* for A.

3.2 Computing Inverses

In Chapter 2, we encountered a formula for computing the inverse of a 2×2 matrix. Unfortunately, formulas are impractical for obtaining inverses of larger matrices. We can apply the notion of matrix factorization to obtain a conceptual view of the computation of inverses. From the section of block-structured matrices, we know that for matrices A, B, and C of the appropriate dimensions, we have

$$B \left[\begin{array}{c|c} A & C \end{array}\right] = \left[\begin{array}{c|c} BA & BC \end{array}\right].$$

In particular, if A is an invertible matrix, then

$$A^{-1} \left[\begin{array}{c|c} A & I \end{array}\right] = \left[\begin{array}{c|c} I & A^{-1} \end{array}\right].$$

Suppose we can transform $\left[\begin{array}{c|c} A & I \end{array}\right]$ to $\left[\begin{array}{c|c} I & B \end{array}\right]$ by a sequence of row operations. We can encode this as premultiplication by a sequence of elementary matrices:

$$E_kE_{k-1}\cdots E_1 \left[\begin{array}{c|c} A & I \end{array}\right] = \left[\begin{array}{c|c} I & B \end{array}\right].$$

Since $E_kE_{k-1}\cdots E_1A = I$, it follows that $E_kE_{k-1}\cdots E_1 = A^{-1}$, but then

$$B = E_kE_{k-1}\cdots E_1I = A^{-1}I = A^{-1}.$$

Thus we can compute A^{-1}, if it exists, by Gaussian elimination. If A is not invertible, we'll see later that it is impossible to transform A to I by elementary row operations, so if we simply set up the augmented matrix $\left[\begin{array}{c|c} A & I \end{array}\right]$ and apply Gaussian elimination to transform A to I, either

1. we get stuck, signaling that A^{-1} does not exist, or
2. the end result is $\left[\begin{array}{c|c} I & A^{-1} \end{array}\right]$.

EXAMPLE: First we try this on a singular matrix, anticipating that we'll get stuck. Let

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

Setting up the system,

$$\left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 2 & 4 & 0 & 1 \end{array}\right],$$

after one elimination step we have

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

and we're done; any attempt to introduce a nonzero in the $2,2$-position will overwrite with a nonzero the new zero in the $2,1$-position.

EXAMPLE: Now for an example in which the inverse exists and the process described above succeeds in transforming A to the identity. Let

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

We begin by setting up the system,

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

Subtracting twice row 1 from row 3, we have

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

Adding three times row 2 to row 3, we have

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

Adding row 3 to row 2, we have

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

Subtracting row 2 from row 1, we have

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

We finally divide row 1 by 2, obtaining

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

Expressing the preceding steps in terms of elementary matrices, we have used a sequence of matrices E_1, E_2, E_3, E_4, and E_5, where

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}, \ E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{bmatrix}, \ E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \ E_4 = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$\text{and } E_5 = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

From the discussion preceding this example, we expect to find

$$A^{-1} = E_5 E_4 E_3 E_2 E_1 = \begin{bmatrix} 3/2 & -2 & -1/2 \\ -2 & 4 & 1 \\ -2 & 3 & 1 \end{bmatrix}.$$

This is easily verified by multiplication.

It should be pointed out once again that computation of inverses is in general both expensive and unreliable. Computation of inverses of small (e.g., 2×2 or 3×3) matrices to solve well-behaved toy problems for academic purposes is routine, and the methods discussed here suffice. In practice, we find alternatives.

3.3 The LU Decomposition

Now suppose that we are faced with the need to solve $A\mathbf{x} = \mathbf{b}$ *repeatedly*, for a fixed $n \times n$ matrix A but for multiple right-hand sides $\mathbf{b}$. We can exploit the factorization discovered in the preceding section. Let's assume first that A can be reduced to row-echelon form *without row interchanges*. The reason for the assumption is that the development is simpler.

So we have elementary matrices $E_1, E_2, \ldots, E_k$, each representing the addition of a multiple of some row i to some row j, where $i > j$, and such that $E_k E_{k-1} \cdots E_1 A = U$, a row-equivalent matrix in row-echelon form. Since E_i^{-1} exists for each $1 \leq i \leq k$, it follows that $A = E_1^{-1} E_2^{-1} \cdots E_k^{-1} U$. By the action of each matrix, we can see that each of these matrices is lower triangular, and it follows that E_i^{-1} is also lower triangular. In fact, the matrices E_i and E_i^{-1} are *unit lower triangular*, which means that their main diagonal entries are all 1's. We already know that the product of unit lower triangular matrices is unit lower triangular, so we know that $E_1^{-1} E_2^{-1} \cdots E_k^{-1}$ is unit lower triangular. Letting L denote this product, we have $A = LU$, the product of a unit lower triangular matrix with an upper triangular matrix. This factorization of A is known as the *LU-decomposition* of A.

So now solving $A\mathbf{x} = \mathbf{b}$ is equivalent to solving $LU\mathbf{x} = \mathbf{b}$. Why is this important news? We can solve $LU\mathbf{x} = \mathbf{b}$ by solving a sequence of triangular systems. We first solve $L\mathbf{y} = \mathbf{b}$, using *forward substitution*, sometimes called *forward elimination*. Solving this system corresponds to preprocessing $\mathbf{b}$ using the same elementary row operations that were used to reduce A to row-echelon form. We now solve $U\mathbf{x} = y$ by back-substitution to recover the solution vector $\mathbf{x}$.

We have already seen that U is the row-echelon form of A produced by Gaussian elimination, but how do we compute L in practice? We know that L is unit lower triangular, so the diagonal and superdiagonal entries are known. We work left-to-right, recording information generated during elimination, starting in column 1. For each $i = 2, \ldots, n$, let $m_{i1} = \dfrac{a_{i1}}{a_{11}}$. By subtracting m_{i1} times row 1 from row i, $2 \leq i \leq n$, we annihilate the $i,1$-entry. Denote by $A^{(2)} = (a_{ij}^{(2)})$ the matrix obtained by annihilating all entries of column 1 below the pivot. (Remember, by our assumption there are no row interchanges.) Now for each $i = 3, \ldots n$, let $m_{i2} = \dfrac{a_{i2}^{(2)}}{a_{22}^{(2)}}$, and subtract m_{i2} times row 2 from row i, $3 \leq i \leq n$. This annihilates all entries of column 2 below the pivot. Continuing in this fashion, we end up with

$$U = \begin{bmatrix} a_{11} & a_{12} & \cdots & \cdots & a_{1n} \\ 0 & a_{22}^{(2)} & \cdots & \cdots & a_{2n}^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 0 & a_{nn}^{(n)} \end{bmatrix}.$$

Now for the best part. It can be shown, and a bit of experimentation will confirm, that

$$L = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ m_{21} & 1 & 0 & \cdots & \cdots & 0 \\ m_{31} & m_{32} & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ m_{n1} & m_{n2} & \cdots & \cdots & m_{n,n-1} & 1 \end{bmatrix}.$$

So computing the LU-decomposition is no more complicated than performing Gaussian elimination on A. We simply record the multipliers as we go.

EXAMPLE: Suppose we want to solve $A\mathbf{x} = \mathbf{b}$, where

$$A = \begin{bmatrix} 2 & 3 & -1 \\ 2 & 4 & 3 \\ 6 & 1 & 7 \end{bmatrix}, \text{ and } \mathbf{b} = \begin{bmatrix} 3 \\ -6 \\ -3 \end{bmatrix}.$$

We begin by finding the LU-decomposition of A. Initially,

$$L = \begin{bmatrix} 1 & 0 & 0 \\ ? & 1 & 0 \\ ? & ? & 1 \end{bmatrix},$$

i.e., we don't specify the entries not yet known. The multipliers for column 1 are $m_{21} = 1$ and $m_{31} = 6/2 = 3$, respectively. So we have
$A^{(2)} = \begin{bmatrix} 2 & 3 & -1 \\ 0 & 1 & 4 \\ 0 & -8 & 10 \end{bmatrix}$, and the current state of L is $L = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 3 & ? & 1 \end{bmatrix}$.
From $A^{(2)}$, we compute $m_{32} = -8$. The final form of L is
$L = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 3 & -8 & 1 \end{bmatrix}$, and the final form of U is given by $U = \begin{bmatrix} 2 & 3 & -1 \\ 0 & 1 & 4 \\ 0 & 0 & 42 \end{bmatrix}$.

We now solve $L\mathbf{y} = \mathbf{b}$, obtaining $\mathbf{y} = \begin{bmatrix} 3 \\ -9 \\ -84 \end{bmatrix}$, and then solve $U\mathbf{x} = \mathbf{y}$, obtaining $\mathbf{x} = \begin{bmatrix} 2 \\ -1 \\ -2 \end{bmatrix}$. It is easy to verify both the decomposition and the computed solution; these are left as exercises.

In summary, we have a way of decomposing an $n \times n$ matrix A into a product $A = LU$, where L is $n \times n$ unit lower triangular and U is an upper triangular matrix that is row-equivalent to A. The value of this is that we can now solve a sequence of k equations of the form $A\mathbf{x_i} = \mathbf{b}_i,\ (1 \leq i \leq k)$ by performing Gaussian elimination once at the beginning, and by then performing forward- and back-substitution once for each vector $\mathbf{b}_i$. The benefit is that, once the factorization of A as $A = LU$ is complete, we can solve $A\mathbf{x_i} = \mathbf{b}_i$ for subsequent right-hand sides $\mathbf{b}_i$ by solving a pair of triangular systems rather than by repeating the elimination. The principal limitation of the technique as described here is that the elimination must not involve row interchanges.

3.4 The Decomposition PA=LU

We now consider the case in which a triangular factorization is desired but the matrix A in question cannot be reduced to row-echelon form without row interchanges. (In hand computation, this occurs when a zero pivot is encountered.) The procedure is simple, in principle. First imagine the effect of performing a row interchange on the augmented matrix $[A|\mathbf{b}]$. If rows i and j are swapped, the interchange also affects entries i and j in $\mathbf{b}$. We can view the process as left-multiplication of $[A|\mathbf{b}]$ by the elementary matrix E_{ij}, the result being the augmented matrix $E_{ij}[A|\mathbf{b}] = [E_{ij}A|E_{ij}\mathbf{b}]$. It follows that $A\mathbf{x} = \mathbf{b}$ if and only if $E_{ij}A\mathbf{x} = E_{ij}\mathbf{b}$. Our problem here is that we want to factor A but wish to postpone consideration of right-hand sides until later. But if we are able to factor $E_{ij}A$ as $E_{ij}A = LU$, and are then presented with a vector $\mathbf{b}$, then by computing $E_{ij}\mathbf{b}$ we have subjected $\mathbf{b}$ to the same interchange already applied to A and are therefore ready to solve $LU\mathbf{x} = E_{ij}\mathbf{b}$. In the most general case, we can interchange a number of row pairs by left-multiplying A by a *permutation matrix* P, which is constructed by applying the necessary interchanges to a single copy of the identity matrix. Alternatively, P can be viewed as a product of elementary matrices, one for each necessary row interchange. As in the case of a single elementary matrix, $A\mathbf{x} = \mathbf{b}$ if and only if $PA\mathbf{x} = P\mathbf{b}$. So if we have in hand the factorization $PA = LU$, we solve $PA\mathbf{x} = P\mathbf{b}$ by first solving $L\mathbf{y} = P\mathbf{b}$, and then solving $U\mathbf{x} = \mathbf{y}$ as before.

EXAMPLE: Let $A = \begin{bmatrix} 1 & 3 & 2 & 1 \\ 1 & 3 & 2 & 2 \\ 1 & 4 & 2 & 2 \\ 2 & 7 & 5 & 6 \end{bmatrix}$. We want an LU-decomposition of A. But after subtracting row one from rows two and three, and twice row one from row four, we have $A^{(1)} = \begin{bmatrix} 1 & 3 & 2 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 4 \end{bmatrix}$. We cannot continue, so we swap rows two and three, obtaining $A^{(2)} = \begin{bmatrix} 1 & 3 & 2 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 4 \end{bmatrix}$. We now subtract row two from row four, obtaining $A^{(3)} = \begin{bmatrix} 1 & 3 & 2 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 3 \end{bmatrix}$. But this is not an upper triangular matrix, so we swap rows three and four, obtaining $U = \begin{bmatrix} 1 & 3 & 2 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. Had we subjected A to these two row swaps in advance, we could have performed elimination without interchanges. Subjecting the identity matrix I to the interchanges in question, we obtain the permutation matrix $P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$. It is easy to verify that $PA = \begin{bmatrix} 1 & 3 & 2 & 1 \\ 1 & 4 & 2 & 2 \\ 2 & 7 & 5 & 6 \\ 1 & 3 & 2 & 2 \end{bmatrix}$ can be reduced to upper triangular form without row interchanges, the result being the factorization

$$PA = LU = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 2 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 2 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Now suppose that we want to solve $A\mathbf{x} = \mathbf{b}$ for $\mathbf{b} = (0, 2, 3, 7)^T$. Since we already have $PA = LU$, we compute $P\mathbf{b} = (0, 3, 7, 2)^T$, and solve $LU\mathbf{x} = P\mathbf{b}$, obtaining the solution vector $\mathbf{x} = (-1, 1, -2, 2)^T$. (Verification is left as an exercise.)

It should be mentioned that, when solving such a system on a computer, the rows would never be physically swapped; instead, row swaps are recorded in a vector $\mathbf{p}$, which in the previous case is initialized as $\mathbf{p} = (1, 2, 3, 4)^T$. In the previous example, the final state of $\mathbf{p}$ would have been $\mathbf{p} = (1, 3, 4, 2)^T$. The code would refer, at each step, to the p_i, j-entry in A rather than the i, j-entry. This is more efficient from the standpoint of minimizing operations.

3.5 The Determinant

It is useful to associate with a square matrix $A = (a_{ij})$ a single number, $\det(A)$, called the *determinant* of A, that tells us (among other things) whether A is singular. The definition of the determinant can take a variety of forms; here we take a recursive approach that has theoretical value and is practical for small matrices. For a matrix of dimensions 1×1, i.e., $A = [a]$, we define the determinant of A by $\det(A) = a$. To define the determinant of an $n \times n$ matrix for $n > 1$, we must first develop some machinery. If A is $n \times n$, the submatrix constructed from A by deleting row i and column j has dimensions $(n-1) \times (n-1)$. The determinant of this smaller matrix is called the $i, j-$*minor* of a_{ij}, denoted by M_{ij}. The signed minor $A_{ij} = (-1)^{i+j} M_{ij}$ is called the *cofactor* of a_{ij}. We now define the determinant of A in terms of cofactors:

$$\det(A) = \sum_{j=1}^{n} a_{1j} A_{1j},$$

that is, $\det(A)$ is the sum of the products of the entries in the first row of A with their corresponding cofactors.

For example, the determinant of a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is the number $\det(A) = a \det([d]) - b \det([c]) = ad - bc$. One can derive a formula for the determinant of a 3×3 matrix using the same approach, but in general it is best to dispense with formulas and simply use the cofactor expansion. In general, computation of the determinant is best avoided altogether; the number of arithmetic operations alone make the process expensive, and it is typically unnecessary. Nevertheless, in academic settings it is useful to become adept at computing determinants of small matrices.

For the 3×3 matrix $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$, the cofactor expansion gives

$$\det(A) = a_{11} \det \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} - a_{12} \det \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix} + a_{13} \det \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix},$$

which involves evaluating three smaller determinants.

EXAMPLE: Given

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

we find $\det(A) = 1(2) - 2(-4) + (-1)(-2) = 12$.

While the definition above calls for expansion along the first row of A, it turns out that we can expand along any row or, in fact, along any column. By choosing, when possible, a row or column that has as many 0's as possible, this simplifies the work. The following theorem expresses this compactly; for a proof see any standard text on linear algebra.

Theorem 5 *For any choice of q or p,*

$$\det(A) = \sum_{j=1}^{n} a_{pj} A_{pj} = \sum_{i=1}^{n} a_{iq} A_{iq}.$$

The following theorem describes useful properties of the determinant, including the effect on the determinant of elementary row operations.

Theorem 6 *Let A be an $n \times n$ matrix. Then*

1. $\det\left(A^T\right) = \det(A)$.

2. *If A has a row (or column) of zeros, then* $\det(A) = 0$.

3. *If any row (or column) of A is multiplied by a scalar c, then the determinant is also multiplied by c, i.e.,*

$$\det\left[\mathbf{a}_1 \cdots c\mathbf{a}_j \cdots \mathbf{a}_n\right] = c \det\left[\mathbf{a}_1 \cdots \mathbf{a}_j \cdots \mathbf{a}_n\right].$$

4. *If any row of A is a multiple of another row, then $\det(A) = 0$. The statement remains true if we replace "row" with "column".*
5. *If a multiple of row i is added to row $j \neq i$, the determinant is unaffected. A similar result holds for columns.*
6. *If two rows (columns) are interchanged, the determinant is multiplied by -1, i.e., $\det(E_{ij}A) = -\det(A)$.*
7. *If A is triangular, then $\det(A) = \Pi_{i=1}^{n} a_{ii}$, i.e., the determinant of a triangular matrix is the product of its diagonal entries.*

From the preceding theorem, we see that the elementary row-interchange matrix has determinant -1, the elementary matrix obtained from I by multiplying row i by c has determinant c, and the elementary matrix in which c times row i is added to row j has determinant 1. Since row operations are equivalent to premultiplication by elementary matrices, it follows that if E is an $n \times n$ elementary matrix and A any $n \times n$ matrix, then $\det(EA) = \det(E)\det(A)$. This leads to several useful results, of which three are presented here.

Theorem 7 *An $n \times n$ matrix A is singular if and only if $\det(A) = 0$.*

Theorem 8 *If A and B are $n \times n$ matrices, then*

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

Corollary 9 *If A is an $n \times n$ nonsingular matrix, then* $\det\left(A^{-1}\right) = (\det A)^{-1}$.

Proof: Suppose A is an $n \times n$ matrix and that $\det A = k \neq 0$. Then

$$1 = \det(I) = \det\left(AA^{-1}\right) = \det(A)\det\left(A^{-1}\right) = k\det\left(A^{-1}\right).$$

Since $k \neq 0$ we have $\det\left(A^{-1}\right) = 1/k = (\det A)^{-1}$. □

Note that if $A = LU$, then since L is unit lower triangular it follows that $\det(L) = 1$. By the multiplicative property of the determinant,

$$\det(A) = \det(LU) = \det(L)\det(U) = \det(U),$$

so we see that A is singular if and only if U contains at least one diagonal zero.

We close the section on determinants with a tool of some utility called Cramer's rule. Suppose A is $n \times n$ and that we want to solve $A\mathbf{x} = \mathbf{b}$. In practice, we'll use Gaussian elimination or the LU-decomposition, but there is another way. Let $A^{(i)}$ denote the matrix obtained from A by replacing column i of A with $\mathbf{b}$. Then x_i, the i^{th} component of the solution vector $\mathbf{x}$, is given by

$$x_i = \frac{\det\left(A^{(i)}\right)}{\det(A)}.$$

EXAMPLE: Suppose we want to solve $A\mathbf{x} = \mathbf{b}$, with

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 3 \\ 2 & -2 & 0 \end{bmatrix}, \text{ and } \mathbf{b} = \begin{bmatrix} 0 \\ 6 \\ 6 \end{bmatrix}.$$

We use the cofactor expansion along column 3 to find

$$\det(A) = -3\det\begin{bmatrix} 1 & 2 \\ 2 & -2 \end{bmatrix} = 18.$$

We now compute

$$x_1 = \frac{1}{18}\begin{vmatrix} 0 & 2 & 0 \\ 6 & 1 & 3 \\ 6 & -2 & 0 \end{vmatrix} = \frac{-2}{18}(-18) = 2,$$

$$x_2 = \frac{1}{18}\begin{vmatrix} 1 & 0 & 0 \\ 2 & 6 & 3 \\ 2 & 6 & 0 \end{vmatrix} = \frac{1}{18}(-18) = -1,$$

and

$$x_3 = \frac{1}{18}\begin{vmatrix} 1 & 2 & 0 \\ 2 & 1 & 6 \\ 2 & -2 & 6 \end{vmatrix} = \frac{1}{18}(18) = 1,$$

and so the computed solution is $\mathbf{x} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$.

Cramer's Rule is more applicable in theory than in practice, since it is impractical to implement the rule for systems of more than modest size. For small systems larger than, say, 3×3, it is less expensive from the point of view of computation to use Gaussian elimination.

3.6 The Cross Product

Many problems arise in $\mathbf{R}^3$ in which it is useful to obtain a vector $\mathbf{z}$ that is perpendicular to two given vectors $\mathbf{u}$ and $\mathbf{v}$. In this setting, it turns out that we can define a *vector* product of $\mathbf{u}$ and $\mathbf{v}$ that does the job. Discussion of this product was postponed to this section for two reasons. The first is that this product is unavailable in $\mathbf{R}^k$ for $k \neq 3$, and so does not play an important role in linear algebra. The second is that it is convenient to be able to bring determinants into the discussion.

DEFINITION: Let $\mathbf{u} = (u_1, u_2, u_3)^T$ and $\mathbf{v} = (v_1, v_2, v_3)^T$ be vectors in $\mathbf{R}^3$. The *cross product* of $\mathbf{u}$ and $\mathbf{v}$ is

$$\mathbf{u} \times \mathbf{v} = (u_2v_3 - u_3v_2, u_3v_1 - u_1v_3, u_1v_2 - u_2v_1)^T.$$

A handy mnemonic device with which one can easily recall the formula for the cross product involves an abuse of the determinant notation. Let $\mathbf{e}_1 = (1,0,0)^T$, $\mathbf{e}_2 = (0,1,0)^T$, and $\mathbf{e}_3 = (0,0,1)^T$. Then we may write

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}.$$

By expanding as if we were taking a determinant, we produce the desired formula. (It is not really a determinant, of course, since the entries in the top row are not real numbers, while the elements in rows two and three are reals.)

Verification that $\mathbf{x} \times \mathbf{y}$ is perpendicular to both $\mathbf{x}$ and $\mathbf{y}$ is a simple exercise in computing the angle between vectors.

EXAMPLE: Let $\mathbf{u} = (1, -2, -1)^T$ and $\mathbf{v} = (2, 3, 1)^T$. Then

$$\mathbf{u}\times\mathbf{v} = ((-2)(1)-3(-1))\mathbf{e}_1+(2(-1)-1(1))\mathbf{e}_2+(1(3)-2(-2))\mathbf{e}_3 = (1, -3, 7)^T.$$

We can verify the result by taking dot products:
$\mathbf{u} \cdot (\mathbf{u} \times \mathbf{v}) = 1 + 6 - 7 = 0$, and $\mathbf{v} \cdot (\mathbf{u} \times \mathbf{v}) = 2 - 9 + 7 = 0$.

A number of useful identities involving the cross product are easily derived. We have already observed that

$$\mathbf{u} \cdot (\mathbf{u} \times \mathbf{v}) = \mathbf{v} \cdot (\mathbf{u} \times \mathbf{v}) = 0.$$

It is easy to show, for any vector $\mathbf{u}$, that

$$\mathbf{u} \times \mathbf{0} = \mathbf{0} \times \mathbf{u} = \mathbf{0},$$

and that

$$\mathbf{u} \times \mathbf{u} = \mathbf{0}.$$

A glance at the mnemonic that packages the cross product as a 3×3 determinant suggests that, for any vectors $\mathbf{u}$ and $\mathbf{v}$,

$$\mathbf{u} \times \mathbf{v} = -(\mathbf{v} \times \mathbf{u}).$$

This is easily verified. We can also exploit properties of determinants to show that if m is any scalar, then

$$m\mathbf{u} \times \mathbf{v} = \mathbf{u} \times m\mathbf{v} = m(\mathbf{u} \times \mathbf{v}).$$

Somewhat surprisingly, we have a pair of distributive laws for the cross product over addition:

$$\begin{aligned} \mathbf{u} \times (\mathbf{v} + \mathbf{w}) &= \mathbf{u} \times \mathbf{v} + \mathbf{u} \times \mathbf{w}, \text{ and} \\ (\mathbf{u} + \mathbf{v}) \times \mathbf{w} &= \mathbf{u} \times \mathbf{w} + \mathbf{v} \times \mathbf{w}. \end{aligned}$$

A straightforward (but tedious) bit of algebra reveals an identity due to Lagrange, which states that

$$||\mathbf{u} \times \mathbf{v}||^2 = ||\mathbf{u}||^2||\mathbf{v}||^2 - (\mathbf{u} \cdot \mathbf{v})^2.$$

Just as the dot product has a geometric interpretation in terms of the angle between vectors, we can provide a geometric interpretation of the cross product. We know that $\mathbf{u} \cdot \mathbf{v} = ||\mathbf{u}||||\mathbf{v}|| \cos\theta$, where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$. Squaring both sides, we have

$$\begin{aligned} (\mathbf{u} \cdot \mathbf{v})^2 &= ||\mathbf{u}||^2||\mathbf{v}||^2 \cos^2\theta \\ &= ||\mathbf{u}||^2||\mathbf{v}||^2(1 - \sin^2\theta) \\ &= ||\mathbf{u}||^2||\mathbf{v}||^2 - ||\mathbf{u}||^2||\mathbf{v}||^2 \sin^2\theta \end{aligned}$$

By rearranging Lagrange's identity, we have

$$(\mathbf{u} \cdot \mathbf{v})^2 = ||\mathbf{u}||^2||\mathbf{v}||^2 - ||\mathbf{u} \times \mathbf{v}||^2.$$

Comparing terms, we have

$$||\mathbf{u} \times \mathbf{v}||^2 = ||\mathbf{u}||^2||\mathbf{v}||^2 \sin^2\theta;$$

taking square roots, we finally have

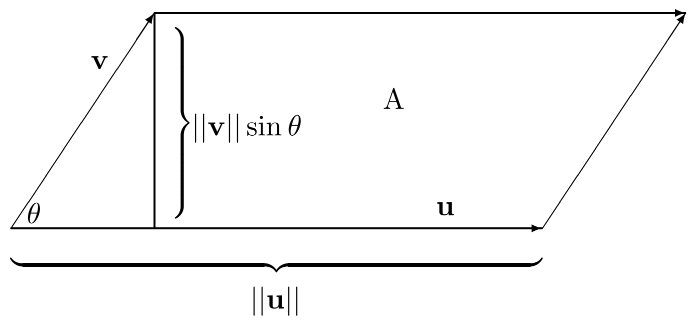

Figure 3.1: The area of the parallelogram determined by **u** and **v** is $A = ||\mathbf{u}||||\mathbf{v}|| \sin\theta = ||\mathbf{u} \times \mathbf{v}||$.

$$||\mathbf{u} \times \mathbf{v}|| = ||\mathbf{u}||||\mathbf{v}|| \sin\theta.$$

But the right-hand side of this equation can be interpreted as the area of a parallelogram determined by **u** and **v**. See Figure 3.1 for an illustration. The parallelogram lies in the plane containing **u** and **v**; for convenience, we may place **u** in standard position colinear with the horizontal axis and let θ represent the positive angle between **u** and **v**. The parallelogram in question has as its vertices the origin and the terminal points of **u**, **v**, and $\mathbf{u} + \mathbf{v}$. The width of the parallelogram is $||\mathbf{u}||$, the height is $||\mathbf{v}|| \sin\theta$, and the result follows.

3.7 Exercises for Chapter 3

1. Let $A = \begin{bmatrix} 2 & 1 \\ 5 & 3 \end{bmatrix}$, and $B = \begin{bmatrix} 4 & 3 \\ 7 & 6 \end{bmatrix}$.

   Find A^{-1} and B^{-1}, if they exist.

2. Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 3 & 4 \\ 2 & 3 & 6 \end{bmatrix}$, and $B = \begin{bmatrix} 2 & 0 & 4 \\ 1 & 2 & 1 \\ 2 & 0 & 5 \end{bmatrix}$.

   Find A^{-1} and B^{-1}, if they exist.

3. Verify that the LU factorization on pages 39–40, and the solution to the system in question, are correct.

4. Let $A = \begin{bmatrix} 4 & 3 \\ 7 & 6 \end{bmatrix}$, and $\mathbf{b} = (0, -3)^T$. Find the LU factorization of A, and use this factorization to solve $A\mathbf{x} = \mathbf{b}$.

5. Let $A = \begin{bmatrix} 1 & 1 & 0 \\ 3 & 1 & 4 \\ 2 & 3 & 5 \end{bmatrix}$, and $\mathbf{b} = (1, 3, 9)^T$. Find the LU factorization of A, and use this factorization to solve $A\mathbf{x} = \mathbf{b}$.

6. Let $A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 2 & 4 \\ 3 & 3 & 6 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 1 & 0 \\ 3 & 1 & 4 \\ 2 & 3 & 5 \end{bmatrix}$, and $C = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & -6 \\ -7 & 8 & 9 \end{bmatrix}$.

   Find $\det(A)$, $\det(B)$, and $\det(C)$, using a cofactor expansion.

7. Let $A = \begin{bmatrix} 2 & 4 \\ -1 & 2 \end{bmatrix}$, and let $\mathbf{b} = (2, -5)^T$. Solve $A\mathbf{x} = \mathbf{b}$ using Cramer's rule.

8. Let $A = \begin{bmatrix} 1 & 1 & 0 \\ 3 & 1 & 4 \\ 2 & 3 & 5 \end{bmatrix}$, and let $\mathbf{b} = (-1, 13, 11)^T$. Solve $A\mathbf{x} = \mathbf{b}$ using Cramer's rule.

9. Verify that, for any $\mathbf{x}$ and $\mathbf{y}$ in $\mathbf{R}^3$, the cross product $\mathbf{x} \times \mathbf{y}$ is perpendicular to both $\mathbf{x}$ and $\mathbf{y}$.

10. Let $\mathbf{u} = (2, -1, 1)^T$ and $\mathbf{v} = (-1, 2, 1)^T$. Find the following:

    (a) $\mathbf{u} \times \mathbf{v}$

(b) $\mathbf{v} \times \mathbf{u}$

(c) The area of the parallelogram determined by $\mathbf{u}$ and $\mathbf{v}$.

11. Let $\mathbf{e}_1, \mathbf{e}_2$, and $\mathbf{e}_3$ be as defined in the text. Verify the following:

(a) $\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3$

(b) $\mathbf{e}_2 \times \mathbf{e}_3 = \mathbf{e}_1$

(c) $\mathbf{e}_3 \times \mathbf{e}_1 = \mathbf{e}_2$

(d) $\mathbf{e}_1 \times \mathbf{e}_3 = -\mathbf{e}_2$

(e) $\mathbf{e}_3 \times \mathbf{e}_2 = -\mathbf{e}_1$

(f) $\mathbf{e}_2 \times \mathbf{e}_1 = -\mathbf{e}_3$

Chapter 4

Introduction to Eigenvalues and Eigenvectors

It has been repeatedly emphasized that the determinant, like the inverse of a nonsingular matrix and like the Cramer's rule approach to solving $A\mathbf{x} = \mathbf{b}$, is largely a theoretical tool and that one should not, in general, consider actually computing the determinant. Exceptions to this rule arise, typically in academic settings and in connection with *small* systems of equations. In this section, we see a situation in which the determinant plays an important theoretical role and in which computation of the determinant is useful for solving toy problems. We begin with such a problem.

Consider the following simple population model. Suppose that the population of California at time t_0 is 30 million, and that the population of the U.S. at time t_0 is 270 million. Furthermore, suppose that during any given year 10% of the population of California leaves and that 1% of the outside population moves in. Finally, suppose that the U.S. population is stable. We can let x_k and y_k denote the populations (in millions) of California and of the U.S. outside of California, respectively, at the end of year k. For $k \geq 1$ we now have a system of equations

$$\begin{cases} x_k &= .9x_{k-1} + .01y_{k-1} \\ y_k &= .1x_{k-1} + .99y_{k-1} \end{cases},$$

and a sequence of vectors $\mathbf{x}_0, \mathbf{x}_1, \ldots$, where $\mathbf{x}_k = \begin{bmatrix} x_k \\ y_k \end{bmatrix}$. Thus $\mathbf{x}_0 = \begin{bmatrix} 30 \\ 240 \end{bmatrix}$. Finally, observe that we can compactly write the relationship between $\mathbf{x}_k$ and $\mathbf{x}_{k-1}$ as $A\mathbf{x}_{k-1} = \mathbf{x}_k$, where $A \begin{bmatrix} .9 & .01 \\ .1 & .99 \end{bmatrix}$. It is a simple exercise to compute $\mathbf{x}_k$ for the first few values of k; we find $\mathbf{x}_1 = \begin{bmatrix} 29.4 \\ 240.6 \end{bmatrix}$, $\mathbf{x}_2 \simeq \begin{bmatrix} 28.87 \\ 241.13 \end{bmatrix}$, and $\mathbf{x}_3 \simeq \begin{bmatrix} 28.39 \\ 241.61 \end{bmatrix}$. One might ask whether there is a *steady-state* population, i.e., a population distribution $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ that satisfies $A\mathbf{x} = \mathbf{x}$. First consider what this implies. Since $A\mathbf{x} = \mathbf{x}$, it follows that $(A - I)\mathbf{x} = \mathbf{0}$. Now this equation always has the trivial solution $\mathbf{x} = 0$, but we are looking for a vector whose entries sum to 270 and is therefore nonzero. If such a vector exists, then $A - I$ must be singular. Let's check: $A - I = \begin{bmatrix} -.1 & .01 \\ .1 & -.01 \end{bmatrix}$. By inspection, the second row is a multiple of the first; Gaussian elimination results in the matrix $\hat{A} = \begin{bmatrix} -.1 & .01 \\ 0 & 0 \end{bmatrix}$. Solving the equation represented by the nonzero row, we find $\mathbf{x} = \begin{bmatrix} s \\ 10s \end{bmatrix}$ where s is an arbitrary parameter. So our population vector is of this form *and* has entries whose total is 270, from which we have a new linear equation to solve: $11s = 270$. This one is easy to solve, though, and we find $s = 270/11 \simeq 24.5455$. Thus a close approximation to our steady-state population is (in millions, again) $x = 24.5455, y = 245.455$.

4.1 Eigenvalues and Eigenvectors

Notice that this process began with our search for a particular vector $\mathbf{x}$ that was unchanged by left-multiplication by A. More generally, we might look for a scalar λ *and* a vector $\mathbf{x}$ with the property that $A\mathbf{x} = \lambda\mathbf{x}$ for a given $n \times n$ matrix A. Such a scalar is called an *eigenvalue* of A, and a vector $\mathbf{x}$ with the property that $A\mathbf{x} = \lambda\mathbf{x}$ is called an *eigenvector*. (To be precise, we typically call $\mathbf{x}$ an *eigenvector associated with* λ.) Rearranging, and using the fact that $\mathbf{x} = I\mathbf{x}$, we have

$$(A - \lambda I)\mathbf{x} = \mathbf{0}. \tag{4.1}$$

This equation always has the trivial solution, but, since we are looking for a nontrivial solution, we require that $(A - \lambda I)$ be singular. It is then a simple matter to find an eigenvector; we simply solve the homogeneous equation $(A - \lambda I)\mathbf{x} = \mathbf{0}$. Since $A - \lambda I$ is singular, Gaussian elimination reveals at least one free variable which enables us to construct a nontrivial solution $\mathbf{x}$. Actually, the existence of a nontrivial solution implies the existence of an infinite family of nontrivial solutions; this set of vectors satisfying $A\mathbf{x} = \lambda\mathbf{x}$ is called the *eigenspace* associated with λ. All this is fine, but how do we find λ? In our example at the beginning of the chapter, we experimentally discovered that $\lambda = 1$ was an eigenvalue of the matrix $A = \begin{bmatrix} .9 & .01 \\ .1 & .99 \end{bmatrix}$, but in general we don't yet know where to begin.

4.2 The Characteristic Equation

Until recently, all we knew about singularity was that it was equivalent to the existence of nontrivial solutions to the homogeneous equation. We now know, however, that a matrix A is singular if and only if $\det A = 0$, and equation 4.1 shows that this is the key to finding eigenvalues by hand. Consider the equation,

$$\det(A - \lambda I) = 0.$$

This is the *characteristic equation* of the matrix A. If one uses the cofactor expansion of the determinant to evaluate the left-hand side of the equation, then it is a simple matter to show by induction that $\det(A - \lambda I)$ is a polynomial in λ. This polynomial is called the *characteristic polynomial* of A; if A is $n \times n$, the characteristic polynomial of A has degree n. It follows that finding eigenvalues for A reduces to finding zeros of the characteristic polynomial $\det(A - \lambda I)$. (At least in principle, this is something that we can do, although in practice it presents serious difficulties.) From experience, we know that we might find repeated zeros. For now, though, we assume that all zeros are distinct. The other case is slightly more complicated, and is left for a subsequent course.

Consider the matrix $A = \begin{bmatrix} .9 & .01 \\ .1 & .99 \end{bmatrix}$ from the beginning of the chapter. We know from the earlier discussion that $\lambda_1 = 1$ is an eigenvalue and that $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 10 \end{bmatrix}$ is an eigenvector associated with λ_1. Is there another eigenvalue? We begin by considering

$$\det(A - \lambda I) = \det \begin{bmatrix} .9 - \lambda & .01 \\ .1 & .99 - \lambda \end{bmatrix} = \lambda^2 - 1.89\lambda + .89,$$

whose roots are $\lambda_1 = 1$ and $\lambda_2 = .89$. Aha! The second eigenvalue is revealed. So there must be a vector $\mathbf{x}$ with the property that $A\mathbf{x} = .89\mathbf{x}$.

To find such a vector, we construct the matrix

$$A - .89I = \begin{bmatrix} .01 & .01 \\ .1 & .1 \end{bmatrix}.$$

As in the previous case, this matrix is visibly singular (It had better be singular, right? That's how we derived this technique.) and the equation $(A - .89I)\mathbf{x} = \mathbf{0}$ has infinitely many solutions, which collectively constitute the eigenspace of A associated with $\lambda = .89$. One element of this eigenspace is the vector $\mathbf{x}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. We can scale this, multiplying by the scalar of our choice, but the sum of the entries will always be zero. From this we can conclude that this eigenvalue/eigenvector pair has no significance in our population model, since populations are implicitly nonnegative. Had this matrix arisen in some other context, this pair might have had some significance.

Let's try another 2×2. Let $A = \begin{bmatrix} 2 & 1 \\ 2 & 3 \end{bmatrix}$. The characteristic polynomial is

$$\det \begin{bmatrix} 2-\lambda & 1 \\ 2 & 3-\lambda \end{bmatrix} = \lambda^2 - 5\lambda + 4,$$

with zeros $\lambda_1 = 1$ and $\lambda_2 = 4$. To find an eigenvector associated with λ_1, we construct the matrix

$$A - \lambda_1 I = A - I = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$$

As expected, this is visibly singular; solving $(A - I)\mathbf{x} = \mathbf{0}$ we find $\mathbf{x} = s\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and can take for a representative eigenvector the vector $\mathbf{x}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. It is easy to verify that $A\mathbf{x}_1 = \mathbf{x}_1$. Now we look for an eigenvector associated with $\lambda_2 = 4$. We begin by constructing the matrix

$$A - 4I = \begin{bmatrix} -2 & 1 \\ 2 & -1 \end{bmatrix}$$

This time solving $(A - 4I)\mathbf{x} = \mathbf{0}$, we find $\mathbf{x} = s\begin{bmatrix} 1 \\ 2 \end{bmatrix}$, and can take for our representative eigenvector $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. As before, we check, finding that

$$A\mathbf{x}_2 = \begin{bmatrix} 2 & 1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \end{bmatrix} = 4\mathbf{x}_2,$$

as expected.

We now look at a 3×3 example. Let $A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & -3 & 1 \\ 0 & 0 & 1 \end{bmatrix}$. The characteristic equation, in factored form, is $(\lambda - 2)(\lambda + 3)(\lambda - 1) = 0$, with roots $\lambda_1 = 2, \lambda_2 = -3$, and $\lambda_3 = 1$.

Solving $(A - \lambda_1)\mathbf{x} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -5 & 1 \\ 0 & 0 & -1 \end{bmatrix} \mathbf{x} = \mathbf{0}$, we have $x_3 = x_2 = 0$ and $x_1 = s$, where s is an arbitrary scalar. So we might choose $\mathbf{x}_1 = (1, 0, 0)^T$.

Solving $(A - \lambda_2)\mathbf{x} = \begin{bmatrix} 5 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 4 \end{bmatrix} \mathbf{x} = \mathbf{0}$, we have $x_3 = 0$, $x_2 = s$, and $x_1 = -s/5$, where s is an arbitrary scalar. So we might choose $\mathbf{x}_2 = (-1, 5, 0)^T$.

Finally solving $(A - \lambda_3)\mathbf{x} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -4 & 1 \\ 0 & 0 & 0 \end{bmatrix} \mathbf{x} = \mathbf{0}$, we have $x_3 = s$, $x_2 = s/4$, and $x_1 = -s/4$, where once again s is an arbitrary scalar. So we might choose $\mathbf{x}_3 = (-1, 1, 4)^T$.

To summarize the procedure for finding the eigenvalues and eigenvectors of an $n \times n$ matrix A:

1. Construct the characteristic polynomial $\det(A - \lambda I)$.

2. Solve the characteristic equation $\det(A - \lambda I) = 0$. The roots $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A.

3. For each eigenvalue λ_i, solve the homogeneous equation $(A - \lambda_i I)\mathbf{x} = \mathbf{0}$; since $(A - \lambda_i I)$ is singular, there will be infinitely many solutions, each of which is an eigenvector associated with λ_i.

4.3 Exercises for Chapter 4

1. In each of the following, verify that the given vectors are eigenvectors for the given matrix. What are the associated eigenvalues?

   (a) $A = \begin{bmatrix} -1 & 8 \\ 1 & 1 \end{bmatrix}$, $\mathbf{x}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} -4 \\ 1 \end{bmatrix}$.

   (b) $A = \begin{bmatrix} 1 & 30 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, $\mathbf{x}_1 = \begin{bmatrix} 0 \\ 0 \\ -3 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 25 \\ -5 \\ 1 \end{bmatrix}$ $\mathbf{x}_3 = \begin{bmatrix} 36 \\ 6 \\ 1 \end{bmatrix}$.

2. For each of the following 2×2 matrices,

   (a) Find the characteristic polynomial, in factored form.

   (b) Find all eigenvalues.

   (c) For each eigenvalue λ, find an associated eigenvector $\mathbf{x}$, and verify that $A\mathbf{x} = \lambda\mathbf{x}$.

   (a) $A = \begin{bmatrix} -3 & 4 \\ 1 & 0 \end{bmatrix}$

   (b) $B = \begin{bmatrix} -3 & 12 \\ 1 & 1 \end{bmatrix}$

   (c) $C = \begin{bmatrix} 2 & -16 \\ -1 & -4 \end{bmatrix}$

3. Repeat exercise (2), but with the following matrices:

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

   (b) $B = \begin{bmatrix} 4 & 5 & -6 \\ 1 & 2 & 0 \\ 0 & 1 & 2 \end{bmatrix}$

   (c) $C = \begin{bmatrix} 3 & 1 & 0 \\ 19 & 1 & 1 \\ -20 & 0 & 1 \end{bmatrix}$

Chapter 5

An Introduction to Complex Arithmetic

The need to identify the eigenvalues of a square matrix is one of many problems that call for us to locate the zeros of a polynomial. We know that the quadratic formula can be used to find the real zeros of a quadratic polynomial, if they exist. Consider the equation, $x^2 + 1 = 0$. Applying the quadratic formula, we find $x = \pm\frac{1}{2}\sqrt{-4} = \pm\sqrt{-1}$. Clearly there is no real solution. What can we do?

Recall that a polynomial $p(x)$ with real coefficients is called *irreducible* over the reals if $p(x)$ cannot be factored as a product $p(x) = s(x)t(x)$, where $s(x)$ and $t(x)$ are polynomials with real coefficients and with positive degree smaller than that of $p(x)$. Any linear polynomial is clearly irreducible, since only constants have smaller degree. The polynomial $x^2 + 1$ from the preceding paragraph is an example of an irreducible quadratic polynomial. It turns out that, while $x^2 + 1$ is irreducible over the reals, it is reducible over a larger field that contains the reals. To see how this is done, we must develop some machinery.

5.1 Introduction, Fundamental Operations

The *imaginary unit*, denoted here by i, is given by $i = \sqrt{-1}$. (Always remember that $i^2 = -1$, since this is a key to doing complex arithmetic.) Since $i = \sqrt{-1}$, it follows that if $k > 0$ then

$$\sqrt{-k} = \sqrt{(-1)k} = \sqrt{-1}\sqrt{k} = i\sqrt{k}.$$

An *imaginary number* is any nonzero multiple of i. A *complex number* is any number of the form $z = a + bi$, where a and b are real numbers. We call a the *real part* of z and b the *imaginary part* of z, denoted $\mathrm{re}(z)$ and $\mathrm{im}(z)$, respectively. If $\mathrm{im}(z) = 0$, then z is real; conversely, if $\mathrm{re}(z) = 0$, then z is imaginary. Complex numbers z and w are equal if and only if $\mathrm{re}(z) = \mathrm{re}(w)$ and $\mathrm{im}(z) = \mathrm{im}(w)$, i.e., $a + bi = c + di$ if and only if $a = c$ and $b = d$. Addition and subtraction of complex numbers are performed componentwise, i.e., $(a + bi) \pm (c + di) = (a \pm c) + (b \pm d)i$. If $z = (a + bi)$ and $w = (c + di)$, then their product is computed by treating them as ordinary binomials, i.e.,
$zw = (ac + adi + bci + bdi^2) = (ac - bd) + (ad + bc)i$.
EXAMPLE: Let $z = 2 + 3i$ and $w = 1 - i$. Then $z + w = 3 + 2i$, $z - w = 1 + 4i$, and $zw = 2 - 2i + 3i - 3i^2 = 5 + i$.

5.2 The Complex Plane

We can associate with each complex number a unique point in the Cartesian plane, by treating one axis (usually the horizontal axis) as the real axis and the other axis as the imaginary axis. We refer to the plane, labeled in this fashion, as the *complex plane*. The number $a + bi$ is then associated with the point (a, b). We now have visual reinforcement of the fact that the real numbers are just a special case of the complex numbers, i.e., $\mathbf{R} = \{a + bi | b = 0\}$. A similar comment applies to the imaginary numbers. This also gives us a nice geometric interpretation for addition of complex numbers: if $z = a + bi$ and $w = c + di$, then the four points $\mathbf{0}, z, w$, and $z + w$ become the vertices of a parallelogram.
EXAMPLE: Let $z = 2 + 3i$, and $w = 1 + 2i$. Then $z + w = 3 + 5i$, and points $(0, 0), (2, 3), (1, 2)$, and $(3, 5)$ describe a parallelogram in the complex plane.

If we first of all associate with the number $z = a + bi$ the directed line segment with initial point at the origin and terminal point at (a, b), then it is easy to see that the length of the segment can be calculated using the ordinary Pythagorean distance function; in this context, we call this length the *modulus* of z, denoted $|z|$ and given by $|z| = \sqrt{a^2 + b^2}$. The modulus is also often called the *magnitude*. As a special case, we have the absolute value of a real number x, given by $|x| = \sqrt{x^2}$.

The representation of a complex number described above is typically called the *rectangular form*, since we have described it in the context of rectangular coordinates. There are two other useful representations. We can first use the modulus to help develop the *polar form* of a complex number. Suppose $z = a + bi$, and let θ be the angle between the positive real axis and the directed line segment with initial point at the origin and terminal point at (a, b). Let $r = |z|$. Then it follows that the polar form for z is $z = r[\cos\theta + i\sin\theta]$. The angle θ is called the *argument* of z, and is easily computed by $\theta = \arctan\frac{b}{a}$ [1]. Note that, because of the periodicity of the sine and cosine functions, if $z = r(\cos\theta + i\sin\theta)$, then $z = r(\cos(\theta + 2k\pi) + i\sin(\theta + 2k\pi))$ for any integer k. We can now make geometric sense of multiplication. If $z_1 = r(\cos\theta + i\sin\theta)$ and $z_2 = s(\cos\phi + i\sin\phi)$, then

$$\begin{aligned} z_1 z_2 &= rs(\cos\theta + i\sin\theta)(\cos\phi + i\sin\phi) \\ &= rs(\cos\theta\cos\phi - \sin\theta\sin\phi) + i(\cos\theta\sin\phi + \sin\theta\cos\phi); \end{aligned}$$

It follows from fundamental trigonometric identities that

$$z_1 z_2 = rs\left[\cos(\theta + \phi) + i\sin(\theta + \phi)\right]. \tag{5.1}$$

So the modulus of the product is the product of the moduli, and the argument of the product is the sum of the arguments.

EXAMPLE: Let $z_1 = 2 + 2i$ and $z_2 = 2i$. Then $z_1 = 2\sqrt{2}(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4})$ and $z_2 = 2(\cos\frac{\pi}{2} + i\sin\frac{\pi}{2})$. So $z_1 z_2$ ought to work out to be $z_1 z_2 = 4\sqrt{2}(\cos\frac{3\pi}{4} + i\sin\frac{3\pi}{4})$. Checking via ordinary multiplication, we find

$$z_1 z_2 = 4i + 4i^2 = -4 + 4i = 4\sqrt{2}(\frac{-1}{\sqrt{2}} + i\frac{1}{\sqrt{2}}) = 4\sqrt{2}(\cos\frac{3\pi}{4} + i\sin\frac{3\pi}{4}).$$

If $z = a + bi$ is a complex number, then the *conjugate* of z, denoted $\overline{z}$, is given by $\overline{z} = a - bi$. So we have $\operatorname{im}(\overline{z}) = -\operatorname{im}(z)$. Geometrically, it follows that we have reflected z across the real axis; if $z = r(\cos\theta + i\sin\theta)$, then $\overline{z} = r(\cos\theta - i\sin\theta) = r(\cos(-\theta) + i\sin(-\theta))$.

From the ealier discussion, we have $\arg(\overline{z}) = -\arg(z)$. It should be clear that $|\overline{z}| = |z|$. Note, too, that for any complex number $z = a + bi$, the product $z\overline{z} = (a + bi)(a - bi) = a^2 - b^2 i^2 = a^2 + b^2$ is real. So now we have a compact representation of the modulus: for any complex number z, $|z| = \sqrt{z\overline{z}}$.

We can now consider complex division. If $z = a + bi$, then

$$z^{-1} = \frac{1}{z} = \frac{\overline{z}}{z\overline{z}} = \frac{a - bi}{a^2 + b^2}.$$

[1] If θ does not lie in the interval $[-\pi/2, \pi/2]$, the angle returned by the arctangent function will be in error; many computer languages include a function called arctan2, which takes as its arguments the sin and cosine of θ, thus ensuring that the angle returned lies in the correct quadrant.

For example, consider the quotient, $\frac{2+i}{1-2i}$. By the preceding paragraph, we have

$$\frac{2+i}{1-2i} = \frac{(2+i)(1+2i)}{(1-2i)(1+2i)} = \frac{2+4i+i-2}{1+4} = \frac{5i}{5} = i.$$

Recall that this chapter began with an example in which we were faced with the irreducible quadratic polynomial x^2+1. We are now in a position to factor x^2+1 as the product of two linear polynomials with complex coefficients: $x^2+1=(x+i)(x-i)$. This is no accident, as shown by the following theorem.

Theorem 10 [The Fundamental Theorem of Algebra] *Every polynomial with complex coefficients factors completely as a product of linear polynomials with complex coefficients.*

As a corollary, we get that every polynomial with *real* coefficients factors completely as a product of linear polynomials with complex coefficients.

EXAMPLE: Let $p(x)=x^5-x^4+3x^3-3x^2+2x-2$. We see experimentally that $x=1$ is a zero; using long or synthetic division we can divide $p(x)$ by $x-1$, obtaining the factorization $p(x)=(x-1)(x^4+3x^2+2)$. The factor of degree four is itself factorable, and we obtain $p(x)=(x-1)(x^2+2)(x^2+1)$. If we insist on real coefficients, we can go no further, since the two quadratic factors are irreducible over the reals. But if we allow complex coefficients, the theorem guarantees success. Applying the results above, we finally have a complete factorization:

$$p(x)=(x-1)(x-i\sqrt{2})(x+i\sqrt{2})(x-i)(x+i).$$

5.3 DeMoivre's Theorem and Euler's Identity

Note that if $z=r(\cos\theta+i\sin\theta)$, then it follows from (5.1) that

$$\begin{aligned} z^2 &= r^2(\cos 2\theta + i\sin 2\theta), \\ z^3 &= r^3(\cos 3\theta + i\sin 3\theta), \end{aligned}$$

and, by induction,

$$z^n = r^n(\cos n\theta + i\sin n\theta) \qquad (5.2)$$

This extremely useful result is known as DeMoivre's Theorem, and gives us one way to approach the matter of taking roots of complex numbers.

5.3.1 Powers and Roots

Powers of complex numbers are easily calculated, *if* we are in polar form, by applying DeMoivre's Theorem. They can be somewhat awkward otherwise. Roots are a bit more work in any case, although their existence is guaranteed by the Fundamental Theorem of Algebra. We first consider roots of real numbers, then imaginary numbers, and finally complex numbers in general.

Suppose we want to solve the equation, $x^n = c$, i.e., we want the n^{th} root(s) of c. First suppose that c is real. We know that we get either no real roots (if $c < 0$ and n is even) or one real root (if n is odd or $c = 0$), or two real roots (if n is even and $c > 0$). There is no obvious pattern in these results, but, by allowing complex roots, a remarkable pattern emerges.

Let's start with $c = 1$. We know, from the preceding discussion, that

$$x^n = 1 = \begin{cases} \cos 0 + i \sin 0 \\ \cos 2\pi + i \sin 2\pi \\ \cos 4\pi + i \sin 4\pi \\ \vdots \\ \cos 2(n-1)\pi + i \sin 2(n-1)\pi \end{cases}.$$

By DeMoivre's Theorem, we have

$$x = \sqrt[n]{x^n} = \begin{cases} \cos 0 + i \sin 0 \\ \cos \frac{2\pi}{n} + i \sin \frac{2\pi}{n} \\ \cos \frac{4\pi}{n} + i \sin \frac{4\pi}{n} \\ \vdots \\ \cos 2\frac{(n-1)\pi}{n} + i \sin 2\frac{(n-1)\pi}{n} \end{cases}.$$

So the n^{th} *roots of unity* are evenly spaced about the unit circle in the complex plane. If n is even, we get two *real* roots ± 1; if n is odd, we get one real root, namely 1 itself.

For example, suppose we want to find the sixth roots of unity. By the preceding discussion, these are
$1, -1, \cos\frac{\pi}{3} + i \sin\frac{\pi}{3}, \cos\frac{2\pi}{3} + i \sin\frac{2\pi}{3}, \cos\frac{4\pi}{3} + i \sin\frac{4\pi}{3}$, and $\cos\frac{5\pi}{3} + i \sin\frac{5\pi}{3}$.

What if $c \neq 1$? Still assuming that c is real, we move from the unit circle to the circle with radius equal to the positive, real n^{th} root of the modulus (absolute value) of c and proceed in the same fashion. For example, the fifth roots of 32 are
$2, 2(\cos\frac{2\pi}{5} + i \sin\frac{2\pi}{5}), 2(\cos\frac{4\pi}{5} + i \sin\frac{4\pi}{5}), 2(\cos\frac{6\pi}{5} + i \sin\frac{6\pi}{5})$, and $2(\cos\frac{8\pi}{5} + i \sin\frac{8\pi}{5})$.

If c is not real, then things generalize nicely. For example, suppose we want to find the cube roots of $8i$. We know that

$$8i = \begin{cases} 8(\cos\frac{\pi}{2} + i\sin\frac{\pi}{2}) \\ 8(\cos(\frac{\pi}{2}+2\pi) + i\sin(\frac{\pi}{2}+2\pi)) = 8(\cos\frac{5\pi}{2} + i\sin\frac{5\pi}{2}) \\ 8(\cos(\frac{\pi}{2}+4\pi) + i\sin(\frac{\pi}{2}+4\pi)) = 8(\cos\frac{9\pi}{2} + i\sin\frac{9\pi}{2}) \end{cases}$$

Applying our new tools, we have

$$\sqrt[3]{8i} = \begin{cases} 2(\cos\frac{\pi}{6} + i\sin\frac{\pi}{6}) \\ 2(\cos\frac{5\pi}{6} + i\sin\frac{5\pi}{6}) \\ 2(\cos\frac{9\pi}{6} + i\sin\frac{9\pi}{6}) = 2(\cos\frac{3\pi}{2} + i\sin\frac{3\pi}{2}) \end{cases}.$$

The most general case occurs when we want to find the nth roots of a complex number z, where both re(z) and im(z) are nonzero. We can exploit the polar form to find these roots quite easily, as formalized in the following theorem:

Theorem 11 *Every nonzero complex number $z = r(\cos\theta + i\sin\theta)$ has n distinct n^{th} roots for any positive integer n. Each has modulus $\sqrt[n]{r}$, and the arguments are the angles $\dfrac{\theta + 2k\pi}{n}$ for $k = 0, 1, \ldots, n-1$.*

EXAMPLE: Suppose we want to find the cube roots of $z = 1 + i$. In polar form, $z = \sqrt{2}(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4})$. It follows from the preceding theorem that we will find three cube roots of z; these are

$$\sqrt[6]{2}\left(\cos\frac{\pi}{12} + i\sin\frac{\pi}{12}\right), \sqrt[6]{2}\left(\cos\frac{3\pi}{4} + i\sin\frac{3\pi}{4}\right), \text{ and } \sqrt[6]{2}\left(\cos\frac{17\pi}{12} + i\sin\frac{17\pi}{12}\right).$$

We close with one more representation for a complex number. Students familiar with calculus will recall that the MacLaurin series expansions for the cosine and sine functions are, respectively,

$$\cos x = \sum_{k=0}^{\infty}(-1)^k\frac{x^{2k}}{(2k)!} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots$$

$$\text{and } \sin x = \sum_{k=0}^{\infty}(-1)^k\frac{x^{2k+1}}{(2k+1)!} = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots,$$

and that the Maclaurin expansion for the natural exponential function is given by

$$e^x = \sum_{k=0}^{\infty}\frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots.$$

So consider the polar form of a complex number. We have

$$\begin{aligned}
\cos x + i\sin x &= \sum_{k=0}^{\infty}(-1)^k\frac{x^{2k}}{(2k)!} + i\sum_{k=0}^{\infty}(-1)^k\frac{x^{2k+1}}{(2k+1)!} \\
&= \sum_{k=0}^{\infty}(-1)^k\left(\frac{x^{2k}}{(2k)!} + i\frac{x^{2k+1}}{(2k+1)!}\right) \\
&= 1 + ix - \frac{x^2}{2!} - \frac{ix^3}{3!} + \frac{x^4}{4!} + \frac{ix^5}{5!} - \frac{x^6}{6!} - \frac{ix^7}{7!} + \cdots \\
&= 1 + ix + \frac{(ix)^2}{2!} + \frac{(ix)^3}{3!} + \frac{(ix)^4}{4!} + \frac{(ix)^5}{5!} + \frac{(ix)^6}{6!} + \frac{(ix)^7}{7!} + \cdots \\
&= e^{ix},
\end{aligned}$$

an *exponential*, which leads us to *Euler's identity*, given by

$$e^{i\theta} = \cos\theta + i\sin\theta. \tag{5.3}$$

This is an extremely useful result. Computing roots, powers, products, and quotients can now be reduced to familiar exponential operations.

EXAMPLE: The exponential representations for $i, -1$, and $-i$ are $e^{i\frac{\pi}{2}}$, $e^{i\pi}$, and $e^{i\frac{3\pi}{2}}$, respectively.

EXAMPLE: The fifth roots of 32 are the numbers $2e^{\frac{i2k\pi}{5}}$ $(0 \le k \le 4)$, which are $2, 2e^{\frac{i2\pi}{5}}, 2e^{\frac{i4\pi}{5}}, 2e^{\frac{i6\pi}{5}}$, and $2e^{\frac{i8\pi}{5}}$.

EXAMPLE: The product of $(1+i)$ and $(\sqrt{3}+i)$, in rectangular form, is

$$(1+i)(\sqrt{3}+i) = (\sqrt{3}-1) + (\sqrt{3}+1)i.$$

In polar form, we have

$$\begin{aligned}
\left(\sqrt{2}\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right)\right)\left(2\left(\cos\frac{\pi}{6} + i\sin\frac{\pi}{6}\right)\right) &= 2\sqrt{2}\left(\cos\left(\frac{\pi}{4}+\frac{\pi}{6}\right) + i\sin\left(\frac{\pi}{4}+\frac{\pi}{6}\right)\right) \\
&= 2\sqrt{2}\left(\cos\frac{5\pi}{12} + i\sin\frac{5\pi}{12}\right).
\end{aligned}$$

In exponential form, we have

$$\sqrt{2}e^{i\frac{\pi}{4}}2e^{i\frac{\pi}{6}} = 2\sqrt{2}e^{i(\frac{\pi}{4}+\frac{\pi}{6})} = 2\sqrt{2}e^{i\frac{5\pi}{12}}.$$

5.4 Complex Eigenvalues and Complex Eigenvectors

In Chapter 4, we gave an introduction to eigenvalues and eigenvectors in which the method for finding eigenvalues of a matrix A involved finding the roots of the characteristic equation of A. We now have the tools in hand to consider eigenvalues and eigenvectors of matrices whose characteristic equations possess nonreal complex roots. For example, consider $A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$. The characteristic equation is

$$\lambda^2 - 2\lambda + 2 = 0,$$

and the roots are the complex eigenvalues

$$\lambda_1 = 1 + i, \ \lambda_2 = 1 - i.$$

We first look for an eigenvector associated with $\lambda_1 = 1 + i$, by forming the matrix

$$A - \lambda_1 I = \begin{bmatrix} -i & 1 \\ -1 & -i \end{bmatrix}.$$

The second row is $+i$ times the first, so the matrix is singular, as expected. The first row gives $ix_1 = x_2$, so any eigenvector is of the form $\mathbf{x} = s \begin{bmatrix} 1 \\ i \end{bmatrix}$; letting $s = 1$, we have $\mathbf{x}_1 = \begin{bmatrix} 1 \\ i \end{bmatrix}$.

Now we look for an eigenvector associated with $\lambda_2 = 1 - i$. We form the matrix

$$A - \lambda_2 I = \begin{bmatrix} i & 1 \\ -1 & i \end{bmatrix}.$$

As before, the matrix is singular; the second row is i times the first. The second row gives $x_1 = ix_2$, so any eigenvector associated with λ_2 therefore has the form $\mathbf{x} = s \begin{bmatrix} i \\ 1 \end{bmatrix}$. Letting $s = 1$, we have the eigenvector $\mathbf{x}_2 = \begin{bmatrix} i \\ 1 \end{bmatrix}$.

We leave as an exercise the verification that $A\mathbf{x}_1 = \lambda_1 \mathbf{x}_1$ and show here that $A\mathbf{x}_2 = \lambda_2 \mathbf{x}_2$.

$$A\mathbf{x}_2 = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} i \\ 1 \end{bmatrix} = \begin{bmatrix} 1+i \\ 1-i \end{bmatrix} = (1-i) \begin{bmatrix} i \\ 1 \end{bmatrix}.$$

If $\mathbf{x}$ is a vector with (possibly) complex components, we denote by $\overline{\mathbf{x}}$ the vector whose components are the conjugates of the components of $\mathbf{x}$, that is, if $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$, then $\overline{\mathbf{x}} = (\overline{x_1}, \overline{x_2}, \ldots, \overline{x_n})^T$.

As long as the entries in a matrix A are real, then the quadratic formula guarantees that complex eigenvalues occur in conjugate pairs. It might not be immediately obvious that the corresponding eigenvectors also occur in conjugate pairs. But this is easily verified. Let A be a real matrix, with a complex eigenvalue λ and corresponding eigenvector $\mathbf{x}$. Then

$$\begin{aligned} A\overline{\mathbf{x}} &= \overline{A}\overline{\mathbf{x}} && \text{(since A is real, then } \overline{A} = A) \\ &= \overline{A\mathbf{x}} && \text{(the conjugate of a product is the product of the conjugates)} \\ &= \overline{\lambda\mathbf{x}} && (A\mathbf{x} = \lambda\mathbf{x}) \\ &= \overline{\lambda}\overline{\mathbf{x}} \end{aligned}$$

So if A has a conjugate pair of complex eigenvalues, the work involved in finding the corresponding eigenvectors is cut in half. This can be seen in the preceding example, and again in the following example.

EXAMPLE: Let $A = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$. The characteristic equation of A is

$$\lambda^2 - 2\lambda + 5 = 0,$$

with roots $\lambda_1 = 1 + 2i$ and $\lambda_2 = 1 - 2i$. Let's first find an eigenvector $\mathbf{x}_1$ for λ_1. We have

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

The first row is easily seen to be i times the second, verifying singularity. Solving $(A - (1+2i)I)\mathbf{x} = 0$, we find that $x_1 + ix_2 = 0$, so $x_1 = -ix_2$. We take $\mathbf{x}_1 = (-i, 1)^T$. It follows from the preceding discussion that an eigenvector for λ_2 is $\mathbf{x}_2 = (i, 1)^T$. Verification is left as an exercise.

5.5 Exercises for Chapter 5

1. Let $z = 2 + 3i$, $w = 1 - 4i$, and $y = 3 - i$. Find the following:

   (a) $z + w, z + y$, and $w + y$

   (b) $z - w, w - z, z - y, y - z, w - y$, and $y - w$

   (c) zw, zy, and wy

2. Let z, w, and y be as in exercise 1. Find $|z|, |w|$, and $|y|$.

3. Let z, w, and y be as in exercise 1. Plot z, w, y and the results of exercise 1 in the complex plane.

4. Let $z = 3 - 4i$. Find $\overline{z}$, and verify that $|z| = \sqrt{z\overline{z}}$.

5. Let $z = 2 + 2i$ and $w = \sqrt{3} - i$. Find z/w and w/z.

6. Let $z = -3 + i$. Find z^{-1}.

7. Let $z = \sqrt{2} + i\sqrt{2}$, $w = 2\sqrt{3} - 2i$, and $y = -\sqrt{3} + 3i$. Find the modulus and argument of each.

8. Express the following in polar form: $z = 1/\sqrt{2} + i/\sqrt{2}$, $w = -1/\sqrt{2} + i/\sqrt{2}$, and $y = 1 + i\sqrt{3}$.

9. Let z, w, and y be as in exercise 7. Use their polar forms to find the products zw, zy, and wy.

10. Let z and w be as in exercise 7. Use their polar forms to find the quotients z/w and w/z.

11. Factor completely as products of linear factors:

    (a) $p(x) = x^2 + 9$.

    (b) $q(x) = x^2 + 6$.

    (c) $r(x) = x^2 - 4x + 13$.

    (d) $s(x) = x^2 + x + 1$.

12. Find the exponential form for each of z, w, and y from exercise 7. Use their exponential forms to find their products zw, zy, and wy.

13. Use the exponential forms from the preceding problem to find $\overline{z}$, $\overline{w}$, and $\overline{y}$.

14. Again using the exponential forms from problem 10, find z/w, w/z, z/y, y/z, w/y, and y/w.

15. Find all roots of the following equations:

    (a) $z^4 = 1$

    (b) $z^3 = 8$

16. Verify that the vectors found in the final example in this chapter are in fact eigenvectors for the given matrix.

17. Let $A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} i \\ 1 \end{bmatrix}$. Verify that $A\mathbf{x} = (1-i)\mathbf{x}$.

18. Let $A = \begin{bmatrix} 0 & -2 \\ 4 & 4 \end{bmatrix}$. Find all eigenvalues and associated eigenspaces of A.

19. As in the preceding problem, but this time using $A = \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}$.

Appendix A

Solutions to the Exercises

A.1 Solutions to Exercises for Chapter 1

1. Only (c) is linear.

2. We are to solve each system by substitution, if possible.

   (a) $$\begin{cases} & & 3x_2 & = & 6 \\ 3x_1 & + & 3x_2 & = & 1 \end{cases}$$

   Solution: From the first equation, $x_2 = 2$. Substituting into the second equation, we have $3x_1 + 3(2) = 3x_1 + 6 = 1$, so $3x_1 = -5$ and $x_1 = -5/3$.

   (b) $$\begin{cases} x_1 & + & 2x_2 & + & x_3 & = & 2 \\ 3x_1 & + & 8x_2 & + & x_3 & = & 12 \\ -x_1 & + & 2x_2 & - & 5x_3 & = & 2 \end{cases}$$

   Solution: No solution exists. If one goes through the steps required to discover a solution, in the end one is faced with an impossibility of the form $0x_2 = 4$.

   (c) $$\begin{cases} x_1 & + & 2x_2 & + & x_3 & = & 2 \\ 3x_1 & + & 8x_2 & + & x_3 & = & 12 \\ -x_1 & + & 2x_2 & - & 3x_3 & = & 2 \end{cases}$$

   Solution: Solving the first equation for x_1, we find $x_1 = 2 - 2x_2 - x_3$. Substituting this for x_1 in the second equation, we have $x_2 - x_3 = 3$, or $x_2 = 3 + x_3$, from which we have now have $x_1 = -4 - 3x_3$. We can then eliminate both x_1

and x_2 from the third equation, arriving at $x_3 = -4$. It follows that $x_2 = 3 - 4 = -1$ and $x_1 = -4 - 3(-4) = 8$.

3. We now use Gaussian elimination to solve systems (a) and (c) from the preceding problem.

(a) $$\begin{cases} \quad\quad 3x_2 = 6 \\ 3x_1 + 3x_2 = 1 \end{cases}$$

Solution: There is really no work to be done here, since x_1 is absent from the first equation.

(c) $$\begin{cases} x_1 + 2x_2 + x_3 = 2 \\ 3x_1 + 8x_2 + x_3 = 12 \\ -x_1 + 2x_2 - 3x_3 = 2 \end{cases}$$

Solution: Subtracting three times row one from row two, and adding row one to row three, we obtain a new system:

$$\begin{cases} x_1 + 2x_2 + x_3 = 2 \\ \quad\quad 2x_2 - 2x_3 = 6 \\ \quad\quad 4x_2 - 2x_3 = 4 \end{cases}$$

Now subtracting twice row two from row three, we have

$$\begin{cases} x_1 + 2x_2 + x_3 = 2 \\ \quad\quad 2x_2 - 2x_3 = 6 \\ \quad\quad\quad\quad 2x_3 = -8 \end{cases}$$

Backsubstitution gives us $x_3 = -4$, $x_2 = \frac{1}{2}(6 - 8) = -1$, and $x_1 = 2 + 2 + 4 = 8$.

4. We are to perform Gaussian elimination on the augmented matrix for each of the following systems, obtaining the general solution.

(a) $$\begin{cases} x_1 + 2x_2 + x_3 = 1 \\ \quad\quad x_2 + x_3 = 1 \end{cases}$$

(b) $$\begin{cases} x_1 + x_2 + x_3 + x_4 + x_5 = 5 \\ x_1 + x_5 = 3 \\ x_1 - x_2 = 3 \end{cases}$$

(c) $$\begin{cases} -x1 + 2x_2 - x_3 = -4 \\ 4x_1 - 6x_2 - x_3 = 7 \\ 3x_1 + 4x_2 + 2x_3 = 15 \end{cases}$$

Solution:

(a) The augmented matrix for system (a) is

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

Elimination has no effect, since the matrix is already in row-echelon form. The free variable is x_3, so we let $x_3 = s$ and proceed with back-substitution: $x_2 = 1 - s$, and $x_1 = 1 - s - 2(1 - s) = -1 + s$. The general solution is then

$$(x_1, x_2, x_3) = (-1 + s, 1 - s, s).$$

(b) The augmented matrix for system (b) is

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 5 \\ 1 & 0 & 0 & 0 & 1 & 3 \\ 1 & -1 & 0 & 0 & 0 & 3 \end{bmatrix}.$$

Introducing zeros below the pivot in the first column, we have

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

Introducing a zero below the pivot in the second column, we have

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 5 \\ 0 & -1 & -1 & -1 & 0 & -2 \\ 0 & 0 & 1 & 1 & -1 & 2 \end{bmatrix},$$

and the elimination is complete. Free variables are x_4 and x_5, so we set $x_4 = s$ and $x_5 = t$, say. It follows that $x_3 = 2 - s + t$, $x_2 = 2 - s - (2 - s + t) = -t$, and $x_1 = 5 + t - (2 - s + t) - s - t = 3 - t$. So the general solution is

$$(x_1, x_2, x_3, x_4, x_5) = (3 - t, -t, 2 - s + t, s, t).$$

(c) The augmented matrix for system (c) is

$$A = \begin{bmatrix} -1 & 2 & -1 & -4 \\ 4 & -6 & -1 & 7 \\ 3 & 4 & 2 & 15 \end{bmatrix}.$$

Introducing zeros below the pivot in the first column, we have

$$A = \begin{bmatrix} -1 & 2 & -1 & -4 \\ 0 & 2 & -5 & -9 \\ 0 & 10 & -1 & 3 \end{bmatrix}.$$

Introducing a zero below the pivot in the second column, we obtain

$$A = \begin{bmatrix} -1 & 2 & -1 & -4 \\ 0 & 2 & -5 & -9 \\ 0 & 0 & 24 & 48 \end{bmatrix}.$$

There are no free variables, so we continue with backsubstitution, obtaining

$$(x_1, x_2, x_3) = (3, 1/2, 2).$$

5. For each of the following systems of two equations in two unknowns, we are to perform the following steps:

   (a) Sketch the system.

   (b) Solve the system.

   (c) Comment on the stability of the solution of either system to small changes in the coefficients.

$$(i) \left\{ \begin{array}{rcl} x + y & = & 3 \\ 1.01x + y & = & 3.01 \end{array} \right. \qquad (ii) \left\{ \begin{array}{rcl} x + y & = & 3 \\ 1.01x + y & = & 3.03 \end{array} \right.$$

Solution:

(a) The lines $y = 3 - x$ and $y = 3.01 = 1.01x$ described by the first system are so nearly identical that any sketch that would fit on this page would appear to contain only one line. The same applies to the lines $y = 3 - x$ and $y = 3.03 - 1.01x$ described by the second system. So the sketches are not presented here.

(b) The solutions are $x = 1, y = 2$ for the first system and $x = 3, y = 0$ for the second.

(c) The second system can be viewed as having been obtained from the first by making a small change in the right-hand side of the second equation; this resulted in a dramatic change in the solution. Thus we may conclude that the solution to either system is extremely sensitive to small changes in the coefficients.

A.2 Solutions to Exercises for Chapter 2

1. We are given $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$, and $\mathbf{z} = \begin{bmatrix} 2 \\ -2 \end{bmatrix}$, and must compute the following:

   (a) $\mathbf{x} + \mathbf{y} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}$, $\mathbf{y} + \mathbf{z} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$, and $\mathbf{x} + \mathbf{z} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$.

   (b) $\mathbf{x} - \mathbf{y} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$, and $\mathbf{y} - \mathbf{x} = \begin{bmatrix} -3 \\ -1 \end{bmatrix}$.

   (c) $2\mathbf{x} - 3\mathbf{y} + \mathbf{z} = \begin{bmatrix} 10 \\ -1 \end{bmatrix}$.

   (d) The dot products:

$$\begin{aligned} x \cdot y &= 1 \cdot (-2) + 2 \cdot 1 = -2 + 2 = 0 \\ x \cdot z &= 1 \cdot 2 + 2 \cdot (-2) = 2 - 4 = -2 \\ y \cdot z &= -2 \cdot 2 + 1 \cdot (-2) = -4 - 2 = -6 \end{aligned}$$

   (e)

$$\begin{aligned} x \cdot y &= 1 \cdot (-2) + 2 \cdot 1 + (-1) \cdot 0 = -2 + 2 + 0 = 0 \\ x \cdot z &= 1 \cdot 2 + 2 \cdot (-2) + (-1) \cdot 3 = 2 - 4 - 3 = -5 \\ y \cdot z &= -2 \cdot 2 + 1 \cdot (-2) + 0 \cdot 3 = -4 - 2 + 0 = -6 \end{aligned}$$

2. We repeat (1), this time using $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$, and $\mathbf{z} = \begin{bmatrix} 2 \\ -2 \\ 3 \end{bmatrix}$, obtaining

(a) $\mathbf{x}+\mathbf{y} = \begin{bmatrix} -1 \\ 3 \\ -1 \end{bmatrix}$, $\mathbf{y}+\mathbf{z} = \begin{bmatrix} 0 \\ -1 \\ 3 \end{bmatrix}$, and $\mathbf{x}+\mathbf{z} = \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix}$.

(b) $\mathbf{x}-\mathbf{y} = \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}$ and $\mathbf{y}-\mathbf{x} = \begin{bmatrix} -3 \\ -1 \\ 1 \end{bmatrix}$.

(c) $2\mathbf{x}-3\mathbf{y}+\mathbf{z} = \begin{bmatrix} 10 \\ -1 \\ 1 \end{bmatrix}$

3. For each pair of vectors from (1), we compute the angle θ between those vectors:

   (a) The angle between $\mathbf{x}$ and $\mathbf{y}$ is $\theta = \cos^{-1} 0 = \dfrac{\pi}{2}$.

   (b) The angle between $\mathbf{y}$ and $\mathbf{z}$ is $\theta = \cos^{-1}\left(\dfrac{-3}{\sqrt{10}}\right) \approx 2.8198$.

   (c) The angle between $\mathbf{x}$ and $\mathbf{z}$ is $\theta = \cos^{-1}\left(\dfrac{-1}{\sqrt{10}}\right) \approx 1.8925$.

4. Repeating (3), using the vectors from (2), we find

   (a) The angle between $\mathbf{x}$ and $\mathbf{y}$ is $\theta = \cos^{-1} 0 = \dfrac{\pi}{2}$.

   (b) The angle between $\mathbf{y}$ and $\mathbf{z}$ is $\theta = \cos^{-1}\left(\dfrac{-6}{\sqrt{85}}\right) \approx 2.2794$.

   (c) The angle between $\mathbf{x}$ and $\mathbf{z}$ is $\theta = \cos^{-1}\left(\dfrac{-5}{\sqrt{102}}\right) \approx 2.0887$.

5. We are given

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

   and are to perform the following:

   (a) Compute the sum, difference, and product of each pair of matrices.

   (b) Compute $2A - 3B + 4C$.

   (c) Find the transpose of each matrix.

(d) For each pair, verify that the transpose of the product is equal to the product of the transposes, but taken in reverse order. (e.g., that $(AB)^T = B^T A^T$)

Solution:

(a) i. The sums: $A+B = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$, $A+C = \begin{bmatrix} 4 & 4 \\ -2 & 2 \end{bmatrix}$, and $B+C = \begin{bmatrix} 4 & 0 \\ -2 & 2 \end{bmatrix}$.

ii. Some differences: $A-B = \begin{bmatrix} 0 & 4 \\ 0 & 0 \end{bmatrix}$, $B-C = \begin{bmatrix} -2 & -4 \\ 2 & 0 \end{bmatrix}$

iii. Some products: $AB = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $BC = \begin{bmatrix} 7 & 0 \\ -2 & 1 \end{bmatrix}$, $CB = \begin{bmatrix} 3 & -4 \\ -2 & 5 \end{bmatrix}$.

(b) $2A - 3B + 4C = \begin{bmatrix} 11 & 18 \\ -8 & 3 \end{bmatrix}$.

(c) The transposes:
$A^T = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$, $B^T = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}$, and $C^T = \begin{bmatrix} 3 & -2 \\ 2 & 1 \end{bmatrix}$.

(d) Using B and C, we get

$$B^T C^T = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 3 & -2 \\ -4 & 5 \end{bmatrix} = (CB)^T.$$

6. We are given $A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix}$, and $B = \begin{bmatrix} 1 & -2 & -2 \\ 0 & 1 & -1 \\ 3 & 2 & 1 \end{bmatrix}$, and are to repeat operations (a), (c), and (d) from the preceding exercise, using A and B.

Solution:

(a) $A+B = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 2 & -2 \\ 4 & 3 & 2 \end{bmatrix}$, $A-B = \begin{bmatrix} 0 & 4 & 6 \\ 0 & 0 & 0 \\ -2 & -1 & 0 \end{bmatrix}$,

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

$$AB = \begin{bmatrix} 13 & 8 & 0 \\ -3 & -1 & -2 \\ 4 & 1 & -2 \end{bmatrix}, \text{ and } BA = \begin{bmatrix} -1 & -2 & 4 \\ -1 & 0 & -2 \\ 4 & 9 & 11 \end{bmatrix}.$$

(c) $A^T = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 4 & -1 & 1 \end{bmatrix}$.

(d) $B^T A^T = \begin{bmatrix} 13 & -3 & 4 \\ 8 & -1 & 1 \\ 0 & -2 & -2 \end{bmatrix} = (AB)^T$.

7. Let $A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 1 & -1 \end{bmatrix}$, and $B = \begin{bmatrix} 1 & -2 \\ 0 & 1 \\ 3 & 2 \end{bmatrix}$. Find each of the following, or explain why the indicated operation is not defined.

   (a) $A + B$ is undefined, since the dimensions of the matrices do not agree.

   (b) $A + B^T = \begin{bmatrix} 2 & 2 & 7 \\ -2 & 2 & 1 \end{bmatrix}$.

   (c) $AB = \begin{bmatrix} 13 & 8 \\ -3 & -1 \end{bmatrix}$.

   (d) $A^T B$ is undefined, since the number of columns in A^T is not equal to the number of rows in B.

8. Yes, in both A and B.

9. The columns of A are a linearly dependent set, but the rows of A are linearly independent. The columns of B are a linearly independent set, while the rows of B are linearly dependent.

10. Simply verify that $AB = BA = I$.

11. (a) The matrices from problem (6) both have rank 3. Those from problem (7) both have rank 2.

    (b) The row echelon form of A is $\begin{bmatrix} 3 & 2 & -4 & 1 & 5 \\ 0 & 0 & 1 & 1 & -9 \\ 0 & 0 & 0 & 0 & 25 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$. From this we see that A has rank 3. The columns of A are linearly dependent, as are the rows. We can also see that the set consisting of columns 1,3, and 5 is linearly independent, as is the set consisting of rows 1,2, and 3.

A.3 Solutions to Exercises for Chapter 3

1. Given $A = \begin{bmatrix} 2 & 1 \\ 5 & 3 \end{bmatrix}$, and $B = \begin{bmatrix} 4 & 3 \\ 7 & 6 \end{bmatrix}$, we must find A^{-1} and B^{-1} if they exist.

   Solution: Both are invertible. Following are the steps for finding A^{-1}; the same procedure applied to B will produce B^{-1}.

   We start with the matrix

$$\left[\, A \,|\, I \,\right] = \left[\begin{array}{cc|cc} 2 & 1 & 1 & 0 \\ 5 & 3 & 0 & 1 \end{array}\right].$$

   After the initial elimination step, we have the matrix

$$\left[\begin{array}{cc|cc} 2 & 1 & 1 & 0 \\ 0 & 1/2 & -5/2 & 1 \end{array}\right].$$

   Multiplying row two by 2 and subtracting from row one, we have

$$\left[\begin{array}{cc|cc} 2 & 0 & 6 & -2 \\ 0 & 1 & -5 & 2 \end{array}\right].$$

   Finally dividing row one by 2, we have

$$\left[\begin{array}{cc|cc} 1 & 0 & 3 & -1 \\ 0 & 1 & -5 & 2 \end{array}\right], \text{ revealing } A^{-1} = \begin{bmatrix} 3 & -1 \\ -5 & 2 \end{bmatrix}.$$

2. Given $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 3 & 4 \\ 2 & 3 & 6 \end{bmatrix}$, and $B = \begin{bmatrix} 2 & 0 & 4 \\ 1 & 2 & 1 \\ 2 & 0 & 5 \end{bmatrix}$, we are again to find A^{-1} and B^{-1} if they exist.

   Solution: As in the preceding problem, both matrices are invertible and, as before, we construct A^{-1}.

Our starting point is

$$\left[\begin{array}{c|c} A & I \end{array}\right] = \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 3 & 4 & 0 & 1 & 0 \\ 2 & 3 & 6 & 0 & 0 & 1 \end{array}\right].$$

Introducing zeros below the pivot in column one, we have

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 3 & 4 & -1 & 1 & 0 \\ 0 & 3 & 6 & -2 & 0 & 1 \end{array}\right].$$

Introducing a zero below the pivot in column two leads to

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

Subtracting row three from row two to put a zero above the pivot in column three, we have

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

Finally, dividing row three by 2 and row two by 3, we obtain

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1/3 & 1 & -2/3 \\ 0 & 0 & 1 & -1/2 & -1/2 & 1/2 \end{array}\right],$$

from which we know that

$$A^{-1} = \left[\begin{array}{ccc} 1 & 0 & 0 \\ 1/3 & 1 & -2/3 \\ -1/2 & -1/2 & 1/2 \end{array}\right].$$

3. Verify that the LU factorization on pages 39–40, and the solution to the system in question, are correct.

   Solution: It suffices to show that $LU = A$ and that $A\mathbf{x} = \mathbf{b}$:

$$LU = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 3 & -8 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 & -1 \\ 0 & 1 & 4 \\ 0 & 0 & 42 \end{bmatrix} = \begin{bmatrix} 2 & 3 & -1 \\ 2 & 4 & 3 \\ 6 & 1 & 7 \end{bmatrix} = A,$$

and

$$A\mathbf{x} = \begin{bmatrix} 2 & 3 & -1 \\ 2 & 4 & 3 \\ 6 & 1 & 7 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ -2 \end{bmatrix} = \begin{bmatrix} 3 \\ -6 \\ -3 \end{bmatrix} = \mathbf{b}.$$

4. Let $A = \begin{bmatrix} 4 & 3 \\ 7 & 6 \end{bmatrix}$, and $\mathbf{b} = (0, -3)^T$. Find the LU factorization of A, and use this factorization to solve $A\mathbf{x} = \mathbf{b}$.

Solution: The multiplier used in reducing $A = \begin{bmatrix} 4 & 3 \\ 7 & 6 \end{bmatrix}$ to the upper triangular form $U = \begin{bmatrix} 4 & 3 \\ 0 & 3/4 \end{bmatrix}$ is $m_{21} = 7/4$. It follows that $L = \begin{bmatrix} 1 & 0 \\ 7/4 & 1 \end{bmatrix}$. We now use the decomposition to solve the indicated system: solving $L\mathbf{y} = \mathbf{b}$, we have $\mathbf{y} = (0, -3)^T$. Solving $U\mathbf{x} = \mathbf{y}$, we find $\mathbf{x} = (3, -4)^T$.

5. Let $A = \begin{bmatrix} 1 & 1 & 0 \\ 3 & 1 & 4 \\ 2 & 3 & 5 \end{bmatrix}$, and $\mathbf{b} = (1, 3, 9)^T$. Find the LU factorization of A, and use this factorization to solve $A\mathbf{x} = \mathbf{b}$.

Solution: The multipliers used in reducing $A = \begin{bmatrix} 1 & 1 & 0 \\ 3 & 1 & 4 \\ 2 & 3 & 5 \end{bmatrix}$ to the upper-triangular form $U = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -2 & 4 \\ 0 & 0 & 7 \end{bmatrix}$ are $m_{21} = 3$, $m_{31} = 2$, and $m_{32} = -1/2$, so $L = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & -1/2 & 1 \end{bmatrix}$. Solving $L\mathbf{y} = \mathbf{b}$, we have $\mathbf{y} = (1, 0, 7)^T$. Solving $U\mathbf{x} = \mathbf{y}$, we obtain $\mathbf{x} = (-1, 2, 1)^T$.

6. Given $A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 2 & 4 \\ 3 & 3 & 6 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 1 & 0 \\ 3 & 1 & 4 \\ 2 & 3 & 5 \end{bmatrix}$, and
$C = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & -6 \\ -7 & 8 & 9 \end{bmatrix}$, we must find $\det(A)$, $\det(B)$, and $\det(C)$, using cofactor expansions.

Solution: Expanding along row one to exploit the 0's, we find

$$\det(A) = 2\det \begin{bmatrix} 2 & 4 \\ 3 & 6 \end{bmatrix} = 2(12 - 12) = 0.$$

Either row one or column three would be best for B; using row one, we have

$$\det(B) = \det \begin{bmatrix} 1 & 4 \\ 3 & 5 \end{bmatrix} - \det \begin{bmatrix} 3 & 4 \\ 2 & 5 \end{bmatrix} = -7 - 7 = -14.$$

There being no 0's in C, there are no shortcuts. Using row one, we find

$$\det(C) = \det \begin{bmatrix} 5 & -6 \\ 8 & 9 \end{bmatrix} - 2\det \begin{bmatrix} 4 & -6 \\ -7 & 9 \end{bmatrix} + 3\det \begin{bmatrix} 4 & 5 \\ -7 & 8 \end{bmatrix} = 306.$$

7. Let $A = \begin{bmatrix} 2 & 4 \\ -1 & 2 \end{bmatrix}$, and let $\mathbf{b} = (2, -5)^T$. Solve $A\mathbf{x} = \mathbf{b}$ using Cramer's rule.

Solution: Using the notation from the text, we have

$$A^{(1)} = \begin{bmatrix} 2 & 4 \\ -5 & 2 \end{bmatrix}, \text{ and } A^{(2)} = \begin{bmatrix} 2 & 2 \\ -1 & -5 \end{bmatrix}.$$

With $\det(A) = 8$, $\det(A^{(1)}) = 24$, and $\det(A^{(2)}) = -8$, we have $x_1 = 3$ and $x_2 = -1$, i.e., $\mathbf{x} = (3, -1)^T$.

8. Let $A = \begin{bmatrix} 1 & 1 & 0 \\ 3 & 1 & 4 \\ 2 & 3 & 5 \end{bmatrix}$, and let $\mathbf{b} = (-1, 13, 11)^T$. Solve $A\mathbf{x} = \mathbf{b}$ using Cramer's rule.

Solution: Again using the notation from the text, we have

$$A^{(1)} = \begin{bmatrix} -1 & 1 & 0 \\ 13 & 1 & 4 \\ 11 & 3 & 5 \end{bmatrix},\ A^{(2)} = \begin{bmatrix} 1 & -1 & 0 \\ 3 & 13 & 4 \\ 2 & 11 & 5 \end{bmatrix},\ \text{and } A(3) = \begin{bmatrix} 1 & 1 & -1 \\ 3 & 1 & 13 \\ 2 & 3 & 11 \end{bmatrix}.$$

Their determinants are $-14, 28$, and -42, respectively, and the determinant of A is -14. It follows that $x_1 = 1$, $x_2 = -2$, and $x_3 = 3$, so $\mathbf{x} = (1, -2, 3)^T$.

9. We verify that $(\mathbf{x} \times \mathbf{y}) \cdot \mathbf{x} = 0$:

$$\begin{aligned} (\mathbf{x} \times \mathbf{y}) \cdot \mathbf{x} &= (x_2y_3 - x_3y_2, x_3y_1 - x_1y_3, x_1y_2 - x_2y_1)^T \cdot (x_1, x_2, x_3)^T \\ &= x_1x_2y_3 - x_1x_3y_2 + x_2x_3y_1 - x_1x_2y_3 + x_1y_2x_3 - x_2y_1x_3 \\ &= x_1x_2y_3 - x_1x_2y_3 + x_1x_3y_2 - x_1x_3y_2 + x_2x_3y_1 - x_2x_3y_1 \\ &= 0. \end{aligned}$$

10. With $\mathbf{u} = (2, -1, 1)^T$ and $\mathbf{v} = (-1, 2, 1)^T$, we have

    (a) $\mathbf{u} \times \mathbf{v} = (-3, -3, 3)^T$.

    (b) $\mathbf{v} \times \mathbf{u} = (3, 3, -3)^T$.

    (c) We can do this in a couple of ways. First, since we already know from (a) that $\mathbf{u} \times \mathbf{v} = (-3, -3, 3)^T$, we can easily compute $||\mathbf{u} \times \mathbf{v}|| = \sqrt{27} = 3\sqrt{3}$. Alternatively, if θ is the angle between $\mathbf{u}$ and $\mathbf{v}$, then $\cos\theta = \dfrac{\mathbf{u} \cdot \mathbf{v}}{||\mathbf{u}||||\mathbf{v}||} = -\dfrac{1}{2}$, so $\theta = \dfrac{2\pi}{3}$. It follows that $\sin\theta = \dfrac{\sqrt{3}}{2}$, and finally the area of the parallelogram determined by $\mathbf{u}$ and $\mathbf{v}$ is $||\mathbf{u} \times \mathbf{v}|| = 3\sqrt{3}$.

11. With $\mathbf{e}_1 = (1, 0, 0)^T$, $\mathbf{e}_2 = (0, 1, 0)^T$, and $\mathbf{e}_3 = (0, 0, 1)^T$, we have $\mathbf{e}_1 \times \mathbf{e}_2 = (0 - 0, 0 - 0, 1 - 0)^T = \mathbf{e}_1$. The other verifications are equally straightforward.

A.4 Solutions to Exercises for Chapter 4

1. In each of the following, we are to verify that the given vectors are eigenvectors for the given matrix and to identify the associated eigenvalues.

(a) $A = \begin{bmatrix} -1 & 8 \\ 1 & 1 \end{bmatrix}$, $\mathbf{x}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} -4 \\ 1 \end{bmatrix}$.

(b) $A = \begin{bmatrix} 1 & 30 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, $\mathbf{x}_1 = \begin{bmatrix} 0 \\ 0 \\ -3 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 25 \\ -5 \\ 1 \end{bmatrix}$ $\mathbf{x}_3 = \begin{bmatrix} 36 \\ 6 \\ 1 \end{bmatrix}$.

Solution:

(a) $A\mathbf{x}_1 = (6,3)^T = 3\mathbf{x}_1$, and $A\mathbf{x}_2 = (12,-3)^T = -3\mathbf{x}_2$; eigenvalues are 3 and -3.

(b) $A\mathbf{x}_1 = (0,0,0)^T = 0\mathbf{x}_1$, $A\mathbf{x}_2 = (-125,25,-5)^T = -5\mathbf{x}_2$, $A\mathbf{x}_3 = (216,36,6)^T = 6\mathbf{x}_3$; eigenvalues are 0, -5, and 6.

2. For each of the following 2×2 matrices, we are to

(a) Find the characteristic polynomial, in factored form.

(b) Find all eigenvalues.

(c) For each eigenvalue λ, find an associated eigenvector $\mathbf{x}$, and verify that $A\mathbf{x} = \lambda\mathbf{x}$.

(a) $A = \begin{bmatrix} -3 & 4 \\ 1 & 0 \end{bmatrix}$

(b) $B = \begin{bmatrix} -3 & 12 \\ 1 & 1 \end{bmatrix}$

(c) $C = \begin{bmatrix} 2 & -16 \\ -1 & -4 \end{bmatrix}$

Solution:

(a) The characteristic polynomial for A is $\det(A - \lambda I) = \lambda^2 + 3\lambda - 4 = (\lambda + 4)(\lambda - 1)$. The eigenvalues are 1 and -4. Solving $(A - I)\mathbf{x} = \mathbf{0}$, we find that any eigenvector is of the form $(s,s)^T$, so $(1,1)^T$ will do nicely as an eigenvector for $\lambda = 1$. Solving $(A + 4I)\mathbf{x} = \mathbf{0}$, we find that the associated eigenvectors are of the form $(4s,-s)^T$, so $(4,-1)$ is an eigenvector for $\lambda = -4$.

(b) The characteristic polynomial for B is
$\lambda^2 + 2\lambda - 15 = (\lambda + 5)(\lambda - 3)$. The eigenvalues are $\lambda_1 = -5$ and $\lambda_2 = 3$. An eigenvector for λ_1 is $(6, -1)^T$, and an eigenvector for λ_2 is $(2, 1)^T$.

(c) The characteristic polynomial for C is
$\lambda^2 + 2\lambda - 24 = (\lambda + 6)(\lambda - 4)$. Eigenvalues are $\lambda_1 = -6$ and $\lambda_2 = 4$. An eigenvector for λ_1 is $(2, 1)^T$, and an eigenvector for λ_2 is $(8, -1)^T$.

3. Repeat exercise (2), but with the following matrices:

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

(b) $B = \begin{bmatrix} 4 & 5 & -6 \\ 1 & 2 & 0 \\ 0 & 1 & 2 \end{bmatrix}$

(c) $C = \begin{bmatrix} 3 & 1 & 0 \\ 19 & 1 & 1 \\ -20 & 0 & 1 \end{bmatrix}$

Solution:

(a) The characteristic polynomial for A is
$(3 - \lambda)(\lambda^2 - 3\lambda - 4) = (\lambda + 1)(\lambda - 3)(\lambda - 4)$; the eigenvalues are $\lambda_1 = -1$, $\lambda_2 = 3$, and $\lambda_3 = 4$. An eigenvector for $\lambda_1 = -1$ is $(0, 2, 1)^T$; an eigenvector for $\lambda_2 = 3$ is $(1, 0, 0)^T$; an eigenvector for $\lambda_3 = 4$ is $(0, 3, -1)^T$.

(b) The characteristic polynomial for B is
$-\lambda^3 + 8\lambda^2 - 15\lambda = (-\lambda)(\lambda - 3)(\lambda - 5)$, with zeros $\lambda_1 = 0$, $\lambda_2 = 3$, and $\lambda_3 = 5$. An eigenvector for $\lambda_1 = 0$ is $(4, -2, 1)^T$; an eigenvector for $\lambda_2 = 3$ is $(1, 1, 1^T)$; an eigenvector for $\lambda_3 = 5$ is $(9, 3, 1)^T$.

(c) The characteristic polynomial for C is
$-\lambda^3 + 5\lambda^2 + 12\lambda - 36 = (\lambda - 2)(\lambda + 3)(\lambda - 6)$, with zeros $\lambda_1 = 2$, $\lambda_2 = -3$, and $\lambda_3 = 6$. An eigenvector for $\lambda_1 = 2$ is $(1, -1, -20)^T$; an eigenvector for $\lambda_2 = -3$ is $(1, -6, 5)^T$; an eigenvector for $\lambda_3 = 6$ is $(1, 3, -4)^T$

A.5 Solutions to Exercises for Chapter 5

1. Given $z = 2 + 3i$, $w = 1 - 4i$, and $y = 3 - i$, we are to find the following:

   (a) $z + w, z + y$, and $w + y$

   (b) $z - w, w - z, z - y, y - z, w - y$, and $y - w$

   (c) zw, zy, and wy

   Solution:

   (a) $z + w = 3 - i,\ z + y = 5 + 2i$, and $w + y = 4 - 5i$.

   (b) $z - w = 1 + 7i,\ w - z = -1 - 7i,\ z - y = -1 + 4i,\ y - z = 1 - 4i,\ w - y = -2 - 3i$, and $y - w = 2 + 3i$.

   (c) $zw = 14 - 5i,\ zy = 9 + 7i$, and $wy = -1 - 13i$.

2. Given z, w, and y be as in exercise 1, we have $|z| = \sqrt{4+9} = \sqrt{13},\ |w|\sqrt{1+16} = \sqrt{17}$, and $|y| = \sqrt{9+1} = \sqrt{10}$.

4. Let $z = 3 - 4i$. Find $\overline{z}$, and verify that $|z| = \sqrt{z\overline{z}}$.

   Solution: $\overline{z} = 3 + 4i$, and $|z| = 5 = \sqrt{(3-4i)(3+4i)}$.

5. Let $z = 2 + 2i$ and $w = \sqrt{3} - i$. Find z/w and w/z.

   Solution:
   $z/w = \dfrac{2+2i}{\sqrt{3}-i} = \dfrac{(2+2i)(\sqrt{3}+i)}{(\sqrt{3}-i)(\sqrt{3}+i)} = \dfrac{(\sqrt{3}-1)+(\sqrt{3}+1)i}{2}$,
   and $w/z = \dfrac{\sqrt{3}-i}{2+2i} = \dfrac{(\sqrt{3}-i)(2-2i)}{(2+2i)(2-2i)} = \dfrac{(\sqrt{3}-1)-(\sqrt{3}+1)i}{4}$.

6. Let $z = -3 + i$. Find z^{-1}.

   Solution: $z^{-}1 = \dfrac{1}{z} = \dfrac{\overline{z}}{z\overline{z}} = \dfrac{-3-i}{10}$.

7. Let $z = \sqrt{2} + i\sqrt{2}$, $w = 2\sqrt{3} - 2i$, and $y = -\sqrt{3} + 3i$. Find the modulus and argument of each.

   Solution: $|z| = \sqrt{2+2} = 2,\ \arg(z) = \dfrac{\pi}{4}$,
   $|w| = \sqrt{16} = 4,\ \arg(w) = \dfrac{-\pi}{6}$,
   $|y| = \sqrt{12} = 2\sqrt{3}$, and $\arg(y) = \dfrac{2\pi}{3}$.

8. Express the following in polar form: $z = 1/\sqrt{2} + i/\sqrt{2}$, $w = -1/\sqrt{2} + i/\sqrt{2}$, and $y = 1 + i\sqrt{3}$.

   Solution: $z = \cos\frac{\pi}{4} + i\sin\frac{\pi}{4}$, $w = \cos\frac{3\pi}{4} + i\sin\frac{3\pi}{4}$, $y = 2(\cos\frac{\pi}{3} + i\sin\frac{\pi}{3})$.

9. Let z, w, and y be as in exercise 7. Use their polar forms to find the products zw, zy, and wy.

   Solution: Applying the results of exercise 7, we have $z = 2\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right)$, $w = 4\left(\cos\frac{-\pi}{6} + i\sin\frac{-\pi}{6}\right)$, and $y = 2\sqrt{3}\left(\cos\frac{2\pi}{3} + i\sin\frac{2\pi}{3}\right)$. It follows that $zw = 8\left(\cos\frac{\pi}{12} + i\sin\frac{\pi}{12}\right)$, $zy = 4\sqrt{3}\left(\cos\frac{11\pi}{12} + i\sin\frac{11\pi}{12}\right)$, and $wy = 8\sqrt{3}\left(\cos\frac{\pi}{2} + i\sin\frac{\pi}{2}\right)$.

10. Let z and w be as in exercise 7. Use their polar forms to find the quotients z/w and w/z.

    Solution: From the preceding problem, we have $z = 2(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4})$ and $w = 4\left(\cos\frac{-\pi}{6} + i\sin\frac{-\pi}{6}\right)$. So $z/w = \frac{1}{2}\left(\cos\frac{5\pi}{12} + i\sin\frac{5\pi}{12}\right)$ and $w/z = 2\left(\cos\frac{-5\pi}{12} + i\sin\frac{-5\pi}{12}\right)$.

11. Factor completely as products of linear factors:

    (a) $p(x) = x^2 + 9 = (x + 3i)(x - 3i)$.

    (b) $q(x) = x^2 + 6 = (x + i\sqrt{6})(x - i\sqrt{6})$.

    (c) $r(x) = x^2 - 4x + 13 = (x - (2 + 3i))(x - (2 - 3i))$.

    (d) $s(x) = x^2 + x + 1 = \left(x + \left(\frac{1 - i\sqrt{3}}{2}\right)\right)\left(x + \left(\frac{1 + i\sqrt{3}}{2}\right)\right)$.

12. Find the exponential form for each of z, w, and y from exercise 7. Use their exponential forms to find their products zw, zy, and wy.

    Solution: $z = 2e^{\frac{i\pi}{4}}$, $w = 4e^{\frac{-i\pi}{6}}$, and $y = 2\sqrt{3}e^{\frac{2i\pi}{3}}$. So $zw = 8e^{\frac{i\pi}{12}}$, $zy = 4\sqrt{3}e^{\frac{11i\pi}{12}}$, and $wy = 8\sqrt{3}e^{\frac{i\pi}{2}}$.

13. Use the exponential forms from the preceding problem to find $\overline{z}$, $\overline{w}$, and $\overline{y}$.

    Solution: $\overline{z} = 2e^{\frac{-i\pi}{4}}$, $\overline{w} = 4e^{\frac{i\pi}{6}}$, and $\overline{y} = 2\sqrt{3}e^{\frac{-2i\pi}{3}}$.

14. Again using the exponential forms from problem 12, find z/w, w/z, z/y, y/z, w/y, and y/w.

    Solution: $z/w = \frac{1}{2}e^{i\left(\frac{\pi}{4}+\frac{\pi}{6}\right)} = \frac{1}{2}e^{\frac{i5\pi}{12}}$, $w/z = 2e^{\frac{-i5\pi}{12}}$, $z/y = \frac{1}{\sqrt{3}}e^{-i\frac{5\pi}{12}}$, $y/z = \sqrt{3}e^{i\frac{5\pi}{12}}$, $w/y = \frac{2}{\sqrt{3}}e^{-i\frac{5\pi}{6}}$, and $y/w = \frac{\sqrt{3}}{2}e^{i\frac{5\pi}{6}}$.

15. Find all roots of the following equations:

    (a) $z^4 = 1$
    (b) $z^3 = 8$

    Solution:

    (a) The roots are $z = \pm 1$ and $z = \pm i$.
    (b) The roots are $z = 2$, $z = 2e^{i\frac{2\pi}{3}}$, and $z = 2e^{i\frac{4\pi}{3}}$.

16. Verify that the vectors found in the final example in this chapter are in fact eigenvectors for the given matrix.

    Solution: The matrix was $A = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$, and the claimed eigenvectors are $\mathbf{x}_1 = (-i, 1)^T$ and $\mathbf{x}_2 = (i, 1)^T$. We check:

    $$A\mathbf{x}_1 = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} -i \\ 1 \end{bmatrix} = \begin{bmatrix} 2-i \\ 1+2i \end{bmatrix} = (1+2i)\mathbf{x}_1,$$

    and

    $$A\mathbf{x}_2 = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} i \\ 1 \end{bmatrix} = \begin{bmatrix} 2+i \\ 1-2i \end{bmatrix} = (1-2i)\mathbf{x}_2.$$

17. Let $A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} i \\ 1 \end{bmatrix}$. Verify that $A\mathbf{x} = (1-i)\mathbf{x}$.

    Solution:
    $A\mathbf{x} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} i \\ 1 \end{bmatrix} = \begin{bmatrix} 1+i \\ 1-i \end{bmatrix} = \begin{bmatrix} (1-i)i \\ (1-i)1 \end{bmatrix} = (1-i) \begin{bmatrix} i \\ 1 \end{bmatrix}.$

18. Let $A = \begin{bmatrix} 0 & -2 \\ 4 & 4 \end{bmatrix}$. Find all eigenvalues and associated eigenspaces of A.

Solution: The characteristic equation is $\lambda^2 - 4\lambda + 8 = 0$, with roots $\lambda = 2 \pm 2i$. Solving $(A - \lambda I)\mathbf{x} = \mathbf{0}$ for $\lambda = 2 + 2i$, we find that $\mathbf{x}$ can be any multiple of $\begin{bmatrix} 2i - 2 \\ 4 \end{bmatrix}$. Repeating this for $\lambda = 2 - 2i$, an associated eigenvector can be any multiple of $\begin{bmatrix} 2i + 2 \\ -4 \end{bmatrix}$.

19. As in the preceding problem, but this time using $A = \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}$.

Solution: The eigenvalues are $\lambda = 0$ and $\lambda = 2$. A representative eigenvector for $\lambda = 0$ is $\mathbf{x} = (i, -1)^T$, while a representative eigenvector for $\lambda = 2$ is $\mathbf{x} = (i, 1)^T$.

Index