Linear Algebra Notes

Elementary Linear Algebra: Applications Version 12th Edition
by Howard Anton, Chris Rorres, and Anton Kaul

Asher Roberts

Contents

Chapter 1

Systems of Linear Equations and Matrices

1.1 Introduction to Systems of Linear Equations

Definition 1.1.1. A linear equation in the n variables $x_1, x_2, \ldots, x_n$ is one that can be expressed in the form

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

where $a_1, a_2, \ldots, a_n$ and b are constants, and the a's are not all zero.
In the special case where $b = 0$, this equation has the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0,$$

which is called a homogeneous linear equation in the variables $x_1, x_2, \ldots, x_n$.

Example 1. The following are linear equations:

$$x + 3y = 7 \qquad x_1 - 2x_2 - 3x_3 + x_4 = 0$$
$$\frac{1}{2}x - y + 3z = -1 \qquad x_1 + x_2 + \cdots + x_n = 1.$$

The following are not linear equations:

$$x + 3y^2 = 4 \qquad 3x + 2y - xy = 5$$
$$\sin x + y = 0 \qquad \sqrt{x_1} + 2x_2 + x_3 = 1.$$

Definition 1.1.2. A finite set of linear equations is called a system of linear equations or, more briefly, a linear system. The variables are called unknowns. A general linear system of m equations in the n unknowns $x_1, x_2, \ldots, x_n$ can be written as

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + \cdots + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + \cdots + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + \cdots + & a_{mn}x_n & = & b_m \end{array}$$

Definition 1.1.3. A solution of a linear system in n unknowns $x_1, x_2, \ldots, x_n$ is a sequence of n numbers $s_1, s_2, \ldots, s_n$ for which the substitution

$$x_1 = s_1, \quad x_2 = s_2, \ldots, \quad x_n = s_n$$

makes each equation a true statement. A solution can be written as

$$(s_1, s_2, \ldots, s_n),$$

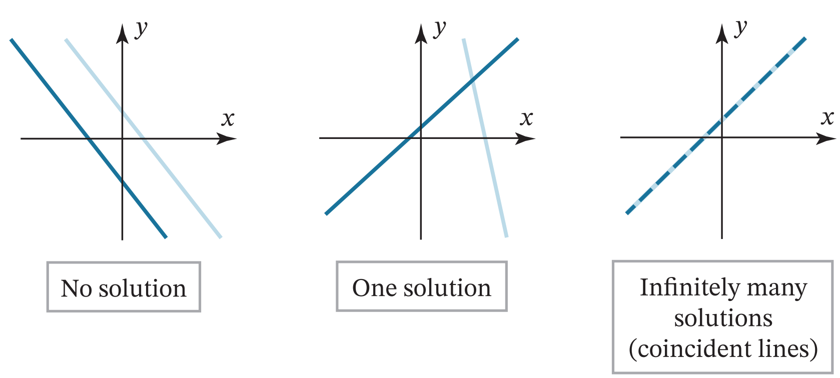
which is called an ordered n-tuple. If $n = 2$, then the n-tuple is called an ordered pair, and if $n = 3$, then it is called an ordered triple.

Remark 1. Consider the linear system

$$a_1x + b_1y = c_1$$
$$a_2x + b_2y = c_2$$

in which the graphs of the equations are lines in the xy-plane. Each solution (x, y) of this system corresponds to a point of intersections of the lines, so there are three possibilities:

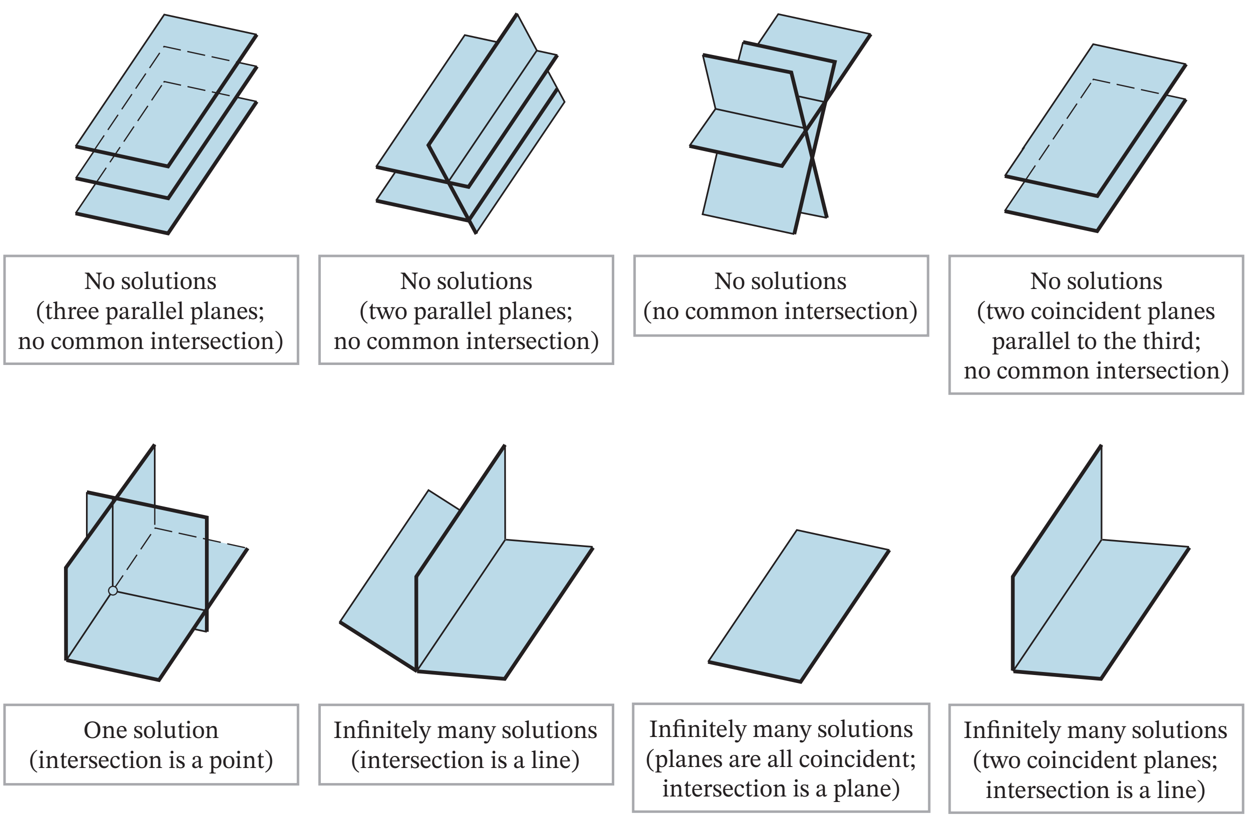
1. The lines may be parallel and distinct, in which case there is no intersection and consequently no solution.
2. The lines may intersect at only one point in which case the system has exactly one solution.
3. The lines may coincide, in which case there are infinitely many points of intersection (the points on the common line) and consequently infinitely many solutions.



In general, we say that a linear system is consistent if it has at least one solution and inconsistent if it has no solutions. Thus, a consistent linear system of two equations in two unknowns has either one solution or infinitely many solutions—there are no other possibilities. The same is true for a linear system of three equations in three unknowns

$$\begin{aligned} a_1x + b_1y + c_1z &= d_1 \\ a_2x + b_2y + c_2z &= d_2 \\ a_3x + b_3y + c_3z &= d_3 \end{aligned}$$

in which the graphs of the equations are planes. The solutions of the system, if any, correspond to points where all three planes intersect, so again we see that there are only three possibilities—no solutions, one solution, or infinitely many solutions.



Theorem 1.1.1. *Every system of linear equations has zero, one, or infinitely many solutions. There are no other possibilities.*

Example 2. Solve the linear system

$$\begin{aligned} x - y &= 1 \\ 2x + y &= 6. \end{aligned}$$

Example 3. Solve the linear system

$$\begin{aligned} x + y &= 4 \\ 3x + 3y &= 6. \end{aligned}$$

Example 4. Solve the linear system

$$\begin{aligned} 4x - 2y &= 1 \\ 16x - 8y &= 4. \end{aligned}$$

Example 5. Solve the linear system

$$\begin{aligned} x - y + 2z &= 5 \\ 2x - 2y + 4z &= 10 \\ 3x - 3y + 6z &= 15. \end{aligned}$$

Definition 1.1.4. The linear system

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

can be abbreviated by writing only the rectangular array of numbers

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix}.$$

This is called the augmented matrix for the system.

Remark 2. The basic method for solving a linear system is to perform algebraic operations on the system that do not alter the solution set and that produce a succession of increasingly simpler systems, until a point is reached where it can be ascertained whether the system is consistent, and if so, what its solutions are. Typically, the algebraic operations are:

1. Multiply an equation through by a nonzero constant.
2. Interchange two equations.
3. Add a constant times one equation to another.

Since the rows (horizontal lines) of an augmented matrix correspond to the equations in the associated system, these three operations correspond to the following operations on the rows of the augmented matrix:

1. Multiply a row through by a nonzero constant.
2. Interchange two rows.
3. Add a constant times one row to another.

These are called elementary row operations on a matrix.

Example 6. Solve the linear system

$$\begin{aligned} x + y + 2z &= 9 \\ 2x + 4y - 3z &= 1 \\ 3x + 6y - 5z &= 0 \end{aligned}$$

by operating on the equations in the system, and by operating on the rows of the augmented matrix for the system.

1.2 Gaussian Elimination

Definition 1.2.1. To be of reduced row echelon form, a matrix must have the following properties:

1. If a row does not consist entirely of zeros, then the first nonzero number in the row is a 1. We call this a leading 1.
2. If there are any rows that consist entirely of zeros, then they are grouped together at the bottom of the matrix.
3. In any two successive rows that do not consist entirely of zeros, the leading 1 in the lower row occurs farther to the right than the leading 1 in the higher row.
4. Each column that contains a leading 1 has zeros everywhere else in that column.

A matrix that has the first three properties is said to be in row echelon form. (Thus, a matrix in reduced row echelon form is of necessity in row echelon form, but not conversely.)

Example 1. The following matrices are in reduced row echelon form.

$$\begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 7 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & -2 & 0 & 1 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

The following matrices are in row echelon form but not reduced row echelon form.

$$\begin{bmatrix} 1 & 4 & -3 & 7 \\ 0 & 1 & 6 & 2 \\ 0 & 0 & 1 & 5 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 2 & 6 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Example 2. With any real numbers substituted for the $*$'s, all matrices of the following types are in row echelon form:

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

All matrices of the following types are in reduced row echelon form:

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

Example 3. Suppose that the augmented matrix for a linear system in the unknowns x_1, x_2, x_3, and x_4 has been reduced by elementary row operations to

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

Solve the system.

Definition 1.2.2. The variables that correspond to the leading 1's in a matrix in reduced row echelon form are called <u>leading variables</u>. The remaining variables are called <u>free variables</u>.

Example 4. In each part, suppose that the augmented matrix for a linear system in the unknowns x, y, and z has been reduced by the elementary row operations to the given reduced row echelon form. Solve the system.

(a) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

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

(c) $\begin{bmatrix} 1 & -5 & 1 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

Definition 1.2.3. If a linear system has infinitely many solutions, then a set of parametric equations from which all solutions can be obtained by assigning numerical values to the parameters is called a general solution of the system.

Theorem 1.2.1. *The following step-by-step elimination procedure can be used to reduce any matrix to reduced row echelon form.*

Step 1. *Locate the leftmost column that does not consist entirely of zeros.*
Step 2. *Interchange the top row with another row, if necessary, to bring a nonzero entry to the top of the column found in Step 1.*
Step 3. *If the entry that is now at the top of the column found in Step 1 is a, multiply the first row by $1/a$ in order to introduce a leading 1.*
Step 4. *Add suitable multiples of the top row to the rows below so that all entries below the leading 1 become zeros.*
Step 5. *Now cover the top row in the matrix and begin again with Step 1 applied to the submatrix that remains. Continue in this way until the entire matrix is in row echelon form.*
Step 6. *Beginning with the last nonzero row and working upward, add suitable multiples of each row to the rows above to introduce zeros above the leading 1's.*

This procedure (or algorithm) is called Gauss-Jordan elimination and consists of two parts, a forward phase in which zeros are introduced below the leading 1's and a backward phase in which zeros are introduced above the leading 1's. If only the forward phase is used, then the procedure produces a row echelon form and is called Gaussian elimination.

Example 5. Solve by Gauss-Jordan elimination.

$$
\begin{aligned}
x_1 + 3x_2 - 2x_3 \phantom{{}- 2x_4} + 2x_5 \phantom{{}- 3x_6} &= 0 \\
2x_1 + 6x_2 - 5x_3 - 2x_4 + 4x_5 - 3x_6 &= -1 \\
5x_3 + 10x_4 + 15x_6 &= 5 \\
2x_1 + 6x_2 + 8x_4 + 4x_5 + 18x_6 &= 6
\end{aligned}
$$

Definition 1.2.4. A system of linear equations is said to be homogeneous if the constant terms are all zero; that is, the system has the form

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

Every homogeneous system of linear equations is consistent because all such systems have $x_1 = 0, x_2 = 0, \ldots, x_n = 0$ as a solution. This solution is called the trivial solution; if there are other solutions, they are called nontrivial solutions.

Remark 1. Because a homogeneous system of linear equations always has the trivial solution, there are only two possibilities for its solutions:

- The system has only the trivial solution.
- The system has infinitely many solutions in addition to the trivial solution.

Example 6. Use Gauss-Jordan elimination to solve the homogeneous system linear system

$$
\begin{aligned}
x_1 + 3x_2 - 2x_3 \qquad\quad + 2x_5 \qquad\quad &= 0 \\
2x_1 + 6x_2 - 5x_3 - 2x_4 + 4x_5 - 3x_6 &= 0 \\
5x_3 + 10x_4 \qquad\quad + 15x_6 &= 0 \\
2x_1 + 6x_2 \qquad\quad + 8x_4 + 4x_5 + 18x_6 &= 0.
\end{aligned}
$$

Theorem 1.2.2 (Free Variable Theorem for Homogeneous Systems)**.** *If a homogeneous linear system has n unknowns, and if the reduced row echelon form of its augmented matrix has r nonzero rows, then the system has $n - r$ free variables.*

Theorem 1.2.3. *A homogeneous linear system with more unknowns than equations has infinitely many solutions.*

Remark 2. For large linear systems that require a computer solution, it is generally more efficient to use Gaussian elimination followed by a technique known as back-substitution to complete the process of solving the system.

Example 7. From the computations in Example 5, a row echelon form of the augmented matrix is

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

Solve the corresponding system of equations.

Example 8. Suppose that the matrices below are augmented matrices for linear systems in the unknowns x_1, x_2, x_3, and x_4. These matrices are all in row echelon form but not reduced row echelon form. Discuss the existence and uniqueness of solutions to the corresponding linear systems.

(a) $\begin{bmatrix} 1 & -3 & 7 & 2 & 5 \\ 0 & 1 & 2 & -4 & 1 \\ 0 & 0 & 1 & 6 & 9 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

(b) $\begin{bmatrix} 1 & -3 & 7 & 2 & 5 \\ 0 & 1 & 2 & -4 & 1 \\ 0 & 0 & 1 & 6 & 9 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & -3 & 7 & 2 & 5 \\ 0 & 1 & 2 & -4 & 1 \\ 0 & 0 & 1 & 6 & 9 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$

Remark 3. There are three facts about row echelon forms and reduced row echelon forms that are important to know:

1. Every matrix has a unique reduced row echelon form.
2. Row echelon forms are not unique.
3. Although row echelon forms are not unique, the reduced row echelon form and all row echelon forms of a matrix A have the same number of zero rows, and the leading 1's always occur in the same positions. Those are called the pivot positions of A. The columns containing leading 1's in a row echelon or reduced row echelon form of A are called the pivot columns of A, and the rows containing the leading 1's are called the pivot rows of A. A *nonzero* entry in a pivot position of A is called a pivot of A.

Example 9. Given that the row echelon form of

$$\begin{bmatrix} 0 & 0 & -2 & 0 & 7 & 12 \\ 2 & 4 & -10 & 6 & 12 & 28 \\ 2 & 4 & -5 & 6 & -5 & -1 \end{bmatrix}$$

is

$$\begin{bmatrix} 1 & 2 & -5 & 3 & 6 & 14 \\ 0 & 0 & 1 & 0 & -\frac{7}{2} & -6 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix}.$$

Determine the pivot positions, pivot columns, pivot rows, and pivots of A.

1.3 Matrices and Matrix Operations

Definition 1.3.1. A matrix is a rectangular array of numbers. The numbers in the array are called the entries.

Example 1. Some examples of matrices are

$$\begin{bmatrix} 1 & 2 \\ 3 & 0 \\ -1 & 4 \end{bmatrix}, \quad \begin{bmatrix} 2 & 1 & 0 & -3 \end{bmatrix}, \quad \begin{bmatrix} e & \pi & -\sqrt{2} \\ 0 & \frac{1}{2} & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \quad [4].$$

Remark 1. The size of a matrix is described in terms of rows (horizontal lines) and columns (vertical lines) it contains. In a size description, the first number always denotes the number of rows, and the second denotes the number of columns. A matrix with only one row is called a row vector (or a row matrix), and a matrix with only one column is called a column vector (or a column matrix).

Remark 2. We will use capital letters to denote matrices and lowercase letters to denote numerical quantities; thus we may write

$$A = \begin{bmatrix} 2 & 1 & 7 \\ 3 & 4 & 2 \end{bmatrix} \quad \text{or} \quad C = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}.$$

When discussing matrices, it is common to refer to numerical quantities as scalars. The entry that occurs in row i and column j of a matrix A will be denoted by a_{ij}. Thus a general $m \times n$ matrix might be written as

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

When a compact notation is desired, the preceding matrix can be written as

$$[a_{ij}]_{m\times n} \quad [a_{ij}].$$

Remark 3. A general $1 \times n$ row vector $\mathbf{a}$ and a general $m \times 1$ column vector $\mathbf{b}$ would be written as

$$\mathbf{a} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

Remark 4. A matrix A with n rows and n columns is called a square matrix of order n, and the shaded entries $a_{11}, a_{22}, \ldots, a_{nn}$ are said to be on the main diagonal of A.

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

Definition 1.3.2. Two matrices are defined to be equal if they have the same size and their corresponding entries are equal.

Example 2. Consider the matrices

$$A = \begin{bmatrix} 2 & 1 \\ 3 & x \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 1 \\ 3 & 5 \end{bmatrix}, \quad C = \begin{bmatrix} 2 & 1 & 0 \\ 3 & 4 & 0 \end{bmatrix}.$$

For what values of x are the matrices equal?

Definition 1.3.3. If A and B are matrices of the same size, then the sum $A+B$ is the matrix obtained by adding the entries of B to the corresponding entries of A, and the difference $A-B$ is the matrix obtained by subtracting the entries of B from the corresponding entries of A. Matrices of different sizes cannot be added or subtracted.

Example 3. Consider the matrices

$$A = \begin{bmatrix} 2 & 1 & 0 & 3 \\ -1 & 0 & 2 & 4 \\ 4 & -2 & 7 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} -4 & 3 & 5 & 1 \\ 2 & 2 & 0 & -1 \\ 3 & 2 & -4 & 5 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}.$$

Find $A+B$, $A+C$, $B+C$, $A-B$, $A-C$, and $B-C$, if possible.

Definition 1.3.4. If A is any matrix and c is any scalar, then the product cA is the matrix obtained by multiplying each entry of the matrix A by c. The matrix cA is said to be a scalar multiple of A.

Example 4. For the matrices

$$A = \begin{bmatrix} 2 & 3 & 4 \\ 1 & 3 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 2 & 7 \\ -1 & 3 & -5 \end{bmatrix}, \quad C = \begin{bmatrix} 9 & -6 & 3 \\ 3 & 0 & 12 \end{bmatrix},$$

find $2A$, $(-1)B$, and $\frac{1}{3}C$.

Definition 1.3.5. If A is an $m \times r$ matrix and B is an $r \times n$ matrix, then the product AB is the $m \times n$ matrix whose entries are determined as follows: To find the entry in row i and column j of AB, single out row i from the matrix A and column j from the matrix B. Multiply the corresponding entries from the row and column together, and then add up the resulting products.

Example 5. Consider the matrices

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

Find AB.

Example 6. Suppose that A, B, and C are matrices with the following sizes:

A	B	C
3×4	4×7	7×3

Determine whether the products AB, AC, BC, BA, CA, and CB are defined.

Remark 5. In general, if $A = [a_{ij}]$ is an $m \times r$ matrix and $B = [b_{ij}]$ is an $r \times n$ matrix, then, as illustrated by the shading in the following display,

$$AB = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1r} \\ a_{21} & a_{22} & \cdots & a_{2r} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{ir} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mr} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1j} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2j} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots & & \vdots \\ b_{r1} & b_{r2} & \cdots & b_{rj} & \cdots & b_{rn} \end{bmatrix}$$

the entry $(AB)_{ij}$ in row i and column j of AB is given by

$$(AB)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \cdots + a_{ir}b_{rj}\,.$$

This is called the row-column rule for matrix multiplication.

Remark 6. A matrix can be subdivided or partitioned into smaller matrices by inserting horizontal and vertical rules between selected rows and columns. The following formulas show how individual column vectors of AB can be obtained by partitioning B into column vectors and how individual row vectors of AB can be obtained by partitioning A into row vectors.

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

(AB computed column by column)

$$AB = \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_m \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 B \\ \mathbf{a}_2 B \\ \vdots \\ \mathbf{a}_m B \end{bmatrix}$$

(AB computed row by row)

Example 7. If A and B are the matrices in Example 5, then compute the second column vector and first row vector of AB.

Definition 1.3.6. If $A_1, A_2, \ldots, A_r$ are matrices of the same size, and if $c_1, c_2, \ldots, c_r$ are scalars, then an expression of the form

$$c_1A_1 + c_2A_2 + \cdots + c_rA_r$$

is called a <u>linear combination</u> of $A_1, A_2, \ldots, A_r$ with <u>coefficients</u> $c_1, c_2, \ldots, c_r$.

Theorem 1.3.1. *If A is an $m \times n$ matrix, and if $\mathbf{x}$ is an $n \times 1$ column vector, then the product $A\mathbf{x}$ can be expressed as a linear combination of the column vectors of A in which the coefficients are the entries of $\mathbf{x}$.*

Proof. Let

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Then

$$A\mathbf{x} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix}$$

□

Example 8. Write the matrix product

$$\begin{bmatrix} -1 & 3 & 2 \\ 1 & 2 & -3 \\ 2 & 1 & -2 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ -9 \\ -3 \end{bmatrix}$$

as a linear combination of column vectors.

Example 9. If A and B are the matrices in Example 5, then write the column vectors of the matrix product AB as linear combinations of column vectors.

Remark 7. Suppose that an $m \times r$ matrix A is partitioned into its r column vectors $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_r$ (each of size $m \times 1$) and an $r \times n$ matrix B is partitioned into its r row vectors $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_r$ (each of size $1 \times n$). Then

$$AB = \mathbf{c}_1\mathbf{r}_1 + \mathbf{c}_2\mathbf{r}_2 + \cdots + \mathbf{c}_r\mathbf{r}_r,$$

and this equation is called the column-row expansion of AB.

Example 10. Find the column-row expansion of the product

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

Remark 8. Consider a system of m linear equations in n unknowns:

$$\begin{array}{ccccccc} a_{11}x_1 & + & a_{12}x_2 & + \cdots + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + \cdots + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + \cdots + & a_{mn}x_n & = & b_m \end{array}$$

We can replace the m equations in this system by the matrix equation

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

If we designate these matrices by A, $\mathbf{x}$, and $\mathbf{b}$, respectively, then we can replace the original system of m equations in n unknowns by the single matrix equation

$$A\mathbf{x} = \mathbf{b}.$$

The matrix A in this equation is called the coefficient matrix of the system. The augmented matrix for the system is obtained by adjoining $\mathbf{b}$ to A as the last column; thus the augmented matrix is

$$[A \mid \mathbf{b}] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right].$$

Definition 1.3.7. If A is any $m \times n$ matrix, then the transpose of A, denoted by A^T, is defined to be the $n \times m$ matrix that results by interchanging the rows and columns of A; that is, the first column of A^T is the first row of A, the second column of A^T is the second row of A, and so forth.

Example 11. Find the transposes of the following matrices:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix}, \quad B = \begin{bmatrix} 2 & 3 \\ 1 & 4 \\ 5 & 6 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 3 & 5 \end{bmatrix}, \quad D = [4].$$

Remark 9. Not only are the columns of A^T the rows of A, but the rows of A^T are the columns of A. Thus the entry in row i and column j of A^T is the entry in row j and column i of A; that is,

$$(A^T)_{ij} = (A)_{ji}.$$

In the special case where A is a square matrix, the transpose of A can be obtained by interchanging entries that are symmetrically positioned about the main diagonal.

Definition 1.3.8. If A is a square matrix, then the trace of A, denoted by $\mathrm{tr}(A)$, is defined to be the sum of the entries on the main diagonal of A. The trace of A is undefined if A is not a square matrix.

Example 12. Find the traces of the following matrices:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 2 & 7 & 0 \\ 3 & 5 & -8 & 4 \\ 1 & 2 & 7 & -3 \\ 4 & -2 & 1 & 0 \end{bmatrix}.$$

1.4 Inverses; Algebraic Properties of Matrices

Theorem 1.4.1 (Properties of Matrix Arithmetic). *Assuming that the sizes of the matrices are such that the indicated operations can be performed, the following rules of matrix arithmetic are valid.*

(a) $A + B = B + A$
(b) $A + (B + C) = (A + B) + C$
(c) $A(BC) = (AB)C$
(d) $A(B + C) = AB + AC$
(e) $(B + C)A = BA + CA$
(f) $A(B - C) = AB - AC$
(g) $(B - C)A = BA - CA$
(h) $a(B + C) = aB + aC$
(i) $a(B - C) = aB - aC$
(j) $(a + b)C = aC + bC$
(k) $(a - b)C = aC - bC$
(l) $a(bC) = (ab)C$
(m) $a(BC) = (aB)C = B(aC)$

Example 1. Consider

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

Compute $(AB)C$ and $A(BC)$.

Example 2. Consider the matrices

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

Compute AB and BA.

Remark 1. A matrix whose entries are all zero is called a zero matrix. Some examples are

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad [0].$$

We will denote a zero matrix by $\mathit{0}$ unless it is important to specify its size, in which case we will denote the $m \times n$ zero matrix by $\mathit{0}_{m \times n}$.

Theorem 1.4.2 (Properties of Zero Matrices)**.** *If c is a scalar, and if the sizes of the matrices are such that the operations can be performed, then:*

(*a*) $A + \mathit{0} = \mathit{0} + A = A$
(*b*) $A - \mathit{0} = A$
(*c*) $A - A = A + (-A) = \mathit{0}$
(*d*) $0A = \mathit{0}$
(*e*) *If $cA = \mathit{0}$, then $c = 0$ or $A = \mathit{0}$.*

Example 3. Consider the matrices

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 \\ 3 & 4 \end{bmatrix}, \quad C = \begin{bmatrix} 2 & 5 \\ 3 & 4 \end{bmatrix}.$$

Compute AB and AC.

Example 4. Consider the matrices

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 3 & 7 \\ 0 & 0 \end{bmatrix}.$$

Compute AB.

Remark 2. A square matrix with 1's on the main diagonal and zeros elsewhere is called an identity matrix. Some examples are

$$[1], \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

An identity matrix is denoted by the letter I. If it is important to emphasize the size, we will write I_n for the $n \times n$ identity matrix.
If A is any $m \times n$ matrix, then

$$AI_n = A \quad \text{and} \quad I_m A = A.$$

Theorem 1.4.3. *If R is the reduced row echelon form of an $n \times n$ matrix A, then either R has a row of zeros or R is the identity matrix I_n.*

Proof. Suppose that the reduced row echelon form of A is

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}.$$

Either the last row in this matrix consists entirely of zeros or it does not. If not, the matrix contains no zero rows, and consequently each of the n rows has a leading entry of 1. Since these leading 1's occur progressively farther to the right as we move down the matrix, each of these 1's must occur on the main diagonal. Since the other entries in the same column as one of these 1's are zero, R must be I_n. Thus, either R has a row of zeros or $R = I_n$. □

Definition 1.4.1. If A is a square matrix, and if a matrix B of the same size can be found such that $AB = BA = I$, then A is said to be invertible (or nonsingular) and B is called an inverse of A. If no such matrix B can be found, then A is said to be singular.

Example 5. Let

$$A = \begin{bmatrix} 2 & -5 \\ -1 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 3 & 5 \\ 1 & 2 \end{bmatrix}.$$

Are A and B inverses of each other?

Example 6. Consider the matrix

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

Is A singular?

Theorem 1.4.4. *If B and C are both inverses of the matrix A, then $B = C$.*

Proof. Since B is an inverse of A, we have $BA = I$. Multiplying both sides on the right by C gives $(BA)C = IC = C$. But it is also true that $(BA)C = B(AC) = BI = B$, so $C = B$. $\square$

Remark 3. If A is invertible, then its inverse will be denoted by the symbol A^{-1}. Thus,

$$AA^{-1} = I \quad \text{and} \quad A^{-1}A = I.$$

Theorem 1.4.5. *The matrix*

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is invertible if and only if $ad - bc \neq 0$, *in which case the inverse is given by the formula*

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

Example 7. In each part, determine whether the matrix is invertible. If so, find its inverse.

(a) $A = \begin{bmatrix} 6 & 1 \\ 5 & 2 \end{bmatrix}$

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

Example 8. Solve the equations

$$\begin{aligned} u &= ax + by \\ v &= cx + dy \end{aligned}$$

for x and y in terms of u and v.

Theorem 1.4.6. *If A and B are invertible matrices with the same size, then AB is invertible and*

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

Proof. We can establish the invertibility and obtain the stated formula at the same time by showing that

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

But

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

and similarly, $(B^{-1}A^{-1})(AB) = I$. □

Remark 4. A product of any number of invertible matrices is invertible, and the inverse of the product is the product of the inverses in the reverse order.

Example 9. Consider the matrices

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

Compute AB, $(AB)^{-1}$, and $B^{-1}A^{-1}$.

Remark 5. If A is a *square* matrix, then we define the nonnegative integer powers of A to be

$$A^0 = I \quad \text{and} \quad A^n = \underbrace{AA\cdots A}_{n}$$

and if A is invertible, then we define the negative integer powers of A to be

$$A^{-n} = (A^{-1})^n = \underbrace{A^{-1}A^{-1}\cdots A^{-1}}_{n}.$$

Because these definitions parallel those for real numbers, the usual laws of nonnegative exponents hold; for example,

$$A^rA^s = A^{r+s} \quad \text{and} \quad (A^r)^s = A^{rs}.$$

Theorem 1.4.7. *If A is invertible and n is a nonnegative integer, then:*
(a) A^{-1} *is invertible and* $(A^{-1})^{-1} = A$.
(b) A^n *is invertible and* $(A^n)^{-1} = A^{-n} = (A^{-1})^n$.
(c) kA *is invertible for any nonzero scalar k, and* $(kA)^{-1} = k^{-1}A^{-1}$.

Example 10. Let A be the matrix in Example 9. Compute $(A^{-1})^3$ and $(A^3)^{-1}$.

Example 11. Calculate $(A+B)^2$ for matrices A and B.

Definition 1.4.2. If A is a square matrix, say $n \times n$, and if

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m$$

is any polynomial, then we define the $n \times n$ matrix $p(A)$ to be

$$p(A) = a_0 I + a_1 A + a_2 A^2 + \cdots + a_m A^m$$

where I is the $n \times n$ identity matrix; that is, $p(A)$ is obtained by substituting A for x and replacing the constant term a_0 by the matrix $a_0 I$. An expression of this form is called a matrix polynomial in A.

Example 12. Find $p(A)$ for

$$p(x) = x^2 - 2x - 5 \quad \text{and} \quad A = \begin{bmatrix} -1 & 2 \\ 1 & 3 \end{bmatrix}.$$

Remark 6. For any polynomials p_1 and p_2 we have

$$p_1(A)p_2(A) = p_2(A)p_1(A).$$

Theorem 1.4.8. *If the sizes of the matrices are such that the stated operations can be performed, then:*

(*a*) $(A^T)^T = A$
(*b*) $(A+B)^T = A^T + B^T$
(*c*) $(A-B)^T = A^T - B^T$
(*d*) $(kA)^T = kA^T$
(*e*) $(AB)^T = B^T A^T$

Remark 7. The transpose of a product of any number of matrices is the product of the transposes in the reverse order.

Theorem 1.4.9. *If A is an invertible matrix, then A^T is also invertible and*

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

Proof. We can establish the invertibility and obtain the formula at the same time by showing that

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

But from part (e) of Theorem 1.4.8 and the fact that $I^T = I$, we have

$$\begin{aligned} A^T(A^{-1})^T &= (A^{-1}A)^T = I^T = I \\ (A^{-1})^T A^T &= (AA^{-1})^T = I^T = I. \end{aligned}$$ $\square$

Example 13. Compute $(A^T)^{-1}$ and $(A^{-1})^T$ for a general 2×2 invertible matrix.

1.5 Elementary Matrices and a Method for Finding A^{-1}

Definition 1.5.1. Matrices A and B are said to be row equivalent if either (hence each) can be obtained from the other by a sequence of elementary row operations.

Definition 1.5.2. A matrix E is called an elementary matrix if it can be obtained from an identity matrix by performing a *single* elementary row operation.

Example 1. What are the operations that produce the following elementary matrices?

$$\begin{bmatrix} 1 & 0 \\ 0 & -3 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Theorem 1.5.1 (Row Operations by Matrix Multiplication)**.** *If the elementary matrix E results from performing a certain row operation on I_m and if A is an $m \times n$ matrix, then the product EA is the matrix that results when this same row operation is performed on A.*

Example 2. Consider the matrices

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

Compute the product EA.

Example 3. What are the operations that produce the following elementary matrices, and what are the operations that restore them to the identity matrix?

$$\begin{bmatrix} 1 & 0 \\ 0 & 7 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 5 \\ 0 & 1 \end{bmatrix}$$

Theorem 1.5.2. *Every elementary matrix is invertible, and the inverse is also an elementary matrix.*

Proof. If E is an elementary matrix, then E results by performing some row operation on I. Let E_0 be the matrix that results when the inverse of this operation is performed on I. Applying Theorem 1.5.1 and using the fact that inverse row operations cancel the effect of each other, it follows that

$$E_0E = I \quad \text{and} \quad EE_0 = I.$$

Thus, the elementary matrix E_0 is the inverse of E. $\square$

Theorem 1.5.3 (Equivalent Statements)**.** *If A is an $n \times n$ matrix, then the following statements are equivalent, that is, all true or all false.*

(*a*) *A is invertible.*
(*b*) *$A\mathbf{x} = \mathbf{0}$ has only the trivial solution.*
(*c*) *The reduced row echelon form of A is I_n.*
(*d*) *A is expressible as a product of elementary matrices.*

Remark 1 (Inversion Algorithm)*.* To find the inverse of an invertible matrix A, find a sequence of elementary row operations that reduces A to the identity and then perform that same sequence of operations on I_n to obtain A^{-1}.

Example 4. Find the inverse of

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{bmatrix}.$$

Example 5. Consider the matrix

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

Is this matrix invertible?

Example 6. Use Theorem 1.5.3 to determine whether the given homogeneous system has nontrivial solutions.

(a)
$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 0 \\ 2x_1 + 5x_2 + 3x_3 &= 0 \\ x_1 \qquad\quad + 8x_3 &= 0 \end{aligned}$$

(b)
$$\begin{aligned} x_1 + 6x_2 + 4x_3 &= 0 \\ 2x_1 + 4x_2 - x_3 &= 0 \\ -x_1 + 2x_2 + 5x_3 &= 0 \end{aligned}$$

1.6 More on Linear Systems and Invertible Matrices

Theorem 1.6.1. *A system of linear equations has zero, one, or infinitely many solutions. There are no other possibilities.*

Proof. If $A\mathbf{x} = \mathbf{b}$ is a system of linear equations, exactly one of the following is true: (a) the system has no solutions, (b) the system has exactly one solution, or (c) the system has more than one solution. The proof will be complete if we can show that the system has infinitely many solutions in case (c).
Assume that $A\mathbf{x} = \mathbf{b}$ has more than one solution, and let $\mathbf{x}_0 = \mathbf{x}_1 - \mathbf{x}_2$, where $\mathbf{x}_1$ and $\mathbf{x}_2$ are any two distinct solutions. Because $\mathbf{x}_1$ and $\mathbf{x}_2$ are distinct, the matrix $\mathbf{x}_0$ is nonzero; moreover,

$$A\mathbf{x}_0 = A(\mathbf{x}_1 - \mathbf{x}_2) = A\mathbf{x}_1 - A\mathbf{x}_2 = \mathbf{b} - \mathbf{b} = \mathbf{0}.$$

If we now let k be any scalar, then

$$\begin{aligned} A(\mathbf{x}_1 + k\mathbf{x}_0) &= A\mathbf{x}_1 + A(k\mathbf{x}_0) = A\mathbf{x}_1 + k(A\mathbf{x}_0) \\ &= \mathbf{b} + k\mathbf{0} = \mathbf{b} + \mathbf{0} = \mathbf{b}. \end{aligned}$$

But this says that $\mathbf{x}_1 + k\mathbf{x}_0$ is a solution of $A\mathbf{x} = \mathbf{b}$. Since $\mathbf{x}_0$ is nonzero and there are infinitely many choices for k, the system $A\mathbf{x} = \mathbf{b}$ has infinitely many solutions. □

Theorem 1.6.2. *If A is an invertible $n \times n$ matrix, then for each $n \times 1$ matrix $\mathbf{b}$, the system of equations $A\mathbf{x} = \mathbf{b}$ has exactly one solution, name, $\mathbf{x} = A^{-1}\mathbf{b}$.*

Proof. Since $A(A^{-1}\mathbf{b}) = \mathbf{b}$, it follows that $\mathbf{x} = A^{-1}\mathbf{b}$ is a solution of $A\mathbf{x} = \mathbf{b}$. To show that this is the only solution, we will assume that $\mathbf{x}_0$ is an arbitrary solution and then show that $\mathbf{x}_0$ must be the solution $A^{-1}\mathbf{b}$.
If $\mathbf{x}_0$ is any solution of $A\mathbf{x} = \mathbf{b}$, then $A\mathbf{x}_0 = \mathbf{b}$. Multiplying both sides of this equation by A^{-1}, we obtain $\mathbf{x}_0 = A^{-1}\mathbf{b}$. □

Example 1. Solve the system of linear equations

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 5 \\ 2x_1 + 5x_2 + 3x_3 &= 3 \\ x_1 \phantom{{}+5x_2} + 8x_3 &= 17. \end{aligned}$$

Example 2. Solve the systems

(a) $$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 4 \\ 2x_1 + 5x_2 + 3x_3 &= 5 \\ x_1 \phantom{{}+5x_2} + 8x_3 &= 9 \end{aligned}$$

(b) $$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 1 \\ 2x_1 + 5x_2 + 3x_3 &= 6 \\ x_1 \phantom{{}+5x_2} + 8x_3 &= -6. \end{aligned}$$

Theorem 1.6.3. *Let A be a square matrix.*
(*a*) *If B is a square matrix satisfying $BA = I$, then $B = A^{-1}$.*
(*b*) *If B is a square matrix satisfying $AB = I$, then $B = A^{-1}$.*

Proof. (a) Assume that $BA = I$. If we can show that A is invertible, the proof can be completed by multiplying $BA = I$ on both sides by A^{-1} to obtain

$$BAA^{-1} = IA^{-1} \quad \text{or} \quad BI = IA^{-1} \quad \text{or} \quad B = A^{-1}.$$

To show that A is invertible, it suffices to show that the system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution. Let $\mathbf{x}_0$ be any solution of this system. If we multiply both sides of $A\mathbf{x}_0 = \mathbf{0}$ on the left by B, we obtain $BA\mathbf{x}_0 = B\mathbf{0}$ or $I\mathbf{x}_0 = \mathbf{0}$ or $\mathbf{x}_0 = \mathbf{0}$. Thus, the system of equations $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.
(b) Assume that $AB = I$. By part (a), $A = B^{-1}$. By multiplying $A = B^{-1}$ on both sides by B, we obtain

$$BA = BB^{-1} \quad \text{or} \qquad BA = I.$$

The result then follows by (a). □

Theorem 1.6.4 (Equivalent Statements)**.** *If A is an $n \times n$ matrix, then the following are equivalent.*
(*a*) *A is invertible.*
(*b*) *$A\mathbf{x} = \mathbf{0}$ has only the trivial solution.*
(*c*) *The reduced row echelon form of A is I_n.*
(*d*) *A is expressible as a product of elementary matrices.*
(*e*) *$A\mathbf{x} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.*
(*f*) *$A\mathbf{x} = \mathbf{b}$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.*

Theorem 1.6.5. *Let A and B be square matrices of the same size. If AB is invertible, then A and B must also be invertible.*

Proof. We will show first that B is invertible by showing that the homogeneous system $B\mathbf{x} = \mathbf{0}$ has only the trivial solution. If we assume that $\mathbf{x}_0$ is any solution of this system, then

$$(AB)\mathbf{x}_0 = A(B\mathbf{x}_0) = A\mathbf{0} = \mathbf{0}$$

so $\mathbf{x}_0 = \mathbf{0}$ by parts (a) and (b) of Theorem 1.6.4 applied to the invertible matrix AB. But the invertibility of B implies the invertibility of B^{-1}, which in turn implies that

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

is invertible since the left side is a product of invertible matrices. □

Example 3. What conditions must b_1, b_2, and b_3 satisfy in order for the system of equations

$$\begin{aligned} x_1 + x_2 + 2x_3 &= b_1 \\ x_1 \qquad + \ \ x_3 &= b_2 \\ 2x_1 + x_2 + 3x_3 &= b_3 \end{aligned}$$

to be consistent?

Example 4. What conditions must b_1, b_2, and b_3 satisfy in order for the system of equations

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= b_1 \\ 2x_1 + 5x_2 + 3x_3 &= b_2 \\ x_1 + 8x_3 &= b_3 \end{aligned}$$

to be consistent?

1.7 Diagonal, Triangular, and Symmetric Matrices

Definition 1.7.1. A square matrix in which all the entries off the main diagonal are zero is called a diagonal matrix. A general $n \times n$ diagonal matrix D can be written as

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

Remark 1. A diagonal matrix is invertible if and only if all of its diagonal entries are nonzero; in this case the inverse of the diagonal matrix D is

$$D^{-1} = \begin{bmatrix} 1/d_1 & 0 & \cdots & 0 \\ 0 & 1/d_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1/d_n \end{bmatrix}.$$

If k is a positive integer, then

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

Example 1. Compute A^{-1}, A^5, and A^{-5} for

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

Remark 2. To multiply a matrix A on the left by a diagonal matrix D, multiply successive rows of A by the successive diagonal entries of D, and to multiply A on the right by D, multiply successive columns of A by the successive diagonal entries of D.

Definition 1.7.2. A square matrix in which all the entries above the main diagonal are zero is called lower triangular, and a square matrix in which all the entries below the main diagonal are zero is called upper triangular. A matrix that is either upper triangular or lower triangular is called triangular.

Example 2. What are general 4×4 upper and lower triangular matrices?

Remark 3. Observe that diagonal matrices are both upper triangular and lower triangular since they have zeros below and above the main diagonal. Observe also that a *square* matrix in row echelon form is upper triangular since it has zeros below the main diagonal.

Theorem 1.7.1.

(*a*) *The transpose of a lower triangular matrix is upper triangular, and the transpose of an upper triangular matrix is lower triangular.*

(*b*) *The product of lower triangular matrices is lower triangular, and the product of upper triangular matrices is upper triangular.*

(*c*) *A triangular matrix is invertible if and only if its diagonal entries are all nonzero.*

(*d*) *The inverse of an invertible lower triangular matrix is lower triangular, and the inverse of an invertible upper triangular matrix is upper triangular.*

Example 3. Consider the upper triangular matrices

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

What can you say about A^{-1}, B^{-1}, AB, and BA?

Definition 1.7.3. A square matrix is said to be <u>symmetric</u> if $A = A^T$.

Example 4. Which of the following matrices are symmetric?

$$\begin{bmatrix} 7 & -3 \\ -3 & 5 \end{bmatrix} \quad \begin{bmatrix} 1 & 4 & 5 \\ 4 & -3 & 0 \\ 5 & 0 & 7 \end{bmatrix} \quad \begin{bmatrix} d_1 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 \\ 0 & 0 & d_3 & 0 \\ 0 & 0 & 0 & d_4 \end{bmatrix}$$

Theorem 1.7.2. *If A and B are symmetric matrices with the same size, and if k is any scalar, then:*
(*a*) A^T *is symmetric.*
(*b*) $A + B$ *and* $A - B$ *are symmetric.*
(*c*) kA *is symmetric.*

Theorem 1.7.3. *The product of two symmetric matrices is symmetric if and only if the matrices commute.*

Proof. Let A and B be symmetric matrices with the same size. Then

$$(AB)^T = B^T A^T = BA.$$

Thus, $(AB)^T = AB$ if and only if $AB = BA$, that is, if and only if A and B commute. $\square$

Example 5. Which of these products is symmetric?

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

Theorem 1.7.4. *If A is an invertible symmetric matrix, then A^{-1} is symmetric.*

Proof. Assume that A is symmetric and invertible. Then

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

Theorem 1.7.5. *If A is an invertible matrix, then AA^T and A^TA are also invertible.*

Proof. Since A is invertible, so is A^T. Thus AA^T and A^TA are invertible, since they are the products of invertible matrices. □

Remark 4. If A is an $m \times n$ matrix, then A^T is an $n \times m$ matrix, so the products AA^T and A^TA are both square matrices—the matrix AA^T has size $m \times m$, and the matrix A^TA has size $n \times n$. Such products are always symmetric since

$$(AA^T)^T = (A^T)^T A^T = AA^T \quad \text{and} \quad (A^TA)^T = A^T(A^T)^T = A^TA.$$

Example 6. Compute A^TA and AA^T for the 2×3 matrix

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

1.8 Introduction to Linear Transformations

Remark 1. The set of all ordered n-tuples of real numbers is denoted by the symbol R^n. The elements of R^n are called vectors and are denoted in boldface type. Ordered n-tuples can be expressed as

$$(s_1, s_2, \ldots, s_n),$$

called the comma-delimited form of a vector, or as the matrix

$$\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix},$$

called the column-vector form. For each $i = 1, 2, \ldots, n$, let $\mathbf{e}_i$ denote the vector in R^n with a 1 in the ith position and zeros elsewhere. In column form these vectors are

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

We call the vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ the standard basis vectors for R^n. They are termed "basis vectors" because all other vectors in R^n are expressible in exactly one way as a linear combination of them. For example, if

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

then we can express $\mathbf{x}$ as

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

Remark 2. A function is a rule that associates with each element of a set A one and only one element in a set B. If f associates the element b with the element a, then we write

$$b = f(a)$$

and we say that b is the image of a under f and the set B the codomain of f. The subset of the codomain that consists of all images of elements in the domain is called the range of f.

Definition 1.8.1. If T is a function with domain R^n and codomain R^m, then we say that T is a transformation from R^n to R^m or that T maps from R^n to R^m, which we denote by writing

$$T : R^n \to R^m.$$

In the special case where $m = n$, a transformation is sometimes called an operator on R^n.

Remark 3. Suppose that we have the system of linear equations written in matrix notation as

$$\mathbf{w} = A\mathbf{x},$$

which we can view as a transformation that maps a vector $\mathbf{x}$ in R^n into the vector $\mathbf{w}$ in R^m by multiplying x on the left by A. We call this a matrix transformation (or matrix operator in the special case where $m = n$). We denote it by

$$T_A : R^n \to R^m.$$

In situations where specifiying the domain and codomain is not essential, we will write

$$\mathbf{w} = T_A(\mathbf{x}).$$

We call the transformation T_A multiplication by A. On occasion we will find it convenient to express this in the schematic form

$$\mathbf{x} \xrightarrow{T_A} \mathbf{w},$$

which is read "T_A maps $\mathbf{x}$ into $\mathbf{w}$."

Example 1. Find the image of the vector

$$\mathbf{x} = \begin{bmatrix} 1 \\ -3 \\ 0 \\ 2 \end{bmatrix}$$

under the transformation from R^4 to R^3 defined by the equations

$$\begin{aligned} w_1 &= 2x_1 - 3x_2 + x_3 - 5x_4 \\ w_2 &= 4x_1 + x_2 - 2x_3 + x_4 \\ w_3 &= 5x_1 - x_2 + 4x_3 \qquad\quad . \end{aligned}$$

Example 2. Find $T_0(\mathbf{x})$ for an arbitrary vector $\mathbf{x}$ in R^n.

Example 3. Find $T_I(\mathbf{x})$ for an arbitrary vector $\mathbf{x}$ in R^n.

Theorem 1.8.1. *For every matrix A the matrix transformation $T_A : R^n \to R^m$ has the following properties for all vectors $\mathbf{u}$ and $\mathbf{v}$ and for every scalar k:*

(a) $T_A(\mathbf{0}) = \mathbf{0}$
(b) $T_A(k\mathbf{u}) = kT_A(\mathbf{u})$
(c) $T_A(\mathbf{u} + \mathbf{v}) = T_A(\mathbf{u}) + T_A(\mathbf{v})$
(d) $T_A(\mathbf{u} - \mathbf{v}) = T_A(\mathbf{u}) - T_A(\mathbf{v})$

Theorem 1.8.2. *$T : R^n \to R^m$ is a matrix transformation if and only if the following relationships hold for all vectors $\mathbf{u}$ and $\mathbf{v}$ in R^n and for every scalar k:*

(i) $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$
(ii) $T(k\mathbf{u}) = kT(\mathbf{u})$

Proof. If T is a matrix transformation, then properties (i) and (ii) follow respectively from parts (c) and (b) of Theorem 1.8.1.
Conversely, assume that properties (i) and (ii) hold. We must show that there exists an $m \times n$ matrix A such that

$$T(\mathbf{x}) = A\mathbf{x}$$

for every vector $\mathbf{x}$ in R^n. Using the additivity and homogeneity properties of T_A, we get

$$T(k_1\mathbf{u}_1 + k_2\mathbf{u}_2 + \cdots + k_r\mathbf{u}_r) = k_1T(\mathbf{u}_1) + k_2T(\mathbf{u}_2) + \cdots + k_rT(\mathbf{u}_r)$$

for all scalars $k_1, k_2, \ldots, k_r$ and all vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r$ in R^n. Let A be the matrix

$$A = [T(\mathbf{e}_1) \mid T(\mathbf{e}_2) \mid \cdots \mid T(\mathbf{e}_n)]$$

where $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ are the standard basis vectors for R^n. Thus $A\mathbf{x}$ is a linear combination of the columns of A in which the successive cofficients are the entries $x_1, x_2, \ldots, x_n$ of $\mathbf{x}$. That is,

$$\begin{aligned} A\mathbf{x} &= x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + \cdots + x_nT(\mathbf{e}_n) \\ &= T(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \cdots + x_n\mathbf{e}_n) \\ &= T(\mathbf{x}). \end{aligned}$$

□

Remark 4. The additivity and homogeneity properties in Theorem 1.8.2 are called linearity conditions, and a transformation that satisfies these conditions is called a linear transformation. Using this terminology Theorem 1.8.2 can be restated as follows.

Theorem 1.8.3. *Every linear transformation from R^n to R^m is a matrix transformation, and conversely, every matrix transformation from R^n to R^m is a linear transformation.*

Theorem 1.8.4. *If $T_A : R^n \to R^m$ and $T_B : R^n \to R^m$ are matrix transformations, and if $T_A(\mathbf{x}) = T_B(\mathbf{x})$ for every vector $\mathbf{x}$ in R^n, then $A = B$.*

Proof. To say that $T_A(\mathbf{x}) = T_B(\mathbf{x})$ for every vector in R^n is the same as saying that $A\mathbf{x} = B\mathbf{x}$ for every vector $\mathbf{x}$ in R^n. This will be true, in particular, if $\mathbf{x}$ is any of the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ for R^n; that is,

$$A\mathbf{e}_j = B\mathbf{e}_j \quad (j = 1, 2, \dots, n).$$

Since every entry of $\mathbf{e}_j$ is 0 except for the jth, which is 1, it follows that $A\mathbf{e}_j$ is the jth column of A and $B\mathbf{e}_j$ is the jth column of B. Since $A\mathbf{e}_j = B\mathbf{e}_j$, this implies that corresponding columns of A and B are the same, and hence that $A = B$. □

Remark 5. Theorem 1.8.4 tells us that every $m \times n$ matrix A produces exactly one matrix transformation (multiplication by A) and every matrix transformation from R^n to R^m arises from exactly one $m \times n$ matrix; we call that matrix the <u>standard matrix</u> for the transformation.

Remark 6 (Finding the Standard Matrix for a Matrix Transformation)*.*

Step 1. Find the images of the standard basis vectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ for R^n.
Step 2. Construct the matrix that has the images obtained in Step 1 as its successive columns. This matrix is the standard matrix for the transformation.

Example 4. Find the standard matrix A for the linear transformation $T : R^2 \to R^3$ defined by the formula

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 2x_1 + \ \ x_2 \\ x_1 - 3x_2 \\ -x_1 + \ \ x_2 \end{bmatrix}.$$

Example 5. For the linear transformation in Example 4, use the standard matrix A obtained in that example to find

$$T\left(\begin{bmatrix} 1 \\ 4 \end{bmatrix}\right).$$

Example 6. Rewrite the transformation $T(x_1, x_2) = (3x_1 + x_2, 2x_1 - 4x_2)$ in column-vector form and find its standard matrix.

Example 7. Find the standard matrix A for the linear transformation $T : R^2 \to R^2$ for which

$$T\left(\begin{bmatrix} -1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} -5 \\ 5 \end{bmatrix}, \quad T\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right) = \begin{bmatrix} 7 \\ -6 \end{bmatrix}.$$

Remark 7. Some of the most basic operators on R^2 and R^3 are those that map each point into its symmetric image about a fixed line or a fixed plane that contains the origin; these are called reflection operators. Matrix operators on R^2 that move points along arcs of circles centered at the origin are called rotation operators.

Table 1

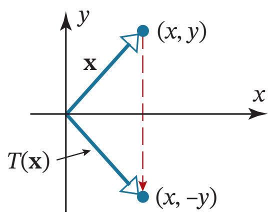
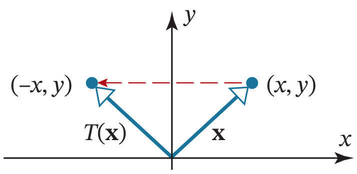
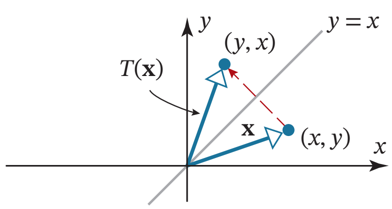
Operator	Illustration	Images of $\mathbf{e}_1$ and $\mathbf{e}_2$	Standard Matrix
Reflection about the x-axis $T(x, y) = (x, -y)$		$T(\mathbf{e}_1) = T(1, 0) = (1, 0)$ $T(\mathbf{e}_2) = T(0, 1) = (0, -1)$	$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$
Reflection about the y-axis $T(x, y) = (-x, y)$		$T(\mathbf{e}_1) = T(1, 0) = (-1, 0)$ $T(\mathbf{e}_2) = T(0, 1) = (0, 1)$	$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$
Reflection about the line $y = x$ $T(x, y) = (y, x)$		$T(\mathbf{e}_1) = T(1, 0) = (0, 1)$ $T(\mathbf{e}_2) = T(0, 1) = (1, 0)$	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

Table 2

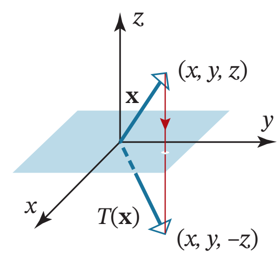
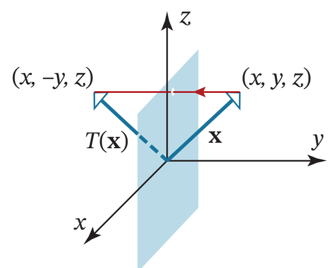
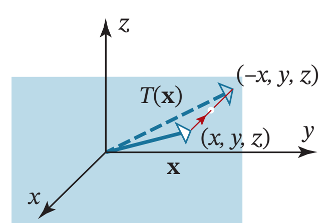
Operator	Illustration	Images of $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$	Standard Matrix
Reflection about the xy-plane $T(x,y,z) = (x,y,-z)$		$T(\mathbf{e}_1) = T(1,0,0) = (1,0,0)$ $T(\mathbf{e}_2) = T(0,1,0) = (0,1,0)$ $T(\mathbf{e}_3) = T(0,0,1) = (0,0,-1)$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$
Reflection about the xz-plane $T(x,y,z) = (x,-y,z)$		$T(\mathbf{e}_1) = T(1,0,0) = (1,0,0)$ $T(\mathbf{e}_2) = T(0,1,0) = (0,-1,0)$ $T(\mathbf{e}_3) = T(0,0,1) = (0,0,1)$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
Reflection about the yz-plane $T(x,y,z) = (-x,y,z)$		$T(\mathbf{e}_1) = T(1,0,0) = (-1,0,0)$ $T(\mathbf{e}_2) = T(0,1,0) = (0,1,0)$ $T(\mathbf{e}_3) = T(0,0,1) = (0,0,1)$	$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Table 3

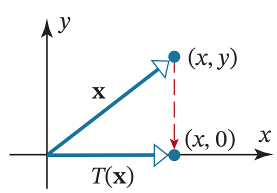
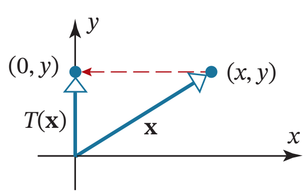
Operator	Illustration	Images of $\mathbf{e}_1$ and $\mathbf{e}_2$	Standard Matrix
Orthogonal projection onto the x-axis $T(x,y) = (x,0)$		$T(\mathbf{e}_1) = T(1,0) = (1,0)$ $T(\mathbf{e}_2) = T(0,1) = (0,0)$	$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$
Orthogonal projection onto the y-axis $T(x,y) = (0,y)$		$T(\mathbf{e}_1) = T(1,0) = (0,0)$ $T(\mathbf{e}_2) = T(0,1) = (0,1)$	$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$

Table 4

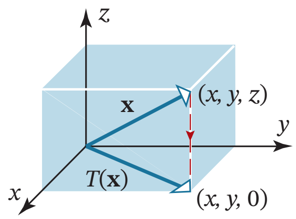
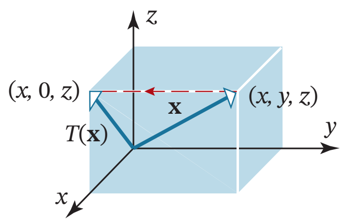
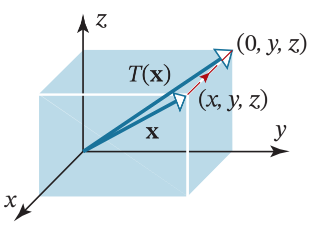
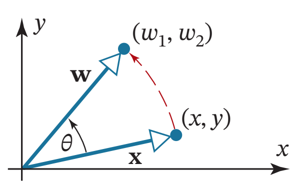
Operator	Illustration	Images of $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$	Standard Matrix
Orthogonal projection onto the xy-plane $T(x, y, z) = (x, y, 0)$		$T(\mathbf{e}_1) = T(1, 0, 0) = (1, 0, 0)$ $T(\mathbf{e}_2) = T(0, 1, 0) = (0, 1, 0)$ $T(\mathbf{e}_3) = T(0, 0, 1) = (0, 0, 0)$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$
Orthogonal projection onto the xz-plane $T(x, y, z) = (x, 0, z)$		$T(\mathbf{e}_1) = T(1, 0, 0) = (1, 0, 0)$ $T(\mathbf{e}_2) = T(0, 1, 0) = (0, 0, 0)$ $T(\mathbf{e}_3) = T(0, 0, 1) = (0, 0, 1)$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
Orthogonal projection onto the yz-plane $T(x, y, z) = (0, y, z)$		$T(\mathbf{e}_1) = T(1, 0, 0) = (0, 0, 0)$ $T(\mathbf{e}_2) = T(0, 1, 0) = (0, 1, 0)$ $T(\mathbf{e}_3) = T(0, 0, 1) = (0, 0, 1)$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Table 5

Operator	Illustration	Images of $\mathbf{e}_1$ and $\mathbf{e}_2$	Standard Matrix
Counterclockwise rotation about the origin through an angle θ		$T(\mathbf{e}_1) = T(1, 0) = (\cos\theta, \sin\theta)$ $T(\mathbf{e}_2) = T(0, 1) = (-\sin\theta, \cos\theta)$	$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$

Example 8. Find the image of $\mathbf{x} = (1, 1)$ under a rotation of $\pi/6$ radians $(= 30°)$ about the origin.

1.9 Compositions of Matrix Transformations

Remark 1. Suppose that T_A is a matrix transformation from R^n to R^k and T_B is a matrix transformation from R^k to R^m. If $\mathbf{x}$ is a vector in R^n, then T_A maps this vector into a vector $T_A(\mathbf{x})$ in R^k, and T_B, in turn, maps that vector into the vector $T_B(T_A(\mathbf{x}))$ in R^m. This process creates a transformation from R^n to R^m that we call the composition of T_B with T_A and denote by the symbol

$$T_B \circ T_A,$$

which is read "T_B circle T_A." The transformation T_A in the formula is performed first; that is,

$$(T_B \circ T_A)(\mathbf{x}) = T_B(T_A(\mathbf{x})).$$

Theorem 1.9.1. *If $T_A : R^n \to R^k$ and $T_B : R^k \to R^m$ are matrix transformations, then $T_B \circ T_A$ is also a matrix transformation and*

$$T_B \circ T_A = T_{BA}.$$

Proof. First we will show that $T_B \circ T_A$ is a linear transformation, thereby establishing that it is a matrix transformation. Then we will show that the standard matrix for this transformation is BA to complete the proof.

To prove that $T_B \circ T_A$ is linear we must show that it has the required additivity and homogeneity properties. For this purpose, let $\mathbf{x}$ and $\mathbf{y}$ be vectors in R^n and observe that

$$\begin{aligned}(T_B \circ T_A)(\mathbf{x}+\mathbf{y}) &= T_B(T_A(\mathbf{x}+\mathbf{y}))\\ &= T_B(T_A(\mathbf{x}) + T_A(\mathbf{y}))\\ &= T_B(T_A(\mathbf{x})) + T_B(T_A(\mathbf{y}))\\ &= (T_B \circ T_A)(\mathbf{x}) + (T_B \circ T_A)(\mathbf{y}),\end{aligned}$$

which proves additivity. Moreover,

$$\begin{aligned}(T_B \circ T_A)(k\mathbf{x}) &= T_B(T_A(k\mathbf{x}))\\ &= T_B(k(T_A(\mathbf{x}))\\ &= kT_B(T_A(\mathbf{x}))\\ &= k(T_B \circ T_A)(\mathbf{x}),\end{aligned}$$

which proves homogeneity and establishes that $T_B \circ T_A$ is a matrix transformation. Thus, there is an $m \times n$ matrix C such that

$$T_B \circ T_A = T_C.$$

To find the appropriate matrix C that satisfies this equation, observe that

$$T_C(\mathbf{x}) = (T_B \circ T_A)(\mathbf{x}) = T_B(T_A(\mathbf{x})) = T_B(A\mathbf{x}) = B(A\mathbf{x}) = (BA)\mathbf{x} = T_{BA}(\mathbf{x}).$$

It now follows that $C = BA$. $\square$

Example 1. Let $T_1 : R^3 \to R^2$ and $T_2 : R^2 \to R^3$ be the linear transformations given by

$$T_1(x, y, z) = (x + 2y, x + 2z - y)$$

and

$$T_2(x, y) = (3x + y, x, x - 2y).$$

Find the standard matrices for $T_2 \circ T_1$ and $T_1 \circ T_2$.

Example 2. Let $T_A : R^2 \to R^2$ be the reflection about the line $y = x$, and let $T_B : R^2 \to R^2$ be the orthogonal projection onto the $y - axis$. What are the standard matrices for $T_A \circ T_B$ and $T_B \circ T_A$?

Example 3. Let $T_{A_1} : R^2 \to R^2$ and $T_{A_2} : R^2 \to R^2$ be the matrix operators that rotate vectors about the origin through the angles θ_1 and θ_2, respectively. Verify that $T_{A_1} \circ T_{A_2} = T_{A_2} \circ T_{A_1}$.

Example 4. Let $T_1 : R^2 \to R^2$ be the reflection about the y-axis, and let $T_2 : R^2 \to R^2$ be the reflection about the x-axis. Verify that $T_1 \circ T_2 = T_2 \circ T_1$.

Remark 2. Compositions can be defined for any finite succession of matrix transformations whose domains and ranges have the appropriate dimensions. For example, consider the matrix transformations

$$T_A : R^n \to R^k, \quad T_B : R^k \to R^l, \quad T_C : R^l \to R^m.$$

We define the composition $(T_C \circ T_B \circ T_A) : R^n \to R^m$ by

$$(T_C \circ T_B \circ T_A)(\mathbf{x}) = T_C(T_B(T_A(\mathbf{x}))).$$

As above, it can be shown that this is a matrix transformation whose standard matrix is CBA and that

$$T_C \circ T_B \circ T_A = T_{CBA}.$$

Example 5. Find the image of a vector

$$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$$

under the matrix transformation that first rotates $\mathbf{x}$ about the origin through an angle $\pi/6$, then reflects the resulting vector about the line $y = x$, and then projects that vector orthogonally onto the y-axis.

Remark 3. If $T_A : R^n \to R^n$ is a matrix operator whose standard matrix A is invertible, then we say that T_A is invertible, and we define the inverse of T_A as

$$T_A^{-1} = T_{A^{-1}},$$

or restated in words, the *inverse of multiplication by A is multiplication by the inverse of A.* Thus, by definition, the standard matrix for T_A^{-1} is A^{-1}, from which it follows that

$$T_A^{-1} \circ T_A = T_{A^{-1}} \circ T_A = T_{A^{-1}A} = T_I.$$

It follows from this that for any vector $\mathbf{x}$ in R^n

$$(T_A^{-1} \circ T_A)(\mathbf{x}) = T_I(\mathbf{x}) = I\mathbf{x} = \mathbf{x}$$

and similarly that $(T_A \circ T_A^{-1})(\mathbf{x}) = \mathbf{x}$. Thus, when T_A and T_A^{-1} are composed in either order they cancel out the effect of one another.

Example 6. Let $T : R^2 \to R^2$ be the operator that rotates each vector in R^2 through the angle θ. Find the standard matrix for T^{-1}.

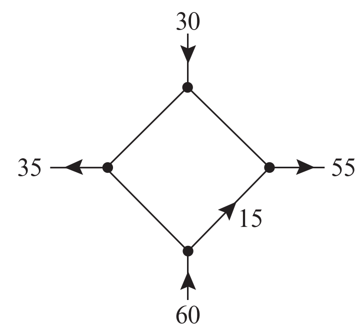
Example 7. Consider the operator $T : R^2 \to R^2$ defined by the equations

$$\begin{aligned} w_1 &= 2x_1 + x_2 \\ w_2 &= 3x_1 + 4x_2. \end{aligned}$$

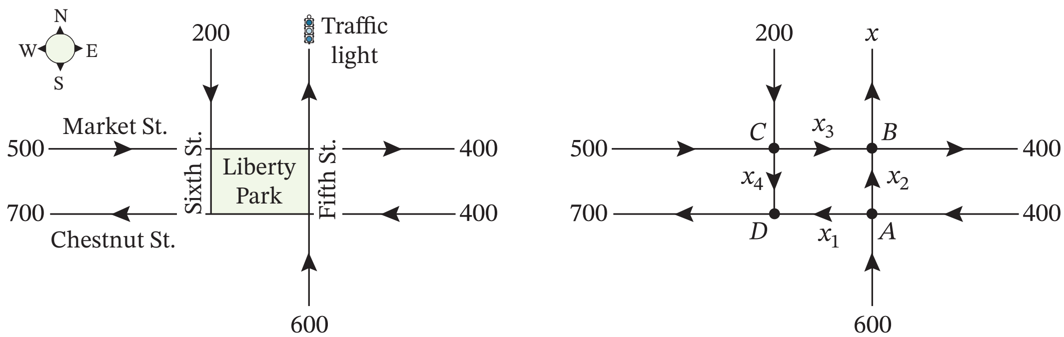
Find $T^{-1}(w_1, w_2)$.

1.10 Applications of Linear Systems

Example 1. The figure shows a network with four nodes in which the flow rate and direction of flow in certain branches are known. Find the flow rates and directions of flow in the remaining branches.



Example 2. The network in the figure shows a proposed plan for the traffic flow around a new park that will house the Liberty Bell in Philadelphia, Pennsylvania. The plan calls for a computerized traffic light at the north exit on Fifth Street, and the diagram indicates the average number of vehicles per hour that are expected to flow in and out of the streets that border the complex. All streets are one-way.



(a) How many vehicles per hour should the traffic light let through to ensure that the average number of vehicles per hour flowing into the complex is the same as the average number of vehicles flowing out?

(b) Assuming that the traffic light has been set to balance the total flow in and out of the complex, what can you say about the average number of vehicles per hour that will flow along the streets that border the complex?

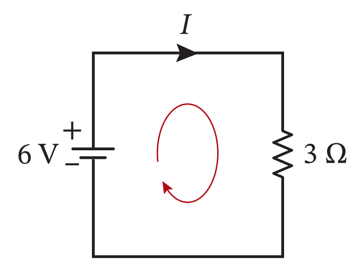
Theorem 1.10.1 (Ohm's Law)**.** *If a current of I amperes passes through a resistor with a resistance of R ohms, then there is a resulting drop of E volts in electrical potential that is the product of the current and resistance; that is,*

$$E = IR.$$

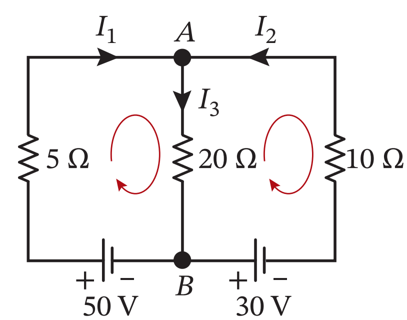
Theorem 1.10.2 (Kirchhoff's Current Law)**.** *The sum of the currents flowing into any node is equal to the sum of the currents flowing out.*

Theorem 1.10.3 (Kirchhoff's Voltage Law)**.** *In one traversal of any closed loop, the sum of the voltage rises equals the sum of the voltage drops.*

Example 3. Determine the current I in the circuit shown in the figure.



Example 4. Determine the currents I_1, I_2, and I_3 in the circuit shown in the figure.



Example 5. Balance the chemical equation

$$\begin{array}{ccccccc} \mathrm{HCl} & + & \mathrm{Na_3PO_4} & \longrightarrow & \mathrm{H_3PO_4} & + & \mathrm{NaCl} \\ \textbf{[hydrochloric acid]} & + & \textbf{[sodium phosphate]} & \longrightarrow & \textbf{[phosphoric acid]} & + & \textbf{[sodium chloride]}. \end{array}$$

Theorem 1.10.4 (Polynomial Interpolation)**.** *Given any n points in the xy-plane that have distinct x-coordinates, there is a unique polynomial of degree $n-1$ or less whose graph passes through those points.*

Example 6. Find a cubic polynomial whose graph passes through the points

$$(1, 3), \quad (2, -2), \quad (3, -5), \quad (4, 0).$$

Example 7. Use polynomial interpolation to approximate the integral

$$\int_0^1 \sin\left(\frac{\pi x^2}{2}\right) dx.$$

1.11 Leontief Input-Output Models

Remark 1. Suppose the open sector of an economy (the sector that does not produce outputs) wants the economy to supply it with goods, products, and utilities with monetary values. The column vector $\mathbf{d}$ that has these numbers as successive components is called the outside demand vector. Since the product-producing sectors consume some of their own output, the monetary value of their output must cover their own needs plus the outside demand. The column vector $\mathbf{x}$ that has these monetary value numbers as successive components is called the production vector for the economy.

By multiplying $\mathbf{x}$ by the consumption matrix C for the economy, whose columns are the inputs required for each output, we obtain $C\mathbf{x}$, the portion of the production vector $\mathbf{x}$ that will be consumed by the productive sectors. The vector $C\mathbf{x}$ is called the intermediate demand vector for the economy. Once the intermediate demand is met, the portion of the production that is left to satisfy the outside demand is $\mathbf{x} - C\mathbf{x}$. Thus $\mathbf{x}$ must satisfy the equation

$$\mathbf{x} - C\mathbf{x} = \mathbf{d},$$

which we will find convenient to rewrite as

$$(I - C)\mathbf{x} = \mathbf{d}.$$

The matrix $I - C$ is called the Leontief matrix and $(I - C)\mathbf{x} = \mathbf{d}$ is called the Leontief equation.

Example 1. Consider the economy described in the table.

Input Required per Dollar Output

Provider		Manufacturing	Agriculture	Utilities
	Manufacturing	\$ 0.50	\$ 0.10	\$ 0.10
	Agriculture	\$ 0.20	\$ 0.50	\$ 0.30
	Utilities	\$ 0.10	\$ 0.30	\$ 0.40

Suppose that the open sector has a demand for \$7900 worth of manufacturing products, \$3950 worth of agricultural products and \$1975 worth of utilities.

(a) Can the economy meet this demand?
(b) If so, find a production vector **x** that will meet it exactly.

Remark 2. In the case where an open economy has n product-producing sectors, the consumption matrix, production vector, and outside demand vector have the form

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{bmatrix}$$

where all entries are nonnegative and

c_{ij} = the monetary value of the output of the ith sector that is needed by the jth sector to produce one unit of output

x_i = the monetary value of the output of the ith sector

d_i = the monetary value of the output of the ith sector that is required to meet the demand of the open sector.

Theorem 1.11.1. *If C is the consumption matrix for an open economy, and if all of the column sums are less than 1, then the matrix $I - C$ is invertible, the entries $(I - C)^{-1}$ are nonnegative, and the economy is productive.*

Example 2. The column sums of the consumption matrix C in Example 1 are less than 1, so $(I - C)^{-1}$ exists and has nonnegative entries. Use a calculating utility to confirm this, and use this inverse to solve the linear system in Example 1.

Chapter 2

Determinants

2.1 Determinants by Cofactor Expansion

Definition 2.1.1. If A is a square matrix, then the minor of entry a_{ij} is denoted by M_{ij} and is defined to be the determinant of the submatrix that remains after the ith row and jth column are deleted from A. The number $(-1)^{i+j}M_{ij}$ is denoted by C_{ij} and is called the cofactor of entry a_{ij}.

Example 1. Let

$$A = \begin{bmatrix} 3 & 1 & -4 \\ 2 & 5 & 6 \\ 1 & 4 & 8 \end{bmatrix}.$$

Find the minors and cofactors of entries a_{11} and a_{32}.

Example 2. Express $\det(A)$ for a 2×2 matrix A in terms of cofactors and entries that all come from the same row or same column of A.

Theorem 2.1.1. *If A is an $n \times n$ matrix, then regardless of which row or column of A is chosen, the number obtained by multiplying the entries in that row or column by the corresponding cofactors and adding the resulting products is always the same.*

Definition 2.1.2. If A is an $n \times n$ matrix, then the number obtained by multiplying the entries in any row or column of A by the corresponding cofactors and adding the resulting products is called the determinant of A, and the sums themselves are called cofactor expansions of A. That is,

$$\det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}$$

and

$$\det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}.$$

Example 3. Find the determinant of the matrix

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

by cofactor expansion along the first row.

Example 4. Let A be the matrix in Example 3, and evaluate $\det(A)$ by cofactor expansion along the first column of A.

Example 5. Find the determinant of the matrix

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

Example 6. Find the determinant of a 4×4 lower triangular matrix.

Theorem 2.1.2. *If A is an $n \times n$ triangular matrix (upper triangular, lower triangular, or diagonal), then* $\det(A)$ *is the product of the entries on the main diagonal of the matrix; that is,* $\det(A) = a_{11}a_{22}\cdots a_{nn}$.

Example 7. Evaluate

$$\begin{vmatrix} 3 & 1 \\ 4 & -2 \end{vmatrix} \quad \text{and} \quad \begin{vmatrix} 1 & 2 & 3 \\ -4 & 5 & 6 \\ 7 & -8 & 9 \end{vmatrix}.$$

2.2 Evaluating Determinants by Row Reduction

Theorem 2.2.1. *Let A be a square matrix. If A has a row of zeros or a column of zeros, then* $\det(A) = 0$.

Proof. Since the determinant of A can be found by a cofactor expansion along any row or column, we can use the row or column of zeros. Thus, if we let $C_1, C_2, \ldots, C_n$ denote the cofactors of A along that row or column, then it follows that

$$\det(A) = 0 \cdot C_1 + 0 \cdot C_2 + \cdots + 0 \cdot C_n = 0. \qquad \square$$

Theorem 2.2.2. *Let A be a square matrix. Then* $\det(A) = \det(A^T)$.

Proof. Since transposing a matrix changes its columns to rows and rows to columns, the cofactor expansion of A along any row is the same as the cofactor expansion of A^T along the corresponding column. Thus, both have the same determinant. $\square$

Theorem 2.2.3. *Let A be an $n \times n$ matrix.*

(*a*) *If B is the matrix that results when a single row or single column of A is multiplied by a scalar k, then* $\det(B) = k\det(A)$.

(*b*) *If B is the matrix that results when two rows or two columns of A are interchanged, then* $\det(B) = -\det(A)$.

(*c*) *If B is the matrix that results when a multiple of one row of A is added to another or when a multiple of one column is added to another, then* $\det(B) = \det(A)$.

Theorem 2.2.4. *Let E be an $n \times n$ elementary matrix.*

(*a*) *If E results from multiplying a row of I_n by a nonzero number k, then* $\det(E) = k$.

(*b*) *If E results from interchanging two rows of I_n, then* $\det(E) = -1$.

(*c*) *If E results from adding a multiple of one row of I_n to another, then* $\det(E) = 1$.

Example 1. Evaluate the following determinants of elementary matrices.

$$\begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} \qquad \begin{vmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{vmatrix} \qquad \begin{vmatrix} 1 & 0 & 0 & 7 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}$$

Theorem 2.2.5. *If A is a square matrix with two proportional rows or two proportional columns, then* $\det(A) = 0$.

Example 2. What are the determinants of the following matrices?

$$\begin{bmatrix} -1 & 4 \\ -2 & 8 \end{bmatrix} \quad \begin{bmatrix} 1 & -2 & 7 \\ -4 & 8 & 5 \\ 2 & -4 & 3 \end{bmatrix} \quad \begin{bmatrix} 3 & -1 & 4 & -5 \\ 6 & -2 & 5 & 2 \\ 5 & 8 & 1 & 4 \\ -9 & 3 & -12 & 15 \end{bmatrix}$$

Example 3. Evaluate $\det(A)$ where

$$A = \begin{bmatrix} 0 & 1 & 5 \\ 3 & -6 & 9 \\ 2 & 6 & 1 \end{bmatrix}.$$

Example 4. Compute the determinant of

$$A = \begin{bmatrix} 1 & 0 & 0 & 3 \\ 2 & 7 & 0 & 6 \\ 0 & 6 & 3 & 0 \\ 7 & 3 & 1 & -5 \end{bmatrix}.$$

Example 5. Evaluate $\det(A)$ where

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

2.3 Properties of Determinants; Cramer's Rule

Remark 1. Suppose that A and B are $n \times n$ matrices and k is any scalar. Since a common factor of any row of a matrix can be moved through the determinant sign, and since each of the n rows in kA has a common factor of k, it follows that

$$\det(kA) = k^n \det(A).$$

Example 1. Consider

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}.$$

Calculate $\det(A)$, $\det(B)$, and $\det(A+B)$.

Theorem 2.3.1. *Let A, B, and C be $n \times n$ matrices that differ only in a single row, say the rth, and assume that the rth row of C can be obtained by adding corresponding entries in the rth rows of A and B. Then*

$$\det(C) = \det(A) + \det(B).$$

The same result holds for columns.

Example 2. Consider

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

Calculate $\det(A)$, $\det(B)$, and $\det(C)$.

Lemma 2.3.2. *If B is an $n \times n$ matrix and E is an $n \times n$ elementary matrix, then*
$$\det(EB) = \det(E)\det(B).$$

Proof. We will consider three cases, each in accordance with the row operation that produces the matrix E.
Case 1: If E results from multiplying a row of I_n by k, then EB results from B by multiplying the corresponding row by k; so we have

$$\det(EB) = k\det(B).$$

But we also have $\det(E) = k$, so

$$\det(EB) = \det(E)\det(B).$$

Cases 2 and 3: The proofs of the cases where E results from interchanging two rows of I_n or from adding a multiple of one row to another follow the same pattern as Case 1. □

Remark 2. It follows by repeated applications of Lemma 2.3.2 that if B is an $n \times n$ matrix and $E_1, E_2, \ldots, E_r$ are $n \times n$ elementary matrices, then

$$\det(E_1 E_2 \cdots E_r B) = \det(E_1)\det(E_2)\cdots\det(E_r)\det(B).$$

Theorem 2.3.3. *A square matrix A is invertible if and only if $\det(A) \neq 0$.*

Proof. Let R be the reduced row echelon form of A. As a preliminary step, we will show that $\det(A)$ and $\det(R)$ are both zero or both nonzero: Let $E_1, E_2, \ldots, E_r$ be the elementary matrices that correspond to the elementary row operations that produce R from A. Thus

$$R = E_r \cdots E_2 E_1 A$$

and so
$$\det(R) = \det(E_r)\cdots\det(E_2)\det(E_1)\det(A).$$

Since the determinant of an elementary matrix is nonzero, it follows that $\det(A)$ and $\det(R)$ are either both zero or both nonzero. If we assume first that A is invertible, then it follows that $R = I$ and hence that $\det(R) = 1$ ($\neq 0$). This, in turn, implies that $\det(A) \neq 0$.
Conversely, assume that $\det(A) \neq 0$. It follows from this that $\det(R) \neq 0$, which tells us that R cannot have a row of zeros. Thus $R = I$ and hence A is invertible. □

Example 3. Is the matrix

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

invertible?

Theorem 2.3.4. *If A and B are square matrices of the same size, then*

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

Proof. We divide the proof into two cases that depend on whether or not A is invertible. If the matrix A is not invertible, then neither is the product AB. Thus we have $\det(AB) = 0$ and $\det(A) = 0$, so it follows that $\det(AB) = \det(A)\det(B)$.
Now assume that A is invertible. Then the matrix A is expressible as a product of elementary matrices, say

$$A = E_1 E_2 \cdots E_r,$$

so

$$AB = E_1 E_2 \cdots E_r B.$$

Therefore,

$$\begin{aligned} \det(AB) &= \det(E_1)\det(E_2)\cdots\det(E_r)\det(B) \\ &= \det(E_1 E_2 \cdots E_r)\det(B) \\ &= \det(A)\det(B). \end{aligned}$$

□

Example 4. Consider the matrices

$$A = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} -1 & 3 \\ 5 & 8 \end{bmatrix}.$$

Calculate $\det(A)$, $\det(B)$, and $\det(AB)$.

Theorem 2.3.5. *If A is invertible, then*

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

Proof. Since $A^{-1}A = I$, it follows that $\det(A^{-1}A) = \det(I)$. Therefore, we must have $\det(A^{-1})\det(A) = 1$. Since $\det(A) \neq 0$, the proof can be completed by dividing through by $\det(A)$. $\square$

Example 5. Let

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

Compute $\det(A)$ using cofactor expansions along the first row and first column, and then compute the sum of the products of the entries in the first row by the corresponding cofactors in the second row, and the sum of the products of the entries in the first column by the corresponding cofactors in the second column.

Definition 2.3.1. If A is any $n \times n$ matrix and C_{ij} is the cofactor of a_{ij}, then the matrix

$$\begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{bmatrix}$$

is called the matrix of cofactors from A. The transpose of this matrix is called the adjoint of A and is denoted by $\text{adj}(A)$.

Example 6. Let

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

Find $\text{adj}(A)$.

Theorem 2.3.6 (Inverse of a Matrix Using Its Adjoint)**.** *If A is an invertible matrix, then*

$$A^{-1} = \frac{1}{\det(A)}\,\mathrm{adj}(A).$$

Proof. We show first that

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

Consider the product

$$A\,\mathrm{adj}(A) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{j1} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{j2} & \cdots & C_{n2} \\ \vdots & \vdots & & \vdots & & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{jn} & \cdots & C_{nn} \end{bmatrix}.$$

The entry in the ith row and jth column of the product $A\,\mathrm{adj}(A)$ is

$$a_{i1}C_{j1} + a_{i2}C_{j2} + \cdots + a_{in}C_{jn}.$$

If $i = j$, then this is the cofactor expansion of $\det(A)$ along the ith row of A, and if $i \neq j$, then the a's and the cofactors come from different rows of A, so the value of this entry is zero. Therefore,

$$A\,\mathrm{adj}(A) = \begin{bmatrix} \det(A) & 0 & \cdots & 0 \\ 0 & \det(A) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \det(A) \end{bmatrix} = \det(A)I.$$

Since A is invertible, $\det(A) \neq 0$. Therefore, we can write

$$\frac{1}{\det(A)}[A\,\mathrm{adj}(A)] = I \quad \text{or} \quad A\left[\frac{1}{\det(A)}\,\mathrm{adj}(A)\right] = I.$$

Multiplying both sides on the left by A^{-1} yields

$$A^{-1} = \frac{1}{\det(A)}\,\mathrm{adj}(A).$$

□

Example 7. Use Theorem 2.3.6 to find the inverse of the matrix A in Example 6.

Theorem 2.3.7 (Cramer's Rule)**.** *If $A\mathbf{x} = \mathbf{b}$ is a system of n linear equations in n unknowns such that $\det(A) \neq 0$, then the system has a unique solution. This solution is*

$$x_1 = \frac{\det(A_1)}{\det(A)}, \quad x_2 = \frac{\det(A_2)}{\det(A)}, \ldots, \quad x_n = \frac{\det(A_n)}{\det(A)},$$

where A_j is the matrix obtained by replacing the entries in the jth column of A by the entries in the matrix

$$\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

Proof. If $\det(A) \neq 0$, then A is invertible and $\mathbf{x} = A^{-1}\mathbf{b}$ is the unique solution of $A\mathbf{x} = \mathbf{b}$. Therefore, we have

$$\begin{aligned} \mathbf{x} = A^{-1}\mathbf{b} = \frac{1}{\det(A)}\operatorname{adj}(A)\mathbf{b} &= \frac{1}{\det(A)} \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \\ &= \frac{1}{\det(A)} \begin{bmatrix} b_1C_{11} + & b_2C_{21} & + \cdots + b_nC_{n1} \\ b_1C_{12} + & b_2C_{22} & + \cdots + b_nC_{n2} \\ \vdots & \vdots & \vdots \\ b_1C_{1n} + & b_2C_{2n}x_2 & + \cdots + b_nC_{nn} \end{bmatrix}. \end{aligned}$$

The entry in the jth row of $\mathbf{x}$ is therefore

$$x_j = \frac{b_1C_{1j} + b_2C_{2j} + \cdots + b_nC_{nj}}{\det(A)}.$$

Now let

$$A_j = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j-1} & b_1 & a_{1j+1} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2j-1} & b_2 & a_{2j+1} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nj-1} & b_n & a_{nj+1} & \cdots & a_{nn} \end{bmatrix}.$$

Since A_j differs from A only in the jth column, it follows that the cofactors of entries $b_1, b_2, \ldots, b_n$ in A_j are the same as the cofactors of the corresponding entries in the jth column of A. The cofactor expansion of $\det(A_j)$ along the jth column is therefore

$$\det(A_j) = b_1C_{1j} + b_2C_{2j} + \cdots + b_nC_{nj}.$$

Substituting this result in gives

$$x_j = \frac{\det(A_j)}{\det(A)}. \qquad \square$$

Example 8. Use Cramer's rule to solve

$$\begin{aligned} x_1 \phantom{{}+4x_2} + 2x_3 &= 6 \\ -3x_1 + 4x_2 + 6x_3 &= 30 \\ -x_1 - 2x_2 + 3x_3 &= 8. \end{aligned}$$

Theorem 2.3.8 (Equivalent Statements)**.** *If A is an $n \times n$ matrix, then the following are equivalent.*

(a) *A is invertible.*
(b) *$A\mathbf{x} = \mathbf{0}$ has only the trivial solution.*
(c) *The reduced row echelon form of A is I_n.*
(d) *A is expressible as a product of elementary matrices.*
(e) *$A\mathbf{x} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.*
(f) *$A\mathbf{x} = \mathbf{b}$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.*
(g) $\det(A) \neq 0$.

Chapter 3

Euclidean Vector Spaces

3.1 Vectors in 2-Space, 3-Space, and n-Space

Remark 1. Geometric vectors in two dimensions (also called 2-space) or in three dimensions (also called 3-space) are represented by arrows. The direction of the arrowhead specifies the direction of the vector and the length of the arrow specifies the magnitude. The tail of the arrow is called the initial point of the vector and the tip the terminal point.

We will denote vectors in boldface type such as **a**, **b**, **v**, **w**, and **x**, and we will denote scalars in lowercase italic type such as a, k, v, w, and x. When we want to indicate that a vector **v** has initial point A and terminal point B, then we will write

$$\mathbf{v} = \overrightarrow{AB}.$$

Vectors with the same length and direction are said to be equivalent. Equivalent vectors are also said to be equal, which we indicate by writing

$$\mathbf{v} = \mathbf{w}.$$

The vector whose initial and terminal points coincide has length zero, so we call this the zero vector and denote it by $\mathbf{0}$.

Definition 3.1.1 (Parallelogram Rule for Vector Addition)**.** If **v** and **w** are vectors in 2-space or 3-space that are positioned so their initial points coincide, then the two vectors form adjacent sides of a parallelogram, and the sum $\mathbf{v} + \mathbf{w}$ is the vector represented by the arrow from the common initial point of **v** and **w** to the opposite vertex of the parallelogram.

Definition 3.1.2 (Triangle Rule for Vector Addition)**.** If $\mathbf{v}$ and $\mathbf{w}$ are vectors in 2-space or 3-space that are positioned so the initial point of $\mathbf{w}$ is at the terminal point of $\mathbf{v}$, then the sum $\mathbf{v} + \mathbf{w}$ is the vector represented by the arrow from the common initial point of $\mathbf{v}$ and $\mathbf{w}$ to the terminal point of $\mathbf{w}$.

Remark 2 (Vector Addition Viewed as Translation)*.* If $\mathbf{v}$, $\mathbf{w}$, and $\mathbf{v} + \mathbf{w}$ are positioned so their initial points coincide, then the terminal point of $\mathbf{v} + \mathbf{w}$ can be viewed in two ways:

1. The terminal point of $\mathbf{v} + \mathbf{w}$ is the point that results when the terminal point of $\mathbf{v}$ is translated in the direction of $\mathbf{w}$ by a distance equal to the length of $\mathbf{w}$.
2. The terminal point of $\mathbf{v} + \mathbf{w}$ is the point that results when the terminal point of $\mathbf{w}$ is translated in the direction of $\mathbf{v}$ by a distance equal to the length of $\mathbf{v}$.

Definition 3.1.3 (Vector Subtraction)**.** The negative of a vector $\mathbf{v}$, denoted by $-\mathbf{v}$, is the vector that has the same length as $\mathbf{v}$ but is oppositely directed, and the difference of $\mathbf{v}$ from $\mathbf{w}$, denoted by $\mathbf{w} - \mathbf{v}$, is taken to be the sum

$$\mathbf{w} - \mathbf{v} = \mathbf{w} + (-\mathbf{v}).$$

Definition 3.1.4 (Scalar Multiplication)**.** If $\mathbf{v}$ is a nonzero vector in 2-space or 3-space, and if k is a nonzero scalar, then we define the scalar product of v by k to be the vector whose length is $|k|$ times the length of $\mathbf{v}$ and whose direction is the same as that of $\mathbf{v}$ if k is positive and opposite to that of $\mathbf{v}$ if k is negative. If $k = 0$ or $\mathbf{v} = \mathbf{0}$, then we define $k\mathbf{v}$ to be $\mathbf{0}$.

Remark 3. Observe that $(-1)\mathbf{v}$ has the same length as $\mathbf{v}$ but is oppositely directed; therefore,

$$(-1)\mathbf{v} = -\mathbf{v}.$$

Remark 4. Since translating a vector does not change it, we agree that the terms *parallel* and *collinear* mean the same thing when applied to vectors. We regard the vector $\mathbf{0}$ as parallel to all vectors.

Remark 5. Vector addition satisfies the associative law for addition, that is,

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

Remark 6. If a vector $\mathbf{v}$ in 2-space or 3-space is positioned with its initial point at the origin of a rectangular coordinate system, then the vector is completely determined by the coordinates of its terminal point. We call these coordinates the components of $\mathbf{v}$ relative to the coordinate system. We will write $\mathbf{v} = (v_1, v_2)$ to denote a vector $\mathbf{v}$ in 2-space with components (v_1, v_2), and

$\mathbf{v} = (v_1, v_2, v_3)$ to denote a vector $\mathbf{v}$ in 3-space with components (v_1, v_2, v_3).

Two vectors in 2-space or 3-space are equivalent if and only if they have the same terminal point when their initial points are at the origin. Algebraically, this means that two vectors are equivalent if and only if their corresponding components are equal. Thus, for example, the vectors

$$\mathbf{v} = (v_1, v_2, v_3) \quad \text{and} \quad \mathbf{w} = (w_1, w_2, w_3)$$

in 3-space are equivalent if and only if

$$v_1 = w_1, \quad v_2 = w_2, \quad v_3 = w_3.$$

Remark 7. If $\overrightarrow{P_1P_2}$ denotes the vector with initial point $P_1(x_1, y_1)$ and terminal point $P_2(x_2, y_2)$, then the components of this vector are given by the formula

$$\overrightarrow{P_1P_2} = (x_2 - x_1, y_2 - y_1).$$

The components of a vector in 3-space that has initial point $P_1(x_1, y_1, z_1)$ and terminal point $P_2(x_2, y_2, z_2)$ are given by

$$\overrightarrow{P_1P_2} = (x_2 - x_1, y_2 - y_1, z_2 - z_1).$$

Example 1. What are the components of the vector $\mathbf{v} = \overrightarrow{P_1P_2}$ with initial point $P_1(2, -1, 4)$ and terminal point $P_2(7, 5, -8)$?

Definition 3.1.5. If n is a positive integer, then an ordered n-tuple is a sequence of n real numbers $(v_1, v_2, \ldots, v_n)$. The set of all ordered n-tuples is called n-space and is denoted by R^n.

Definition 3.1.6. Vectors $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ and $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ in R^n are said to be equivalent (also called equal) if

$$v_1 = w_1, \quad v_2 = w_2, \ldots, \quad v_n = w_n.$$

We indicate this by writing $\mathbf{v} = \mathbf{w}$.

Example 2. When is

$$(a, b, c, d) = (1, -4, 2, 7)$$

true?

Definition 3.1.7. If $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ and $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ are vectors in R^n, and if k is any scalar, then we define

$$\begin{aligned}
&\mathbf{v} + \mathbf{w} = (v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n) \\
&k\mathbf{v} = (kv_1, kv_2, \ldots, kv_n) \\
&-\mathbf{v} = (-v_1, -v_2, \ldots, -v_n) \\
&\mathbf{w} - \mathbf{v} = \mathbf{w} + (-\mathbf{v}) = (w_1 - v_1, w_2 - v_2, \ldots, w_n - v_n).
\end{aligned}$$

Example 3. If $\mathbf{v} = (1, -3, 2)$ and $\mathbf{w} = (4, 2, 1)$, then find $\mathbf{v} + \mathbf{w}$, $2\mathbf{v}$, $-\mathbf{w}$, and $\mathbf{v} - \mathbf{w}$.

Theorem 3.1.1. *If* $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *are vectors in* R^n, *and if* k *and* m *are scalars, then:*

(*a*) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
(*b*) $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$
(*c*) $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$
(*d*) $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$
(*e*) $k(\mathbf{u} + \mathbf{v}) = k\mathbf{u} + k\mathbf{v}$
(*f*) $(k + m)\mathbf{u} = k\mathbf{u} + m\mathbf{u}$
(*g*) $k(m\mathbf{u}) = (km)\mathbf{u}$
(*h*) $1\mathbf{u} = \mathbf{u}$

Theorem 3.1.2. *If* $\mathbf{v}$ *is a vector in* R^n *and* k *is a scalar, then:*

(*a*) $0\mathbf{v} = \mathbf{0}$
(*b*) $k\mathbf{0} = \mathbf{0}$
(*c*) $(-1)\mathbf{v} = -\mathbf{v}$

Definition 3.1.8. If $\mathbf{w}$ is a vector in R^n, then $\mathbf{w}$ is said to be a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ in R^n if it can be expressed in the form

$$\mathbf{w} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_r\mathbf{v}_r$$

where $k_1, k_2, \ldots, k_r$ are scalars. These scalars are called the coefficients of the linear combination. In the case where $r = 1$, this formula becomes $\mathbf{w} = k_1\mathbf{v}_1$, so that a linear combination of a single vector is just a scalar multiple of that vector.

3.2 Norm, Dot Product, and Distance in R^n

Definition 3.2.1. If $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ is a vector in R^n, then the norm of $\mathbf{v}$ (also called the length of $\mathbf{v}$ or the magnitude of $\mathbf{v}$) is denoted by $\|\mathbf{v}\|$, and is defined by the formula

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

Example 1. Find the norm of the vector $\mathbf{v} = (-3, 2, 1)$ in R^3 and the norm of the vector $\mathbf{v} = (2, -1, 3, -5)$ in R^4.

Theorem 3.2.1. *If $\mathbf{v}$ is a vector in R^n, and if k is any scalar, then:*
(a) $\|\mathbf{v}\| \geq 0$
(b) $\|\mathbf{v}\| = 0$ *if and only if* $\mathbf{v} = \mathbf{0}$
(c) $\|k\mathbf{v}\| = |k|\|\mathbf{v}\|$

Remark 1. A vector of norm 1 is called a unit vector. If $\mathbf{v}$ is any nonzero vector in R^n, then

$$\mathbf{u} = \frac{1}{\|\mathbf{v}\|}\mathbf{v}$$

defines a unit vector that is in the same direction as $\mathbf{v}$. The process of multiplying a nonzero vector by the reciprocal of its length to obtain a unit vector is called normalizing $\mathbf{v}$.

Example 2. Find the unit vector $\mathbf{u}$ that has the same direction as $\mathbf{v} = (2, 2, -1)$.

Remark 2. When a rectangular coordinate system is introduced in R^2 or R^3, the unit vectors in the positive directions of the coordinate axes are called the standard unit vectors. In R^2 these vectors are denoted by

$$\mathbf{i} = (1,0) \quad \text{and} \quad \mathbf{j} = (0,1)$$

and in R^3 by

$$\mathbf{i} = (1,0,0), \quad \mathbf{j} = (0,1,0), \quad \text{and} \quad \mathbf{k} = (0,0,1).$$

Every vector $\mathbf{v} = (v_1, v_2)$ in R^2 and every vector $\mathbf{v} = (v_1, v_2, v_3)$ in R^3 can be expressed as a linear combination of standard unit vectors by writing

$$\begin{aligned} \mathbf{v} &= (v_1, v_2) = v_1(1,0) + v_2(0,1) = v_1\mathbf{i} + v_2\mathbf{j} \\ \mathbf{v} &= (v_1, v_2, v_3) = v_1(1,0,0) + v_2(0,1,0) + v_3(0,0,1) = v_1\mathbf{i} + v_2\mathbf{j} + v_3\mathbf{k}. \end{aligned}$$

Moreover, we can generalize these formulas to R^n by defining the standard unit vectors in R^n to be

$$\mathbf{e}_1 = (1,0,0,\ldots,0), \quad \mathbf{e}_2 = (0,1,0,\ldots,0), \ldots, \quad \mathbf{e}_n = (0,0,0,\ldots,1)$$

in which case every vector $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ in R^n can be expressed as

$$\mathbf{v} = (v_1, v_2, \ldots, v_n) = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + \cdots + v_n\mathbf{e}_n.$$

Example 3. Write the vectors $(2, -3, 4)$ and $(7, 3, -4, 5)$ as linear combinations of standard unit vectors.

Definition 3.2.2. If $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ are points in R^n, then we denote the distance between $\mathbf{u}$ and $\mathbf{v}$ by $d(\mathbf{u}, \mathbf{v})$ and define it to be

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_n - v_n)^2}.$$

Example 4. If

$$\mathbf{u} = (1, 3, -2, 7) \quad \text{and} \quad \mathbf{v} = (0, 7, 2, 2)$$

then find the distance between $\mathbf{u}$ and $\mathbf{v}$.

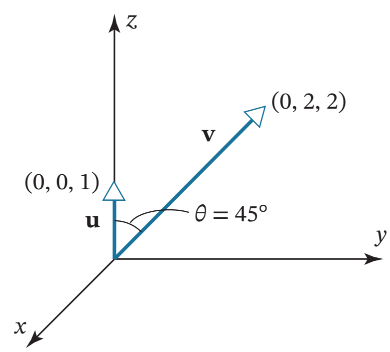
Remark 3. Let $\mathbf{u}$ and $\mathbf{v}$ be nonzero vectors in R^2 or R^3 that have been positioned so that their initial points coincide. We define the angle between u and v to be the angle θ determined by $\mathbf{u}$ and $\mathbf{v}$ that satisfies the inequalities $0 \leq \theta \leq \pi$.

Definition 3.2.3. If $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors in R^2 or R^3, and if θ is the angle between $\mathbf{u}$ and $\mathbf{v}$, then the dot product (also called the Euclidean inner product) of $\mathbf{u}$ and $\mathbf{v}$ is denoted by $\mathbf{u} \cdot \mathbf{v}$ and is defined as

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

If $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$, then we define $\mathbf{u} \cdot \mathbf{v}$ to be 0.

Example 5. Find the dot product of the vectors shown in the figure.



Definition 3.2.4. If $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ are vectors in R^n, then the dot product (also called the Euclidean inner product) of $\mathbf{u}$ and $\mathbf{v}$ is denoted by $\mathbf{u} \cdot \mathbf{v}$ and is defined by

$$\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n.$$

Example 6.

(a) Use Definition 3.2.4 to compute the dot product of the vectors $\mathbf{u}$ and $\mathbf{v}$ in Example 5.

(b) Calculate $\mathbf{u} \cdot \mathbf{v}$ for the following vectors in R^4:

$$\mathbf{u} = (-1, 3, 5, 7), \quad \mathbf{v} = (-3, -4, 1, 0).$$

Example 7. Find the angle between a diagonal of a cube and one of its edges.

Remark 4. In the special case where $\mathbf{u} = \mathbf{v}$ in Definition 3.2.4, we obtain the relationship

$$\mathbf{v} \cdot \mathbf{v} = v_1^2 + v_2^2 + \cdots + v_n^2 = \|\mathbf{v}\|^2.$$

This yields the following formula for expressing the length of a vector in terms of a dot product:

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}.$$

Theorem 3.2.2. *If* $\mathbf{u}$*,* $\mathbf{v}$*, and* $\mathbf{w}$ *are vectors in* R^n*, and if* k *is a scalar, then:*

(a) $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
(b) $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$
(c) $k(\mathbf{u} \cdot \mathbf{v}) = (k\mathbf{u}) \cdot \mathbf{v}$
(d) $\mathbf{v} \cdot \mathbf{v} \geq 0$ *and* $\mathbf{v} \cdot \mathbf{v} = 0$ *if and only if* $\mathbf{v} = \mathbf{0}$

Theorem 3.2.3. *If* $\mathbf{u}$*,* $\mathbf{v}$*, and* $\mathbf{w}$ *are vectors in* R^n*, and if* k *is a scalar, then:*

(a) $\mathbf{0} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{0} = 0$
(b) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$
(c) $\mathbf{u} \cdot (\mathbf{v} - \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} - \mathbf{u} \cdot \mathbf{w}$
(d) $(\mathbf{u} - \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} - \mathbf{v} \cdot \mathbf{w}$
(e) $k(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (k\mathbf{v})$

Example 8. Calculate $(\mathbf{u} - 2\mathbf{v}) \cdot (3\mathbf{u} + 4\mathbf{v})$.

Theorem 3.2.4 (Cauchy-Schwarz Inequality)**.** *If* $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ *and* $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ *are vectors in* R^n*, then*

$$|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\|\|\mathbf{v}\|$$

or in terms of components

$$|u_1v_1 + u_2v_2 + \cdots + u_nv_n| \leq (u_1^2 + u_2^2 + \cdots + u_n^2)^{1/2}(v_1^2 + v_2^2 + \cdots + v_n^2)^{1/2}.$$

Theorem 3.2.5. *If* $\mathbf{u}$*,* $\mathbf{v}$*, and* $\mathbf{w}$ *are vectors in* R^n*, then:*

(a) $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$
(b) $d(\mathbf{u}, \mathbf{v}) \leq d(\mathbf{u}, \mathbf{w}) + d(\mathbf{w}, \mathbf{v})$

Proof. (a)

$$\begin{aligned}
\|\mathbf{u} + \mathbf{v}\|^2 &= (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) = (\mathbf{u} \cdot \mathbf{u}) + 2(\mathbf{u} \cdot \mathbf{v}) + (\mathbf{v} \cdot \mathbf{v}) \\
&= \|\mathbf{u}\|^2 + 2(\mathbf{u} \cdot \mathbf{v}) + \|\mathbf{v}\|^2 \\
&\leq \|\mathbf{u}\|^2 + 2|\mathbf{u} \cdot \mathbf{v}| + \|\mathbf{v}\|^2 \\
&\leq \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| + \|\mathbf{v}\|^2 \\
&= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2
\end{aligned}$$

(b)

$$\begin{aligned}
d(\mathbf{u}, \mathbf{v}) &= \|\mathbf{u} - \mathbf{v}\| = \|(\mathbf{u} - \mathbf{w}) + (\mathbf{w} - \mathbf{v})\| \\
&\leq \|\mathbf{u} - \mathbf{w}\| + \|\mathbf{w} - \mathbf{v}\| = d(\mathbf{u}, \mathbf{w}) + d(\mathbf{w}, \mathbf{v}).
\end{aligned}$$

□

Theorem 3.2.6 (Parallelogram Equation for Vectors)**.** *If* $\mathbf{u}$ *and* $\mathbf{v}$ *are vectors in* R^n*, then*

$$\|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2 = 2\left(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2\right).$$

Proof.

$$\begin{aligned}
\|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2 &= (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) + (\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v}) \\
&= 2(\mathbf{u} \cdot \mathbf{u}) + 2(\mathbf{v} \cdot \mathbf{v}) \\
&= 2\left(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2\right).
\end{aligned}$$

□

Theorem 3.2.7. *If* $\mathbf{u}$ *and* $\mathbf{v}$ *are vectors in* R^n *with the Euclidean inner product, then*

$$\mathbf{u} \cdot \mathbf{v} = \frac{1}{4}\|\mathbf{u} + \mathbf{v}\|^2 - \frac{1}{4}\|\mathbf{u} - \mathbf{v}\|^2.$$

Proof.

$$\begin{aligned}
\|\mathbf{u} + \mathbf{v}\|^2 &= (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) = \|\mathbf{u}\|^2 + 2(\mathbf{u} \cdot \mathbf{v}) + \|\mathbf{v}\|^2 \\
\|\mathbf{u} - \mathbf{v}\|^2 &= (\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v}) = \|\mathbf{u}\|^2 - 2(\mathbf{u} \cdot \mathbf{v}) + \|\mathbf{v}\|^2
\end{aligned}$$

from which the result follows by simple algebra. □

Remark 5. If A is an $n \times n$ matrix and $\mathbf{u}$ and $\mathbf{v}$ are $n \times 1$ matrices, then

$$A\mathbf{u} \cdot \mathbf{v} = \mathbf{u} \cdot A^T\mathbf{v}$$
$$\mathbf{u} \cdot A\mathbf{v} = A^T\mathbf{u} \cdot \mathbf{v}.$$

Example 9. Suppose that

$$A = \begin{bmatrix} 1 & -2 & 3 \\ 2 & 4 & 1 \\ -1 & 0 & 1 \end{bmatrix}, \quad \mathbf{u} = \begin{bmatrix} -1 \\ 2 \\ 4 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} -2 \\ 0 \\ 5 \end{bmatrix}.$$

Verify that $A\mathbf{u} \cdot \mathbf{v} = \mathbf{u} \cdot A^T\mathbf{v}$.

3.3 Orthogonality

Definition 3.3.1. Two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ in R^n are said to be orthogonal (or perpendicular) if $\mathbf{u} \cdot \mathbf{v} = 0$. We will also agree that the zero vector in R^n is orthogonal to *every* vector in R^n.

Example 1.

(a) Show that $\mathbf{u} = (-2, 3, 1, 4)$ and $\mathbf{v} = (1, 2, 0, -1)$ are orthogonal vectors in R^4.

(b) Let $S = \{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ be the set of standard unit vectors in R^3. Show that each ordered pair of vectors in S is orthogonal.

Remark 1. If $\mathbf{n}$ is a *nonzero* vector, called a normal, that is orthogonal to a line or plane, then

$$a(x - x_0) + b(y - y_0) = 0$$
$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$$

are called the point-normal equations of the line through the point $P_0(x_0, y_0)$ that has normal $\mathbf{n} = (a, b)$ and the plane through the point $P_0(x_0, y_0, z_0)$ that has normal $\mathbf{n} = (a, b, c)$.

Example 2. Write equations that represent the line through the point $(3, -7)$ with normal $\mathbf{n} = (6, 1)$ and the plane through the point $(3, 0, 7)$ with normal $\mathbf{n} = (4, 2, -5)$.

Theorem 3.3.1.

(a) *If a and b are constants that are not both zero, then an equation of the form*

$$ax + by + c = 0$$

represents a line in R^2 with normal $\mathbf{n} = (a, b)$.

(b) *If a, b, and c are constants that are not all zero, then an equation of the form*

$$ax + by + cz + d = 0$$

represents a plane in R^3 with normal $\mathbf{n} = (a, b, c)$.

Example 3.

(a) The equation $ax+by = 0$ represents a line through the origin in R^2. Show that the vector $\mathbf{n}_1 = (a, b)$ formed from the coefficients of the equation is orthogonal to the line, that is, orthogonal to every vector along the line.

(b) The equation $ax + by + cz = 0$ represents a plane through the origin in R^3. Show that the vector $\mathbf{n}_2 = (a, b, c)$ formed from the coefficients of the equation is orthogonal to the plane, that is, orthogonal to every vector that lies in the plane.

Theorem 3.3.2 (Projection Theorem)**.** *If* $\mathbf{u}$ *and* $\mathbf{a}$ *are vectors in* R^n*, and* $\mathbf{a} \neq \mathbf{0}$*, then* $\mathbf{u}$ *can be expressed in exactly one way in the form* $\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2$*, where* $\mathbf{w}_1$ *is a scalar multiple of* $\mathbf{a}$ *and* $\mathbf{w}_2$ *is orthogonal to* $\mathbf{a}$*.*

Proof. Since the vector $\mathbf{w}_1$ is to be a scalar multiple of $\mathbf{a}$, it must have the form

$$\mathbf{w}_1 = k\mathbf{a}.$$

Our goal is to find a value of the scalar k and a vector $\mathbf{w}_2$ that is orthogonal to $\mathbf{a}$ such that

$$\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2.$$

We can determine k by writing

$$\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2 = k\mathbf{a} + \mathbf{w}_2$$

and thus

$$\mathbf{u} \cdot \mathbf{a} = (k\mathbf{a} + \mathbf{w}_2) \cdot \mathbf{a} = k\|\mathbf{a}\|^2 + (\mathbf{w}_2 \cdot \mathbf{a}).$$

Since $\mathbf{w}_2$ is to be orthogonal to $\mathbf{a}$, $\mathbf{w}_2 \cdot \mathbf{a}$ must be 0, and hence k must satisfy the equation

$$\mathbf{u} \cdot \mathbf{a} = k\|\mathbf{a}\|^2$$

from which we obtain

$$k = \frac{\mathbf{u} \cdot \mathbf{a}}{\|\mathbf{a}\|^2}$$

as the only possible value for k. Then writing

$$\mathbf{w}_2 = \mathbf{u} - \mathbf{w}_1 = \mathbf{u} - k\mathbf{a} = \mathbf{u} - \frac{\mathbf{u} \cdot \mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a}$$

we see that

$$\mathbf{w}_2 \cdot \mathbf{a} = \left(\mathbf{u} - \frac{\mathbf{u} \cdot \mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a}\right) \cdot \mathbf{a} = \mathbf{u} \cdot \mathbf{a} - \mathbf{u} \cdot \mathbf{a} = 0.$$ □

Remark 2. The vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ in the Projection Theorem have associated names—the vector $\mathbf{w}_1$ is called the <u>orthogonal projection of $\mathbf{u}$ on $\mathbf{a}$</u> or sometimes the <u>vector component of $\mathbf{u}$ along $\mathbf{a}$</u>, and the vector $\mathbf{w}_2$ is called the <u>vector component of $\mathbf{u}$ orthogonal to $\mathbf{a}$</u>. The vector $\mathbf{w}_1$ is commonly denoted by

$$\operatorname{proj}_{\mathbf{a}} \mathbf{u} = \frac{\mathbf{u} \cdot \mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a},$$

in which case it follows that $\mathbf{w}_2$ is

$$\mathbf{u} - \operatorname{proj}_{\mathbf{a}} \mathbf{u} = \mathbf{u} - \frac{\mathbf{u} \cdot \mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a}.$$

Example 4. Let $\mathbf{u} = (2, -1, 3)$ and $\mathbf{a} = (4, -1, 2)$. Find the vector component of $\mathbf{u}$ along $\mathbf{a}$ and the vector component of $\mathbf{u}$ orthogonal to $\mathbf{a}$.

Example 5.

(a) Find the orthogonal projections of the vectors $\mathbf{e}_1 = (1, 0)$ and $\mathbf{e}_2 = (0, 1)$ on the line L that makes an angle θ with the positive x-axis.

(b) Use the result in part (a) to find the standard matrix for the operator $T : R^2 \to R^2$ that maps each point orthogonally onto L.

Example 6. Use part (b) of Example 5 to find the orthogonal projection of the vector $\mathbf{x} = (1, 5)$ onto the line through the origin that makes an angle of $\pi/6$ $(= 30^\circ)$ with the positive x-axis.

Remark 3. The reflection about a line L through the origin that makes an angle θ with the positive x-axis is given by

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

Example 7. Find the reflection of the vector $\mathbf{x} = (1, 5)$ about the line through the origin that makes an angle of $\pi/6$ $(= 30^\circ)$ with the x-axis.

Remark 4. A formula for the norm of the vector component of $\mathbf{u}$ along $\mathbf{a}$ can be derived as follows:

$$\|\operatorname{proj}_{\mathbf{a}} \mathbf{u}\| = \left\| \frac{\mathbf{u} \cdot \mathbf{a}}{\|\mathbf{a}\|^2} \mathbf{a} \right\| = \left| \frac{\mathbf{u} \cdot \mathbf{a}}{\|\mathbf{a}\|^2} \right| \|\mathbf{a}\| = \frac{|\mathbf{u} \cdot \mathbf{a}|}{\|\mathbf{a}\|^2} \|\mathbf{a}\|.$$

Thus,

$$\|\operatorname{proj}_{\mathbf{a}} \mathbf{u}\| = \frac{|\mathbf{u} \cdot \mathbf{a}|}{\|\mathbf{a}\|}.$$

Theorem 3.3.3 (Theorem of Pythagoras in R^n)**.** *If* $\mathbf{u}$ *and* $\mathbf{v}$ *are orthogonal vectors in* R^n *with the Euclidean inner product, then*

$$\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

Proof. Since $\mathbf{u}$ and $\mathbf{v}$ are orthogonal, we have $\mathbf{u}\cdot\mathbf{v} = 0$, from which it follows that

$$\|\mathbf{u}+\mathbf{v}\|^2 = (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v}) = \|\mathbf{u}\|^2 + 2(\mathbf{u}\cdot\mathbf{v}) + \|\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2. \qquad \square$$

Example 8. We showed in Example 1 that the vectors

$$\mathbf{u} = (-2,3,1,4) \quad \text{and} \quad \mathbf{v} = (1,2,0,-1)$$

are orthogonal. Verify the Theorem of Pythagoras for these vectors.

Theorem 3.3.4.

(*a*) *In* R^2 *the distance* D *between the point* $P_0(x_0, y_0)$ *and the line* $ax+by+c=0$ *is*

$$D = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}.$$

(*b*) *In* R^3 *the distance* D *between the point* $P_0(x_0, y_0, z_0)$ *and the plane* $ax + by + cz + d = 0$ *is*

$$D = \frac{|ax_0 + by_0 + cz_0 + d|}{\sqrt{a^2 + b^2 + c^2}}.$$

Example 9. Find the distance D between the point $(1, -4, -3)$ and the plane $2x - 3y + 6z = -1$.

Example 10. The planes

$$x + 2y - 2z = 3 \quad \text{and} \quad 2x + 4y - 4z = 7$$

are parallel since their normals, $(1, 2, -2)$ and $(2, 4, -4)$, are parallel vectors. Find the distance between these planes.

3.4 The Geometry of Linear Systems

Theorem 3.4.1. *Let L be the line in R^2 or R^3 that contains the point $\mathbf{x}_0$ and is parallel to the nonzero vector $\mathbf{v}$. Then the equation of the line through $\mathbf{x}_0$ that is parallel to $\mathbf{v}$ is*

$$\mathbf{x} = \mathbf{x}_0 + t\mathbf{v},$$

where the variable t is called a parameter. If $\mathbf{x}_0 = \mathbf{0}$, then the line passes through the origin and the equation has the form

$$\mathbf{x} = t\mathbf{v}.$$

Theorem 3.4.2. *Let W be the plane in R^3 that contains the point $\mathbf{x}_0$ and is parallel to the noncollinear vectors $\mathbf{v}_1$ and $\mathbf{v}_2$. Then an equation of the plane through $\mathbf{x}_0$ that is parallel to $\mathbf{v}_1$ and $\mathbf{v}_2$ is given by*

$$\mathbf{x} = \mathbf{x}_0 + t_1\mathbf{v}_1 + t_2\mathbf{v}_2,$$

where the variables t_1 and t_2 are called parameters. If $\mathbf{x}_0 = \mathbf{0}$, then the plane passes through the origin and the equation has the form

$$\mathbf{x} = t_1\mathbf{v}_1 + t_2\mathbf{v}_2.$$

Definition 3.4.1. If $\mathbf{x}_0$ and $\mathbf{v}$ are vectors in R^n, and if $\mathbf{v}$ is nonzero, then the equation

$$\mathbf{x} = \mathbf{x}_0 + t\mathbf{v}$$

defines the line through $\mathbf{x}_0$ that is parallel to $\mathbf{v}$. In the special case where $\mathbf{x}_0 = \mathbf{0}$, the line is said to pass through the origin.

Definition 3.4.2. If $\mathbf{x}_0$, $\mathbf{v}_1$, and $\mathbf{v}_2$ are vectors in R^n, and if $\mathbf{v}_1$ and $\mathbf{v}_2$ are not collinear, then the equation

$$\mathbf{x} = \mathbf{x}_0 + t_1\mathbf{v}_1 + t_2\mathbf{v}_2$$

defines the plane through $\mathbf{x}_0$ that is parallel to $\mathbf{v}_1$ and $\mathbf{v}_2$. In the special case where $\mathbf{x}_0 = \mathbf{0}$, the plane is said to pass through the origin.

Example 1.

(a) Find a vector equation and parametric equations of the line in R^2 that passes through the origin and is parallel to the vector $\mathbf{v} = (-2, 3)$.

(b) Find a vector equation and parametric equations of the line in R^3 that passes through the point $P_0(1, 2, -3)$ and is parallel to the vector $\mathbf{v} = (4, -5, 1)$.

(c) Use the vector equation obtained in part (b) to find two points on the line that are different from P_0.

Example 2. Find vector and parametric equations of the plane $x - y + 2z = 5$.

Example 3.

(a) Find vector and parametric equations of the line through the origin of R^4 that is parallel to the vector $\mathbf{v} = (5, -3, 6, 1)$.

(b) Find vector and parametric equations of the plane in R^4 that passes through the point $\mathbf{x}_0 = (2, -1, 0, 3)$ and is parallel to both $\mathbf{v}_1 = (1, 5, 2, -4)$ and $\mathbf{v}_2 = (0, 7, -8, 6)$.

Remark 1. If $\mathbf{x}_0$ and $\mathbf{x}_1$ are distinct points in R^n, then the line determined by these points is parallel to the vector $\mathbf{v} = \mathbf{x}_1 - \mathbf{x}_0$, so it follows that the line can be expressed in vector form as

$$\mathbf{x} = \mathbf{x}_0 + t(\mathbf{x}_1 - \mathbf{x}_0)$$

or, equivalently, as

$$\mathbf{x} = (1 - t)\mathbf{x}_0 + t\mathbf{x}_1.$$

These are called the <u>two-point vector equations</u> of a line in R^n.

Example 4. Find vector and parametric equations for the line in R^2 that passes through the points $P(0,7)$ and $Q(5,0)$.

Definition 3.4.3. If $\mathbf{x}_0$ and $\mathbf{x}_1$ are vectors in R^n, then the equation

$$\mathbf{x} = \mathbf{x}_0 + t(\mathbf{x}_1 - \mathbf{x}_0) \quad (0 \leq t \leq 1)$$

defines the line segment from $\mathbf{x}_0$ to $\mathbf{x}_1$. When convenient, this equation can be written as

$$\mathbf{x} = (1-t)\mathbf{x}_0 + t\mathbf{x}_1 \quad (0 \leq t \leq 1).$$

Example 5. Find equations for the line segment in R^2 from $\mathbf{x}_0 = (1,-3)$ to $\mathbf{x}_1 = (5,6)$.

Theorem 3.4.3. *If A is an $m \times n$ matrix, then the solution set of the homogeneous linear system $A\mathbf{x} = \mathbf{0}$ consists of all vectors in R^n that are orthogonal to every row vector of A.*

Example 6. We showed in Example 6 of Section 1.2 that the general solution of the homogeneous linear system

$$\begin{bmatrix} 1 & 3 & -2 & 0 & 2 & 0 \\ 2 & 6 & -5 & -2 & 4 & -3 \\ 0 & 0 & 5 & 10 & 0 & 15 \\ 2 & 6 & 0 & 8 & 4 & 18 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

is

$$x_1 = -3r - 4s - 2t, \quad x_2 = r, \quad x_3 = -2s \quad x_4 = s, \quad x_5 = t, \quad x_6 = 0.$$

Verify Theorem 3.4.3 for this system.

Theorem 3.4.4. *The general solution of a consistent linear system* $A\mathbf{x} = \mathbf{b}$ *can be obtained by adding any specific solution of* $A\mathbf{x} = \mathbf{b}$ *to the general solution of* $A\mathbf{x} = \mathbf{0}$.

Proof. Let $\mathbf{x}_0$ be any specific solution of $A\mathbf{x} = \mathbf{b}$, let W denote the solution set of $A\mathbf{x} = \mathbf{0}$, and let $\mathbf{x}_0 + W$ denote the set of all vectors that result by adding $\mathbf{x}_0$ to each vector in W. We must show that if $\mathbf{x}$ is a vector in $\mathbf{x}_0 + W$, then $\mathbf{x}$ is a solution of $A\mathbf{x} = \mathbf{b}$, and conversely that every solution of $A\mathbf{x} = \mathbf{b}$ is in the set $\mathbf{x}_0 + W$.

Assume first that $\mathbf{x}$ is a vector in $\mathbf{x}_0 + W$. This implies that $\mathbf{x}$ is expressible in the form $\mathbf{x} = \mathbf{x}_0 + \mathbf{w}$, where $A\mathbf{x}_0 = \mathbf{b}$ and $A\mathbf{w} = \mathbf{0}$. Thus,

$$A\mathbf{x} = A(\mathbf{x}_0 + \mathbf{w}) = A\mathbf{x}_0 + A\mathbf{w} = \mathbf{b} + \mathbf{0} = \mathbf{b},$$

which shows that $\mathbf{x}$ is a solution of $A\mathbf{x} = \mathbf{b}$.

Conversely, let $\mathbf{x}$ be any solution of $A\mathbf{x} = \mathbf{b}$. To show that $\mathbf{x}$ is in the set $\mathbf{x}_0 + W$ we must show that $\mathbf{x}$ is expressible in the form

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{w}$$

where $\mathbf{w}$ is in W (i.e., $A\mathbf{w} = \mathbf{0}$). We can do this by taking $\mathbf{w} = \mathbf{x} - \mathbf{x}_0$. This vector obviously satisfies $\mathbf{x} = \mathbf{x}_0 + \mathbf{w}$, and it is in W since

$$A\mathbf{w} = A(\mathbf{x} - \mathbf{x}_0) = A\mathbf{x} - A\mathbf{x}_0 = \mathbf{b} - \mathbf{b} = \mathbf{0}. \qquad \square$$

3.5 Cross Product

Definition 3.5.1. If $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$ are vectors in 3-space, then the cross product $\mathbf{u} \times \mathbf{v}$ is the vector defined by

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

or, in determinant notation,

$$\mathbf{u} \times \mathbf{v} = \left(\begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix}, - \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix}, \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \right).$$

Example 1. Find $\mathbf{u} \times \mathbf{v}$, where $\mathbf{u} = (1, 2, -2)$ and $\mathbf{v} = (3, 0, 1)$.

Theorem 3.5.1 (Relationships Involving Cross Product and Dot Product)**.** *If* $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *are vectors in* 3*-space, then*

(*a*) $\mathbf{u} \cdot (\mathbf{u} \times \mathbf{v}) = 0$
(*b*) $\mathbf{v} \cdot (\mathbf{u} \times \mathbf{v}) = 0$
(*c*) $\|\mathbf{u} \times \mathbf{v}\|^2 = \|\mathbf{u}\|^2\|\mathbf{v}\|^2 - (\mathbf{u} \cdot \mathbf{v})^2$
(*d*) $\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{u} \cdot \mathbf{v})\mathbf{w}$
(*e*) $(\mathbf{u} \times \mathbf{v}) \times \mathbf{w} = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{v} \cdot \mathbf{w})\mathbf{u}$

Example 2. Consider the vectors

$$\mathbf{u} = (1, 2, -2) \quad \text{and} \quad \mathbf{v} = (3, 0, 1).$$

Verify that $\mathbf{u} \times \mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$.

Theorem 3.5.2 (Properties of Cross Product)**.** *If* $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *are vectors in 3-space and* k *is any scalar, then*

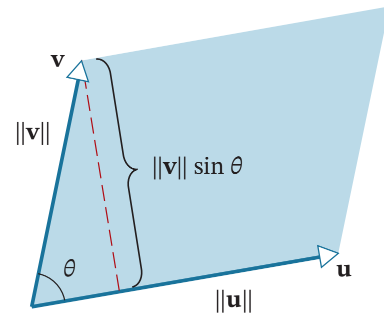
(a) $\mathbf{u} \times \mathbf{v} = -(\mathbf{v} \times \mathbf{u})$
(b) $\mathbf{u} \times (\mathbf{v} + \mathbf{w}) = (\mathbf{u} \times \mathbf{v}) + (\mathbf{u} \times \mathbf{w})$
(c) $(\mathbf{u} + \mathbf{v}) \times \mathbf{w} = (\mathbf{u} \times \mathbf{w}) + (\mathbf{v} \times \mathbf{w})$
(d) $k(\mathbf{u} \times \mathbf{v}) = (k\mathbf{u}) \times \mathbf{v} = \mathbf{u} \times (k\mathbf{v})$
(e) $\mathbf{u} \times \mathbf{0} = \mathbf{0} \times \mathbf{u} = \mathbf{0}$
(f) $\mathbf{u} \times \mathbf{u} = \mathbf{0}$

Example 3. Compute $\mathbf{i} \times \mathbf{j}$.

Theorem 3.5.3. *If* $\mathbf{u}$ *and* $\mathbf{v}$ *are vectors in* 3*-space, then* $\|\mathbf{u} \times \mathbf{v}\|$ *is equal to the area of the parallelogram determined by* $\mathbf{u}$ *and* $\mathbf{v}$*.*

Proof. If θ denotes the angle between $\mathbf{u}$ and $\mathbf{v}$, then

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



Since $0 \leq \theta \leq \pi$, it follows that $\sin\theta \geq 0$, so this can be rewritten as

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

But $\|\mathbf{v}\| \sin\theta$ is the altitude of the parallelogram determined by $\mathbf{u}$ and $\mathbf{v}$ (see the figure). Thus the area A of this parallelogram is given by

$$A = (\text{base})(\text{altitude}) = \|\mathbf{u}\|\|\mathbf{v}\| \sin\theta = \|\mathbf{u} \times \mathbf{v}\|. \qquad \square$$

Example 4. Find the area of the triangle determined by the points $P_1(2, 2, 0)$, $P_2(-1, 0, 2)$, and $P_3(0, 4, 3)$.

Definition 3.5.2. If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in 3-space, then

$$\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$$

is called the scalar triple product of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$.

Remark 1. The scalar triple product of $\mathbf{u} = (u_1, u_2, u_3)$, $\mathbf{v} = (v_1, v_2, v_3)$, and $\mathbf{w} = (w_1, w_2, w_3)$ can be calculated from the formula

$$\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}$$

since

$$\begin{aligned} \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) &= \mathbf{u} \cdot \left(\begin{vmatrix} v_2 & v_3 \\ w_2 & w_3 \end{vmatrix} \mathbf{i} - \begin{vmatrix} v_1 & v_3 \\ w_1 & w_3 \end{vmatrix} \mathbf{j} + \begin{vmatrix} v_1 & v_2 \\ w_1 & w_2 \end{vmatrix} \mathbf{k} \right) \\ &= \begin{vmatrix} v_2 & v_3 \\ w_2 & w_3 \end{vmatrix} u_1 - \begin{vmatrix} v_1 & v_3 \\ w_1 & w_3 \end{vmatrix} u_2 + \begin{vmatrix} v_1 & v_2 \\ w_1 & w_2 \end{vmatrix} u_3 \\ &= \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}. \end{aligned}$$

Example 5. Calculate the scalar triple product $\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})$ of the vectors

$$\mathbf{u} = 3\mathbf{i} - 2\mathbf{j} - 5\mathbf{k}, \quad \mathbf{v} = \mathbf{i} + 4\mathbf{j} - 4\mathbf{k}, \quad \mathbf{w} = 3\mathbf{j} + 2\mathbf{k}.$$

Theorem 3.5.4.

(*a*) *The absolute value of the determinant*

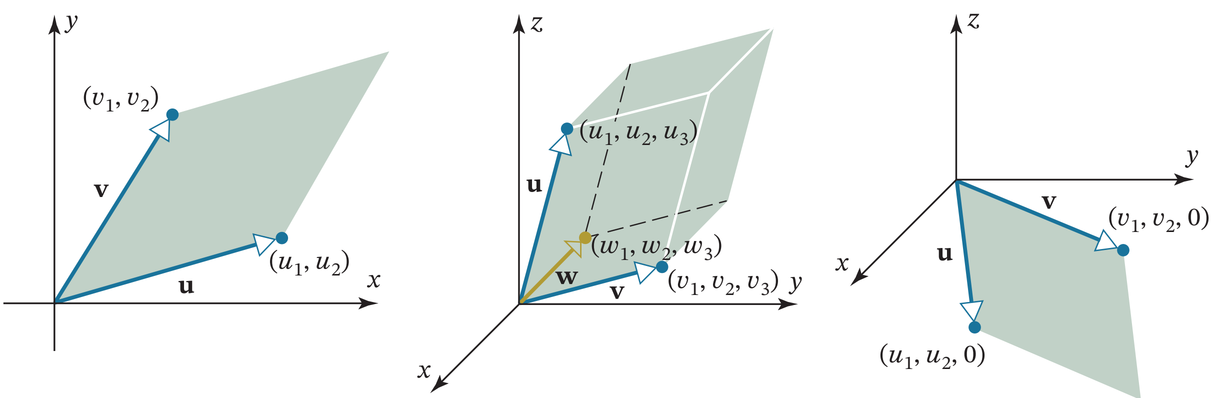
$$\det \begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix}$$

is equal to the area of the parallelogram in 2*-space determined by the vectors* $\mathbf{u} = (u_1, u_2)$ *and* $\mathbf{v} = (v_1, v_2)$. *(See Figure a.)*

(*b*) *The absolute value of the determinant*

$$\det \begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix}$$

is equal to the volume of the parallelepiped in 3*-space determined by the vectors* $\mathbf{u} = (u_1, u_2, u_3)$, $\mathbf{v} = (v_1, v_2, v_3)$, *and* $\mathbf{w} = (w_1, w_2, w_3)$. *(See Figure b.)*



(a) **(b)** **(c)**

Proof. (a) We will view **u** and **v** as vectors in the xy-plane of an xyz-coordinate system (Figure c), in which case these vectors are expressed as $\mathbf{u} = (u_1, u_2, 0)$ and $\mathbf{v} = (v_1, v_2, 0)$. Thus

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & 0 \\ v_1 & v_2 & 0 \end{vmatrix} = \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \mathbf{k} = \det \begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix} \mathbf{k}.$$

It follows from Theorem 3.5.3 and the fact that $\|k\| = 1$ that the area A of the parallelogram determined by **u** and **v** is

$$A = \|\mathbf{u} \times \mathbf{v}\| = \left\| \det \begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix} \mathbf{k} \right\| = \left| \det \begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix} \right| \|\mathbf{k}\| = \left| \det \begin{bmatrix} u_1 & u_2 \\ v_1 & v_2 \end{bmatrix} \right|.$$

(b) Take the base of the parallelepiped determined by **u**, **v**, and **w** to be the parallelogram determined by **v** and **w**. The area of the base is $\|\mathbf{v} \times \mathbf{w}\|$ and the height h of the parallelepiped is the length of the orthogonal projection of **u** on $\mathbf{v} \times \mathbf{w}$. Therefore,

$$h = \| \operatorname{proj}_{\mathbf{v}\times\mathbf{w}} \mathbf{u}\| = \frac{|\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})|}{\|\mathbf{v} \times \mathbf{w}\|}.$$

It follows that the volume V of the parallelepiped is

$$V = (\text{area of base}) \cdot \text{height} = \|\mathbf{v} \times \mathbf{w}\| \frac{|\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})|}{\|\mathbf{v} \times \mathbf{w}\|} = |\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w})|,$$

and so

$$V = \left| \det \begin{bmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{bmatrix} \right|. \qquad \square$$

Theorem 3.5.5. *If the vectors* $\mathbf{u} = (u_1, u_2, u_3)$*,* $\mathbf{v} = (v_1, v_2, v_3)$*, and* $\mathbf{w} = (w_1, w_2, w_3)$ *have the same initial point, then they lie in the same plane if and only if*

$$\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix} = 0.$$

Chapter 4

General Vector Spaces

4.1 Real Vector Spaces

Definition 4.1.1. Let V be an arbitrary nonempty set of objects on which two operations are defined: addition, and multiplication by numbers called scalars. By addition we mean a rule for associating with each pair of objects $\mathbf{u}$ and $\mathbf{v}$ in V an object $\mathbf{u}+\mathbf{v}$, called the sum of $\mathbf{u}$ and $\mathbf{v}$; by scalar multiplication we mean a rule for associating with each scalar k and each object $\mathbf{u}$ in V an object $k\mathbf{u}$ called the scalar multiple of $\mathbf{u}$ by k. If the following axioms are satisfied by all objects $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ in V and all scalars k and m, then we call V a vector space and we call the objects in V vectors.

1. If $\mathbf{u}$ and $\mathbf{v}$ are objects in V, then $\mathbf{u}+\mathbf{v}$ is in V.
2. $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$
3. $\mathbf{u}+(\mathbf{v}+\mathbf{w})=(\mathbf{u}+\mathbf{v})+\mathbf{w}$
4. There is an object $\mathbf{0}$ in V, called a zero vector for V, such that $\mathbf{0}+\mathbf{u}=\mathbf{u}+\mathbf{0}=\mathbf{u}$ for all $\mathbf{u}$ in V.
5. For each $\mathbf{u}$ in V, there is an object $-\mathbf{u}$ in V, called a negative of $\mathbf{u}$, such that $\mathbf{u}+(-\mathbf{u})=(-\mathbf{u})+\mathbf{u}=\mathbf{0}$.
6. If k is any scalar and $\mathbf{u}$ is any object in V, then $k\mathbf{u}$ is in V.
7. $k(\mathbf{u}+\mathbf{v})=k\mathbf{u}+k\mathbf{v}$
8. $(k+m)\mathbf{u}=k\mathbf{u}+m\mathbf{u}$
9. $k(m\mathbf{u})=(km)(\mathbf{u})$
10. $1\mathbf{u}=\mathbf{u}$

Example 1. Let V consist of a single object, which we denote by $\mathbf{0}$, and define

$$\mathbf{0}+\mathbf{0}=\mathbf{0} \quad \text{and} \quad k\mathbf{0}=\mathbf{0}$$

for all scalars k. Check that all the vector space axioms are satisfied.

Example 2. Let $V = R^n$, and define the vector space operations on V to be the usual operations of addition and scalar multiplication of n-tuples, that is,

$$\begin{aligned} \mathbf{u}+\mathbf{v} &= (u_1, u_2, \ldots, u_n) + (v_1, v_2, \ldots, v_n) = (u_1+v_1, u_2+v_2, \ldots, u_n+v_n) \\ k\mathbf{u} &= (ku_1, ku_2, \ldots, ku_n). \end{aligned}$$

Check that all the vector space axioms are satisfied.

Example 3. Let V consist of objects of the form

$$\mathbf{u} = (u_1, u_2, \ldots, u_n, \ldots)$$

in which $u_1, u_2, \ldots, u_n, \ldots$ is an infinite sequence of real numbers. We define two infinite sequences to be *equal* if their corresponding components are equal, and we define addition and scalar multiplication componentwise by

$$\begin{aligned}
\mathbf{u} + \mathbf{v} &= (u_1, u_2, \ldots, u_n, \ldots) + (v_1, v_2, \ldots, v_n, \ldots) \\
&= (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n, \ldots) \\
k\mathbf{u} &= (ku_1, ku_2, \ldots, ku_n, \ldots).
\end{aligned}$$

Confirm that V with these operations is a vector space.

Example 4. Let V be the set of 2×2 matrices with real entries, and take the vector space operations on V to be the usual operations of matrix addition and scalar multiplication; that is,

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} + \begin{bmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{bmatrix} = \begin{bmatrix} u_{11} + v_{11} & u_{12} + v_{12} \\ u_{21} + v_{21} & u_{22} + v_{22} \end{bmatrix}$$

$$k\mathbf{u} = k \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} = \begin{bmatrix} ku_{11} & ku_{12} \\ ku_{21} & ku_{22} \end{bmatrix}.$$

Confirm that V with these operations is a vector space.

Example 5. Confirm that the set V of all $m \times n$ matrices with the usual matrix operations of addition and scalar multiplication is a vector space.

Example 6. Let V be the set of real-valued functions that are defined at each x in the interval $(-\infty, \infty)$. If $\mathbf{f} = f(x)$ and $\mathbf{g} = g(x)$ are two functions in V and if k is any scalar, then define the operations of addition and scalar multiplication by

$$(\mathbf{f} + \mathbf{g})(x) = f(x) + g(x)$$
$$(k\mathbf{f})(x) = kf(x).$$

Confirm that V with these operations is a vector space.

Example 7. Let $V = R^2$ and define addition and scalar multiplication as follows: If $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$, then define

$$\mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2)$$

and if k is any real number, then define

$$k\mathbf{u} = (ku_1, 0).$$

Show that V is not a vector space.

Example 8. Let V be the set of positive real numbers, let $\mathbf{u} = u$ and $\mathbf{v} = v$ be any vectors (i.e., positive real numbers) in V, and let k be any scalar. Define the operations on V to be

$$\begin{aligned} u + v &= uv \\ ku &= u^k \end{aligned}$$

Confirm that V with these operations is a vector space.

Theorem 4.1.1. *Let V be a vector space, $\mathbf{u}$ a vector in V, and k a scalar, then:*

(a) $0\mathbf{u} = \mathbf{0}$
(b) $k\mathbf{0} = \mathbf{0}$
(c) $(-1)\mathbf{u} = -\mathbf{u}$
(d) *If $k\mathbf{u} = \mathbf{0}$, then $k = 0$ or $\mathbf{u} = \mathbf{0}$.*

4.2 Subspaces

Definition 4.2.1. A subset W of a vector space V is called a subspace of V if W is itself a vector space under the addition and scalar multiplication defined on V.

Theorem 4.2.1 (Subspace Test)**.** *If W is a set of one or more vectors in a vector space V, then W is a subspace of V if and only if the following conditions are satisfied.*
(a) *If $\mathbf{u}$ and $\mathbf{v}$ are vectors in W, then $\mathbf{u}+\mathbf{v}$ is in W.*
(b) *If k is a scalar and $\mathbf{u}$ is a vector in W, then $k\mathbf{u}$ is in W.*

Proof. If W is a subspace of V, then all the vector space axioms hold in W, including Axioms 1 and 6, which are precisely conditions (a) and (b).

Conversely, assume that conditions (a) and (b) hold. Since these are Axioms 1 and 6, and since Axioms 2, 3, 7, 8, 9, and 10 are inherited from V, we only need to show that Axioms 4 and 5 hold in W. For this purpose, let $\mathbf{u}$ be any vector in W. It follows from condition (b) that $k\mathbf{u}$ is a vector in W for every scalar k. In particular, $0\mathbf{u} = \mathbf{0}$ and $(-1)\mathbf{u} = -\mathbf{u}$ are in W, which shows that Axioms 4 and 5 hold in W. □

Example 1. If V is any vector space, show that the subset $W = \{\mathbf{0}\}$ of V consisting of the zero vector only is a subspace of V, called the zero subspace of V.

Example 2. Show that lines through the origin are subspaces of R^2 and of R^3.

Example 3. Show that planes through the origin are subspaces of R^3.

Example 4. Let W be the set of all points (x, y) in R^2 for which $x \geq 0$ and $y \geq 0$. Show that this set is not a subspace of R^2.

Example 5. Show that the set of symmetric $n \times n$ matrices is a subspace of M_{nn}.

Example 6. Show that the set of invertible $n \times n$ matrices is not a subspace of M_{nn}.

Example 7. Show that the set of continuous functions on $(-\infty, \infty)$, denoted by $C(-\infty, \infty)$, is a subspace of $F(-\infty, \infty)$.

Example 8. Show that the set of functions with m continuous derivatives on $(-\infty, \infty)$ and the set of functions with derivatives of all orders $(-\infty, \infty)$ are subspaces of $F(-\infty, \infty)$, denoted by $C^m(-\infty, \infty)$ and $C^\infty(-\infty, \infty)$, respectively.

Example 9. Show that the set of all polynomials is a subspace of $F(-\infty, \infty)$, denoted by P_∞.

Example 10. Show that the set of polynomials with positive degree n is not a subspace of $F(-\infty, \infty)$, but that for each non-negative integer n the polynomials of degree n *or less* form a subspace of $F(-\infty, \infty)$, denoted by P_n.

Example 11. Determine whether the indicated set of matrices is a subspace of M_{22}.

(a) The set U consisting of all matrices of the form

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

(b) The set W consisting of all 2×2 matrices A such that

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

Example 12. Determine whether the indicated set of polynomials is a subspace of P_2.

(a) The set U consisting of all polynomials of the form $\mathbf{p} = 1 + ax - ax^2$, where a is a real number.

(b) The set W consisting of all polynomials $\mathbf{p}$ in P_2 such that $\mathbf{p}(2) = 0$.

Theorem 4.2.2. *If $W_1, W_2, \ldots, W_r$ are subspaces of a vector space V, then the intersection of these subspaces is also a subspace of V.*

Proof. Let W be the intersection of the subspaces $W_1, W_2, \ldots, W_r$. This set is not empty because each of these subspaces contains the zero vector of V, and hence so does their intersection. Thus, it remains to show that W is closed under addition and scalar multiplication.

To prove closure under addition, let $\mathbf{u}$ and $\mathbf{v}$ be vectors in W. Since W is the intersection of $W_1, W_2, \ldots, W_r$, it follows that $\mathbf{u}$ and $\mathbf{v}$ also lie in each of these subspaces. Moreover, since these subspaces are closed under addition and scalar multiplication, they also all contain the vectors $\mathbf{u} + \mathbf{v}$ and $k\mathbf{u}$ for every scalar k, and hence so does their intersection W. $\square$

Theorem 4.2.3. *The solution set of a homogeneous linear system $A\mathbf{x} = \mathbf{0}$ of m equations in n unknowns is a subspace of R^n.*

Proof. Let W be the solution set of the system. The set W is not empty because it contains at least the trivial solution $\mathbf{x} = \mathbf{0}$.

To show that W is a subspace of R^n, we must show that it is closed under addition and scalar multiplication. To do this, let $\mathbf{x}_1$ and $\mathbf{x}_2$ be vectors in W. Since these vectors are solutions of $A\mathbf{x} = \mathbf{0}$, we have

$$A\mathbf{x}_1 = \mathbf{0} \quad \text{and} \quad A\mathbf{x}_2 = \mathbf{0}.$$

It follows from these equations and the distributive property of matrix multiplication that

$$A(\mathbf{x}_1 + \mathbf{x}_2) = A\mathbf{x}_1 + A\mathbf{x}_2 = \mathbf{0} + \mathbf{0} = \mathbf{0},$$

so W is closed under addition. Similarly, if k is any scalar then

$$A(k\mathbf{x}_1) = kA\mathbf{x}_1 = k\mathbf{0} = \mathbf{0},$$

so W is also closed under scalar multiplication. $\square$

Example 13. In each part the solution of the linear system is provided. Give a geometric description of the solution set.

(a) $$\begin{bmatrix} 1 & -2 & 3 \\ 2 & -4 & 6 \\ 3 & -6 & 9 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

(b) $$\begin{bmatrix} 1 & -2 & 3 \\ -3 & 7 & -8 \\ -2 & 4 & -6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

(c) $$\begin{bmatrix} 1 & -2 & 3 \\ -3 & 7 & -8 \\ 4 & 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

(d) $$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Theorem 4.2.4. *If A is an $m \times n$ matrix, then the kernel of the matrix transformation $T_A : R^n \to R^m$, the set of vectors in R^n that T_A maps into the zero vector in R^m, is a subspace of R^n.*

4.3 Spanning Sets

Definition 4.3.1. If $\mathbf{w}$ is a vector in a vector space V, then $\mathbf{w}$ is said to be a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ in V if $\mathbf{w}$ can be expressed in the form

$$\mathbf{w} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_r\mathbf{v}_r$$

where $k_1, k_2, \ldots, k_r$ are scalars. These scalars are called the coefficients of the linear combination.

Theorem 4.3.1. *If $S = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r\}$ is a nonempty set of vectors in a vector space V, then:*

(a) *The set W of all possible linear combinations of the vectors in S is a subspace of V.*

(b) *The set W in part* (a) is the "smallest" subspace of V that contains all of the vectors in S in the sense that any other subspace that contains those vectors contains W.

Proof. (a) Let W be the set of all possible linear combinations of the vectors in S. We must show that W is closed under addition and scalar multiplication. To prove closure under addition, let

$$\mathbf{u} = c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + \cdots + c_r\mathbf{w}_r \quad \text{and} \quad \mathbf{v} = k_1\mathbf{w}_1 + k_2\mathbf{w}_2 + \cdots + k_r\mathbf{w}_r$$

be two vectors in W. It follows that their sum can be written as

$$\mathbf{u} + \mathbf{v} = (c_1 + k_1)\mathbf{w}_1 + (c_2 + k_2)\mathbf{w}_2 + \cdots + (c_r + k_r)\mathbf{w}_r,$$

which is a linear combination of the vectors in S. Similarly, if a is any scalar, then

$$a\mathbf{u} = (ac_1)\mathbf{w}_1 + (ac_2)\mathbf{w}_2 + \cdots + (ac_r)\mathbf{w}_r,$$

which is a linear combination of the vectors in S.
(b) Let W' be any subspace of V that contains all of the vectors in S. Since W' is closed under addition and scalar multiplication, it contains all linear combinations of the vectors in S and hence contains W. □

Remark 1. The subspace W in Theorem 4.3.1 is called the subspace of V spanned by S. The vectors $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r$ in S are said to span W, and we write

$$W = \operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r\} \quad \text{or} \quad W = \operatorname{span}(S).$$

Example 1. Show that the standard unit vectors span R^n.

Example 2.

(a) If $\mathbf{v}$ is a nonzero vector in R^2 or R^3 that has its initial point at the origin, what is a geometric description of span$\{\mathbf{v}\}$?

(b) If $\mathbf{v}_1$ and $\mathbf{v}_2$ are nonzero vectors in R^3 that have their initial points at the origin, what is a geometric description of span$\{\mathbf{v}_1, \mathbf{v}_2\}$?

Example 3. Show that the polynomials $1, x, x^2, \ldots, x^n$ span the vector space P_n.

Example 4. Consider the vectors $\mathbf{u} = (1, 2, -1)$ and $\mathbf{v} = (6, 4, 2)$ in R^3. Show that $\mathbf{w} = (9, 2, 7)$ is a linear combination of $\mathbf{u}$ and $\mathbf{v}$ and that $\mathbf{w}' = (4, -1, 8)$ is *not* a linear combination of $\mathbf{u}$ and $\mathbf{v}$.

Example 5. Determine whether the vectors $\mathbf{v}_1 = (1, 1, 2)$, $\mathbf{v}_2 = (1, 0, 1)$, and $\mathbf{v}_3 = (2, 1, 3)$ span the vector space R^3.

Example 6. Determine whether the set S spans P_2.

(a) $S = \{1 + x + x^2, -1 - x, 2 + 2x + x^2\}$

(b) $S = \{x + x^2, x - x^2, 1 + x, 1 - x\}$

Example 7. In each part, determine whether the set S spans M_{22}.

(a) $S = \left\{ \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \right\}$

(b) $S = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix} \right\}$

Theorem 4.3.2. *If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ and $S' = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k\}$ are nonempty sets of vectors in a vector space V, then*

$$\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\} = \operatorname{span}\{\mathbf{w}_1 . \mathbf{w}_2, \ldots, \mathbf{w}_k\}$$

if and only if each vector in S is a linear combination of those in S', and each vector in S' is a linear combination of those in S.

4.4 Linear Independence

Definition 4.4.1. If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is a set of two or more vectors in a vector space V, then S is said to be a linearly independent set if no vector in S can be expressed as a linear combination of the others. A set that is not linearly independent is said to be linearly dependent.

Theorem 4.4.1. *A nonempty set $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ in a vector space V is linearly independent if and only if the only coefficients satisfying the vector equation*

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_r\mathbf{v}_r = \mathbf{0}$$

are $k_1 = 0, k_2 = 0, \ldots, k_r = 0$.

Example 1. Show that the standard unit vectors in R^n are linearly independent.

Example 2. Determine whether the vectors

$$\mathbf{v}_1 = (1, -2, 3), \quad \mathbf{v}_2 = (5, 6, -1), \quad \mathbf{v}_3 = (3, 2, 1)$$

are linearly independent or linearly dependent in R^3.

Example 3. Determine whether the vectors

$$\mathbf{v}_1 = (1, 2, 2, -1), \quad \mathbf{v}_2 = (4, 9, 9, -4), \quad \mathbf{v}_3 = (5, 8, 9, -5)$$

in R^4 are linearly independent or linearly dependent.

Example 4. Show that the polynomials

$$1, \quad x, \quad x^2, \ldots, \quad x^n$$

form a linearly independent set in P_n.

Example 5. Determine whether the polynomials

$$\mathbf{p}_1 = 1 - x, \quad \mathbf{p}_2 = 5 + 3x - 2x^2, \quad \mathbf{p}_3 = 1 + 3x - x^2$$

are linearly dependent or linearly independent in P_2.

Theorem 4.4.2.

(a) *A set with finitely many vectors that contains* $\mathbf{0}$ *is linearly dependent.*
(b) *A set with exactly two vectors is linearly independent if and only if neither vector is a scalar multiple of the other.*

Example 6. Determine whether the functions $\mathbf{f}_1 = x$ and $\mathbf{f}_2 = \sin x$ are linearly independent in $F(-\infty, \infty)$, and whether the functions $\mathbf{g}_1 = \sin 2x$ and $\mathbf{g}_2 = \sin x \cos x$ are linearly independent in $F(-\infty, \infty)$.

Theorem 4.4.3. *Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ be a set of vectors in R^n. If $r > n$, then S is linearly dependent.*

Proof. Suppose that

$$\begin{aligned}
\mathbf{v}_1 &= (v_{11}, v_{12}, \ldots, v_{1n}) \\
\mathbf{v}_2 &= (v_{21}, v_{22}, \ldots, v_{2n}) \\
\vdots \quad & \qquad\quad \vdots \\
\mathbf{v}_r &= (v_{r1}, v_{r2}, \ldots, v_{rn})
\end{aligned}$$

and consider the equation

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_r\mathbf{v}_r = \mathbf{0}.$$

If we express both sides of this equation in terms of components and then equate the corresponding components, we obtain the system

$$\begin{aligned}
v_{11}k_1 + v_{22}k_2 + \cdots + v_{r1}k_r &= 0 \\
v_{12}k_1 + v_{22}k_2 + \cdots + v_{r2}k_r &= 0 \\
\vdots \qquad \vdots \qquad\qquad \vdots \quad &\quad \vdots \\
v_{1n}k_1 + v_{2n}k_2 + \cdots + v_{rn}k_r &= 0.
\end{aligned}$$

This is a homogeneous system of n equations in the r unknowns $k_1, \ldots, k_r$. Since $r > n$, the system has nontrivial solutions. Therefore, $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is a linearly dependent set. □

Example 7. It is an important fact that the nonzero row vectors of a matrix in row echelon or reduced row echelon form are linearly independent. To suggest how a general proof might go, show that the row vectors of the matrix

$$R = \begin{bmatrix} 1 & a_{12} & a_{13} & a_{14} \\ 0 & 1 & a_{23} & a_{24} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

are linearly independent.

Definition 4.4.2. If $\mathbf{f}_1 = f_1(x), \mathbf{f}_2 = f_2(x), \ldots, \mathbf{f}_n = f_n(x)$ are functions that are $n-1$ times differentiable on the interval $(-\infty, \infty)$, then the determinant

$$W(x) = \begin{vmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f_1'(x) & f_2'(x) & \cdots & f_n'(x) \\ \vdots & \vdots & & \vdots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \cdots & f_n^{(n-1)}(x) \end{vmatrix}$$

is called the Wronskian of $f_1, f_2, \ldots, f_n$.

Theorem 4.4.4. *If the functions $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n$ have $n-1$ continuous derivatives on the interval $(-\infty, \infty)$, and if the Wronskian of these functions is not identically zero on $(-\infty, \infty)$, then these functions form a linearly independent set of vectors in $C^{(n-1)}(-\infty, \infty)$.*

Example 8. Use the Wronskian to show that $\mathbf{f}_1 = x$ and $\mathbf{f}_2 = \sin x$ are linearly independent vectors in $C^\infty(-\infty, \infty)$.

Example 9. Use the Wronskian to show that $\mathbf{f}_1 = 1$, $\mathbf{f}_2 = e^x$, and $\mathbf{f}_3 = e^{2x}$ are linearly independent vectors in $C^\infty(-\infty, \infty)$.

4.5 Coordinates and Basis

Remark 1. A vector space V is said to be finite-dimensional if there is a finite set of vectors in V that spans V and is said to be infinite-dimensional if no such set exists.

Definition 4.5.1. If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a set of vectors in a finite-dimensional vector space V, then S is called a basis for V if:

(a) S spans V.
(b) S is linearly independent.

Example 1. Show that the standard unit vectors form a basis for R^n called the standard basis for R^n.

Example 2. Show that $S = \{1, x, x^2, \ldots, x^n\}$ is a basis for the vector space P_n of polynomials of degree n or less.

Example 3. Show that the vectors $\mathbf{v}_1 = (1, 2, 1)$, $\mathbf{v}_2 = (2, 9, 0)$, and $\mathbf{v}_3 = (3, 3, 4)$ form a basis for R^3.

Example 4. Show that the matrices

$$M_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad M_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad M_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad M_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

form a basis for the vector space M_{22} of 2×2 matrices.

Example 5. Show that the vector space P_∞ of all polynomials with real coefficients is infinite-dimensional by showing that it has no finite spanning set.

Example 6. Which of the vector spaces in Examples 1-5 are finite-dimensional, and which are infinite-dimensional?

Theorem 4.5.1 (Uniqueness of Basis Representation)**.** *If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for a vector space V, then every vector $\mathbf{v}$ in V can be expressed in the form $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$ in exactly one way.*

Proof. Since S spans V, it follows from the definition of a spanning set that every vector in V is expressible as a linear combination of the vectors in S. To see that there is only *one* way to express a vector as a linear combination of the vectors in S, suppose that some vector $\mathbf{v}$ can be written as

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

and also as

$$\mathbf{v} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_n\mathbf{v}_n.$$

Subtracting the second equation from the first gives

$$\mathbf{0} = (c_1 - k_1)\mathbf{v}_1 + (c_2 - k_2)\mathbf{v}_2 + \cdots + (c_n - k_n)\mathbf{v}_n.$$

Since the right side of this equation is a linear combination of vectors in S, the linear independence of S implies that

$$c_1 - k_1 = 0, \quad c_2 - k_2 = 0, \ldots, \quad c_n - k_n = 0,$$

that is,

$$c_1 = k_1, \quad c_2 = k_2, \ldots, \quad c_n = k_n. \qquad \square$$

Definition 4.5.2. If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for a vector space V, and

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$

is the expression for a vector $\mathbf{v}$ in terms of the basis S, then the scalars $c_1, c_2, \ldots, c_n$ are called the coordinates of $\mathbf{v}$ relative to the basis S. The vector $(c_1, c_2, \ldots, c_n)$ in R^n constructed from these coordinates is called the coordinate vector of $\mathbf{v}$ relative to S; it is denoted by

$$(\mathbf{v})_S = (c_1, c_2, \ldots, c_n).$$

Example 7. What is the coordinate vector $(\mathbf{v})_S$ where $V = R^n$ and S is the standard basis?

Example 8.

(a) Find the coordinate vector for the polynomial

$$\mathbf{p}(x) = c_0 + c_1x + c_2x^2 + \cdots + c_nx^n$$

relative to the standard basis for the vector space P_n.

(b) Find the coordinate vector of

$$B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

relative to the standard basis for M_{22}.

Example 9.

(a) We showed in Example 3 that the vectors

$$\mathbf{v}_1 = (1, 2, 1), \quad \mathbf{v}_2 = (2, 9, 0), \quad \mathbf{v}_3 = (3, 3, 4)$$

form a basis for R^3. Find the coordinate vector of $\mathbf{v} = (5, -1, 9)$ relative to the basis $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$.

(b) Find the vector $\mathbf{v}$ in R^3 whose coordinate vector relative to S is $(\mathbf{v})_S = (-1, 3, 2)$.

4.6 Dimension

Theorem 4.6.1. *All bases for a finite-dimensional vector space have the same number of vectors.*

Theorem 4.6.2. *Let V be an n-dimensional vector space, and let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be any basis.*
(a) *If a set in V has more than n vectors, then it is linearly dependent.*
(b) *If a set in V has fewer than n vectors, then it does not span V.*

Definition 4.6.1. The dimension of a finite-dimensional vector space V is denoted by $\dim(V)$ and is defined to be the number of vectors in a basis for V. In addition, the zero vector space is defined to have dimension zero.

Example 1. Find the dimensions of R^n, P_n, and M_{mn}.

Example 2. If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is a set of linearly independent vectors, what is $\dim[\text{span}(S)]$?

Example 3. Find a basis for and the dimension of the solution space of the homogeneous system

$$\begin{aligned} x_1 + 3x_2 - 2x_3 + 2x_5 &= 0 \\ 2x_1 + 6x_2 - 5x_3 - 2x_4 + 4x_5 - 3x_6 &= 0 \\ 5x_3 + 10x_4 + 15x_6 &= 0 \\ 2x_1 + 6x_2 + 8x_4 + 4x_5 + 18x_6 &= 0. \end{aligned}$$

Theorem 4.6.3 (Plus/Minus Theorem)**.** *Let S be a nonempty set of vectors in a vector space V.*

(*a*) *If S is a linearly independent set, and if $\mathbf{v}$ is a vector in V that is outside of* span(S)*, then the set $S \cup \{\mathbf{v}\}$ that results by inserting $\mathbf{v}$ into S is still linearly independent.*

(*b*) *If $\mathbf{v}$ is a vector in S that is expressible as a linear combination of other vectors in S, and if $S - \{\mathbf{v}\}$ denotes the set obtained by removing $\mathbf{v}$ from S, then S and $S - \{\mathbf{v}\}$ span the same space; that is,*

$$\operatorname{span}(S) = \operatorname{span}(S - \{\mathbf{v}\}).$$

Example 4. Show that $\mathbf{p}_1 = 1 - x^2$, $\mathbf{p}_2 = 2 - x^2$, and $\mathbf{p}_3 = x^3$ are linearly independent vectors.

Theorem 4.6.4. *Let V be an n-dimensional vector space, and let S be a set in V with exactly n vectors. Then S is a basis for V if and only if S spans V or S is linearly independent.*

Proof. Assume that S has exactly n vectors and spans V. To prove that S is a basis, we must show that S is a linearly independent set. But if this is not so, them some vector $\mathbf{v}$ in S is a linear combination of the remaining vectors. If we remove this vector from S, then it follows that the remaining set of $n-1$ vectors still spans V. But this is impossible since no set with fewer than n vectors can span an n-dimensional vector space. Thus S is linearly independent.

Assume that S has exactly n vectors and is a linearly independent set. To prove that S is a basis, we must show that S spans V. But if this is not so, then there is some vector $\mathbf{v}$ in V that is not in span(S). If we insert this vector into S, then this set of $n + 1$ vectors is still linearly independent. But this is impossible, since no set with more than n vectors in an n-dimensional vector space can be linearly independent. Thus S spans V. □

Example 5.

(a) Explain why the vectors $\mathbf{v}_1 = (-3, 7)$ and $\mathbf{v}_2 = (5, 5)$ form a basis for R^2.

(b) Explain why the vectors $\mathbf{v}_1 = (2, 0, -1)$, $\mathbf{v}_2 = (4, 0, 7)$, and $\mathbf{v}_3 = (-1, 1, 4)$ form a basis for R^3.

Theorem 4.6.5. *Let S be a finite set of vectors in a finite-dimensional vector space V.*

(a) If S spans V but is not a basis for V, then S can be reduced to a basis for V by removing appropriate vectors from S.

(b) If S is a linearly independent set that is not already a basis for V, then S can be enlarged to a basis for V by inserting appropriate vectors into S.

Theorem 4.6.6. *If W is a subspace of a finite-dimensional vector space V, then:*

(a) W is finite-dimensional.

(b) $\dim(W) \leq \dim(V)$.

(c) $W = V$ if and only if $\dim(W) = \dim(V)$.

Proof. (a) Since V is finite-dimensional, there exists a finite set S spanning V. Since W is a subspace of V, $W \subset V$. Therefore S also spans W, so W is finite-dimensional.
(b) Part (a) shows that W is finite-dimensional, so it has a basis

$$S = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}.$$

Either S is also a basis for V or it is not. If so, then $\dim(V) = m$, which means that $\dim(V) = \dim(W)$. If not, then because S is a linearly independent set it can be enlarged to a basis for V. But this implies that $\dim(W) < \dim(V)$, so we have shown that $\dim(W) \leq \dim(V)$ in all cases.
(c) Assume that $\dim(W) = \dim(V)$ and that $S = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$ is a basis for W. If S is not also a basis for V, then being linearly independent S can be extended to a basis for V. But this would mean that $\dim(V) > \dim(W)$, which contradicts our hypothesis. Thus S must also be a basis for V, which means that $W = V$. The converse is obvious. $\square$

4.7 Change of Basis

Remark 1. If $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is a basis for a finite-dimensional vector space V, and if

$$(\mathbf{v})_S = (c_1, c_2, \dots, c_n)$$

is the coordinate vector of $\mathbf{v}$ relative to S, then the mapping

$$\mathbf{v} \to (\mathbf{v})_S$$

creates a connection (a one-to-one correspondence) between vectors in the *general* vector space V and vectors in the *Euclidean* vector space R^n. We call this the coordinate map relative to S from V to R^n.

Remark 2 (Solution of the Change-of-Basis Problem). If we change the basis for a vector space V from an old basis $B = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ to a new basis $B' = \{\mathbf{u}'_1, \mathbf{u}'_2, \dots, \mathbf{u}'_n\}$, then for each vector $\mathbf{v}$ in V, the new coordinate vector $[\mathbf{v}]_{B'}$ is related to the old coordinate vector $[\mathbf{v}]_B$ by the equation

$$[\mathbf{v}]_{B'} = P[\mathbf{v}]_B$$

where the columns of P are the coordinate vectors of the old basis vectors relative to the new basis; that is,

$$P = \left[[\mathbf{u}_1]_{B'} \mid [\mathbf{u}_2]_{B'} \mid \dots \mid [\mathbf{u}_n]_{B'}\right].$$

Example 1. Consider the bases $B = \{\mathbf{u}_1, \mathbf{u}_2\}$ and $B' = \{\mathbf{u}'_1, \mathbf{u}'_2\}$ for R^2, where

$$\mathbf{u}_1 = (1,0), \quad \mathbf{u}_2 = (0,1), \quad \mathbf{u}'_1 = (1,1), \quad \mathbf{u}'_2(2,1).$$

(a) Find the transition matrix $P_{B \to B'}$ from B to B'.

(b) Find the transition matrix $P_{B' \to B}$ from B' to B.

Example 2. Let B and B' be the bases in Example 1. Use an appropriate formula to find $[\mathbf{v}]_{B'}$ given that

$$[\mathbf{v}]_B = \begin{bmatrix} -3 \\ 5 \end{bmatrix}.$$

Theorem 4.7.1. *If P is the transition matrix from a basis B to a basis B' for a finite-dimensional vector space V, then P is invertible and P^{-1} is the transition matrix from B' to B.*

Remark 3 (A Procedure for Computing Transition Matrices).

Step 1. Form the partitioned matrix [**new basis | old basis**] in which the basis vectors are in column form.

Step 2. Use elementary row operations to reduce the matrix in Step 1 to reduced row echelon form.

Step 3. The resulting matrix will be [***I*** **| transition matrix from old to new**] where I is an identity matrix.

Step 4. Extract the matrix on the right side of the matrix obtained in Step 3.

Example 3. In Example 1 we considered the bases $B = \{\mathbf{u}_1, \mathbf{u}_2\}$ and $B' = \{\mathbf{u}'_1, \mathbf{u}'_2\}$ for R^2, where

$$\mathbf{u}_1 = (1,0), \quad \mathbf{u}_2 = (0,1), \quad \mathbf{u}'_1 = (1,1), \quad \mathbf{u}'_2 = (2,1).$$

(a) Use Remark 3 to find the transition matrix from B to B'.

(b) Use Remark 3 to find the transition matrix from B' to B.

Theorem 4.7.2. *Let $B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be any basis for R^n and let $S = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ be the standard basis for R^n. If the vectors in these bases are written in column form, then*

$$P_{B \to S} = [\mathbf{u}_1 \mid \mathbf{u}_2 \mid \cdots \mid \mathbf{u}_n].$$

4.8 Row Space, Column Space, and Null Space

Example 1. Let

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

What are the row and column vectors of A?

Definition 4.8.1. If A is an $m \times n$ matrix, then the subspace of R^n spanned by the row vectors of A is called the row space of A, and the subspace of R^n spanned by the column vectors of A is called the column space of A. The solution space of the homogeneous system of equations $A\mathbf{x} = \mathbf{0}$, which is a subspace of R^n, is called the null space of A.

Theorem 4.8.1. *A system of linear equations $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b}$ is in the column space of A.*

Example 2. Let $A\mathbf{x} = \mathbf{b}$ be the linear system

$$\begin{bmatrix} -1 & 3 & 2 \\ 1 & 2 & -3 \\ 2 & 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ -9 \\ -3 \end{bmatrix}.$$

Show that $\mathbf{b}$ is in the column space of A by expressing it as a linear combination of the column vectors of A.

Theorem 4.8.2. *If $\mathbf{x}_0$ is any solution of a consistent linear system $A\mathbf{x} = \mathbf{b}$, and if $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is a basis for the null space of A, then every solution of $A\mathbf{x} = \mathbf{b}$ can be expressed in the form*

$$\mathbf{x} = \mathbf{x}_0 + c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k.$$

Conversely, for all choices of scalars $c_1, c_2, \ldots, c_k$, the vector $\mathbf{x}$ in this formula is a solution of $A\mathbf{x} = \mathbf{b}$.

Proof. Let $\mathbf{x}_0$ be any solution of $A\mathbf{x} = \mathbf{b}$, let W denote the null space of $A\mathbf{x} = \mathbf{0}$, and let $\mathbf{x}_0 + W$ be the set of all vectors that result by adding $\mathbf{x}_0$ to each vector in W. Thus, the vectors in $\mathbf{x}_0 + W$ are those that are expressible in the form

$$\mathbf{x} = \mathbf{x}_0 + c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k.$$

We must show that if $\mathbf{x}$ is a vector in $\mathbf{x}_0 + W$, then $\mathbf{x}$ is a solution of $A\mathbf{x} = \mathbf{b}$, and conversely that every solution of $A\mathbf{x} = \mathbf{b}$ is in the set $\mathbf{x}_0 + W$.

Assume first that $\mathbf{x}$ is a vector in $\mathbf{x}_0 + W$. This implies that $\mathbf{x}$ is expressible in the form $\mathbf{x} = \mathbf{x}_0 + \mathbf{w}$, where $A\mathbf{x}_0 = \mathbf{b}$ and $A\mathbf{w} = \mathbf{0}$. Thus,

$$A\mathbf{x} = A(\mathbf{x}_0 + \mathbf{w}) = A\mathbf{x}_0 + A\mathbf{w} = \mathbf{b} + \mathbf{0} = \mathbf{b},$$

which shows that $\mathbf{x}$ is a solution of $A\mathbf{x} = \mathbf{b}$.

Conversely, let $\mathbf{x}$ be any solution of $A\mathbf{x} = \mathbf{b}$. To show that $\mathbf{x}$ is in the set $\mathbf{x}_0 + W$ we must show that $\mathbf{x}$ is expressible in the form

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{w}$$

where $\mathbf{w}$ is in W (i.e., $A\mathbf{w} = \mathbf{0}$). We can do this by taking $\mathbf{w} = \mathbf{x} - \mathbf{x}_0$. This vector obviously satisfies $\mathbf{x} = \mathbf{x}_0 + \mathbf{w}$, and it is in W since

$$A\mathbf{w} = A(\mathbf{x} - \mathbf{x}_0) = A\mathbf{x} - A\mathbf{x}_0 = \mathbf{b} - \mathbf{b} = \mathbf{0}. \qquad \square$$

Remark 1. The vector $\mathbf{x}_0$ in Theorem 4.7.2 is called a particular solution of $A\mathbf{x} = \mathbf{b}$, and the remaining part of the formula is called the general solution of $A\mathbf{x} = \mathbf{0}$.

Theorem 4.8.3.

(a) *Row equivalent matrices have the same row space.*
(b) *Row equivalent matrices have the same null space.*

Proof. (a) If A and B are row equivalent then each can be obtained from the other by elementary row operations. As these operations involve only scalar multiplication (multiply a row by a scalar) and linear combinations (add a scalar multiple of one row to another), it follows that the row space of each is a subspace of the other, so the two row spaces must be the same.
(b) If A and B are row equivalent then each can be obtained from the other by elementary row operations. But elementary row operations do not change the solution set of a linear system, so the solution sets of $A\mathbf{x} = \mathbf{0}$ and $B\mathbf{x} = \mathbf{0}$ must be the same. That is, A and B have the same null space. □

Theorem 4.8.4. *If a matrix R is in row echelon form, then the row vectors with the leading 1's (the nonzero row vectors) form a basis for the row space of R, and the column vectors with the leading 1's of the row vectors form a basis for the column space of R.*

Example 3. Find bases for the row and column spaces of the matrix

$$R = \begin{bmatrix} 1 & -2 & 5 & 0 & 3 \\ 0 & 1 & 3 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Example 4. Find a basis for the row space of the matrix

$$A = \begin{bmatrix} 1 & -3 & 4 & -2 & 5 & 4 \\ 2 & -6 & 9 & -1 & 8 & 2 \\ 2 & -6 & 9 & -1 & 9 & 7 \\ -1 & 3 & -4 & 2 & -5 & -4 \end{bmatrix}.$$

Theorem 4.8.5. *If A and B are row equivalent matrices, then:*

(a) *A given set of column vectors of A is linearly independent if and only if the corresponding column vectors of B are linearly independent.*

(b) *A given set of column vectors of A forms a basis for the column space of A if and only if the corresponding column vectors of B form a basis for the column space of B.*

Example 5. Find a basis for the column space of the matrix

$$A = \begin{bmatrix} 1 & -3 & 4 & -2 & 5 & 4 \\ 2 & -6 & 9 & -1 & 8 & 2 \\ 2 & -6 & 9 & -1 & 9 & 7 \\ -1 & 3 & -4 & 2 & -5 & -4 \end{bmatrix}.$$

Example 6. Find a basis for the row space of

$$A = \begin{bmatrix} 1 & -2 & 0 & 0 & 3 \\ 2 & -5 & -3 & -2 & 6 \\ 0 & 5 & 15 & 10 & 0 \\ 2 & 6 & 18 & 8 & 6 \end{bmatrix}$$

consisting entirely of row vectors from A.

Example 7. The following vectors span a subspace of R^4. Find a subset of these vectors that forms a basis of this subspace.

$$\begin{aligned} \mathbf{v}_1 &= (1, 2, 2, -1), & \mathbf{v}_2 &= (-3, -6, -6, 3), \\ \mathbf{v}_3 &= (4, 9, 9, -4), & \mathbf{v}_4 &= (-2, -1, -1, 2), \\ \mathbf{v}_5 &= (5, 8, 9, -5), & \mathbf{v}_6 &= (4, 2, 7, -4). \end{aligned}$$

Example 8.

(a) Find a subset of the vectors

$$\mathbf{v}_1 = (1, -2, 0, 3), \quad \mathbf{v}_2 = (2, -5, -3, 6),$$
$$\mathbf{v}_3 = (0, 1, 3, 0), \quad \mathbf{v}_4 = (2, -1, 4, -7), \quad \mathbf{v}_5 = (5, -8, 1, 2)$$

that forms a basis for the subspace of R^4 spanned by these vectors.

(b) Express each vector not in the basis as a linear combination of the basis vectors.

4.9 Rank, Nullity, and the Fundamental Matrix Spaces

Theorem 4.9.1. *The row space and the column space of a matrix A have the same dimension.*

Proof. Elementary row operations do not change the dimension of the row space or of the column space of a matrix. Thus, if R is any row echelon form of A, it must be true that

$$\begin{aligned}\dim(\text{row space of } A) &= \dim(\text{row space of } R)\\ \dim(\text{column space of } A) &= \dim(\text{column space of } R)\end{aligned}$$

so it suffices to show that the row and column spaces of R have the same dimension. But the dimension of the row space of R is the number of nonzero rows, and the dimension of the column space of R is the number of leading 1's. Since these two numbers are the same, the row and column space have the same dimension. $\square$

Definition 4.9.1. The common dimension of the row space and column space of a matrix A is called the rank of A and is denoted by $\text{rank}(A)$; the dimension of the null space of A is called the nullity of A and is denoted by $\text{nullity}(A)$.

Example 1. Find the rank and nullity of the matrix

$$A = \begin{bmatrix} -1 & 2 & 0 & 4 & 5 & -3 \\ 3 & -7 & 2 & 0 & 1 & 4 \\ 2 & -5 & 2 & 4 & 6 & 1 \\ 4 & -9 & 2 & -4 & -4 & 7 \end{bmatrix}.$$

Example 2. What is the maximum possible rank of an $m \times n$ matrix A that is not square?

Theorem 4.9.2 (Dimension Theorem for Matrices)**.** *If A is a matrix with n columns, then*

$$\operatorname{rank}(A) + \operatorname{nullity}(A) = n.$$

Proof. Since A has n columns, the homogeneous linear system $A\mathbf{x} = \mathbf{0}$ has n unknowns. These fall into two distinct categories: the leading variables and the free variables. Thus,

$$\begin{bmatrix} \text{number of leading} \\ \text{variables} \end{bmatrix} + \begin{bmatrix} \text{number of free} \\ \text{variables} \end{bmatrix} = n.$$

But the number of leading variables is the same as the number of leading 1's in any row echelon form of A, which is the same as the dimension of the row space of A, which is the same as the rank of A. Also, the number of free variables in the general solution of $A\mathbf{x} = \mathbf{0}$ is the same as the number of parameters in that solution, which is the same as the dimension of the solution space of $A\mathbf{x} = \mathbf{0}$, which is the same as the nullity of A. □

Example 3. Verify Theorem 4.8.2 for the matrix in Example 1.

Theorem 4.9.3. *If A is an $m \times n$ matrix, then*

(a) $\operatorname{rank}(A)$ = *the number of leading variables in the general solution of* $A\mathbf{x} = \mathbf{0}$.

(b) $\operatorname{nullity}(A)$ = *the number of parameters in the general solution of* $A\mathbf{x} = \mathbf{0}$.

Example 4.

(a) Find the number of parameters in the general solution of $A\mathbf{x} = \mathbf{0}$ if A is a 5×7 matrix of rank 3.

(b) Find the rank of a 5×7 matrix A for which $A\mathbf{x} = \mathbf{0}$ has a two-dimensional solution space.

Theorem 4.9.4. *If $A\mathbf{x} = \mathbf{b}$ is a consistent linear system of m equations in n unknowns, and if A has rank r, then the general solution of the system contains $n - r$ parameters.*

Remark 1. The following spaces associated with a matrix A and its transpose A^T are called the fundamental spaces of a matrix A:

row space of A	column space of A
null space of A	null space of A^T

The row space and null space of A are subspaces of R^n, whereas the column space of A and the null space of A^T are subspaces of R^m. The null space of A^T is also called the left null space of A because transposing both sides of the equation $A^T\mathbf{x} = \mathbf{0}$ produces the equation $\mathbf{x}^T A = \mathbf{0}^T$ in which the unknown is on the left. The dimension of the left null space of A is called the left nullity of A.

Theorem 4.9.5. *If A is any matrix, then* $\operatorname{rank}(A) = \operatorname{rank}(A^T)$.

Proof.

$$\operatorname{rank}(A) = \dim(\text{row space of } A) = \dim(\text{column space of } A^T) = \operatorname{rank}(A^T).$$

□

Example 5. Find bases for the fundamental spaces of the matrix

$$A = \begin{bmatrix} -1 & 2 & 0 & 4 & 5 & -3 \\ 3 & -7 & 2 & 0 & 1 & 4 \\ 2 & -5 & 2 & 4 & 6 & 1 \\ 4 & -9 & 2 & -4 & -4 & 7 \end{bmatrix}.$$

Definition 4.9.2. If W is a subspace of R^n, then the set of all vectors in R^n that are orthogonal to every vector in W is called the orthogonal complement of W and is denoted by the symbol $W^{\perp}$.

Theorem 4.9.6. *If W is a subspace of R^n, then:*

(a) *$W^{\perp}$ is a subspace of R^n.*
(b) *The only vector common to W and $W^{\perp}$ is $\mathbf{0}$.*
(c) *The orthogonal complement of $W^{\perp}$ is W.*

Example 6. What is the orthogonal complement of a line W through the origin in R^2? What is the orthogonal complement of a plane W through the origin in R^3?

Theorem 4.9.7. *If A is an $m \times n$ matrix, then:*

(a) *The null space of A and the row space of A are orthogonal complements in R^n.*

(b) *The null space of A^T and the column space of A are orthogonal complements in R^m.*

Theorem 4.9.8 (Equivalent Statements)**.** *If A is an $n \times n$ matrix, then the following statements are equivalent.*

(a) *A is invertible.*

(b) *$A\mathbf{x} = \mathbf{0}$ has only the trivial solution.*

(c) *The reduced row echelon form of A is I_n.*

(d) *A is expressible as a product of elementary matrices.*

(e) *$A\mathbf{x} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.*

(f) *$A\mathbf{x} = \mathbf{b}$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.*

(g) $\det(A) \neq 0$.

(h) *The column vectors of A are linearly independent.*

(i) *The row vectors of A are linearly independent.*

(j) *The column vectors of A span R^n.*

(k) *The row vectors of A span R^n.*

(l) *The column vectors of A form a basis for R^n.*

(m) *The row vectors of A form a basis for R^n.*

(n) *A has rank n.*

(o) *A has nullity 0.*

(p) *The orthogonal complement of the null space of A is R^n.*

(q) *The orthogonal complement of the row space of A is $\{\mathbf{0}\}$.*

Remark 2. A linear system with more constraints than unknowns is called an overdetermined system. A linear system with fewer constraints than unknowns is called an underdetermined system.

Theorem 4.9.9. *Let A be an $m \times n$ matrix.*

(*a*) *(Overdetermined Case). If $m > n$, then the linear system $A\mathbf{x} = \mathbf{b}$ is inconsistent for at least one vector $\mathbf{b}$ in R^n.*

(*b*) *(Underdetermined Case). If $m < n$, then for each vector $\mathbf{b}$ in R^m the linear system $A\mathbf{x} = \mathbf{b}$ is either inconsistent or has infinitely many solutions.*

Proof. (a) Assume that $m > n$, in which case the column vectors of A cannot span R^m. Thus, there is at least one vector $\mathbf{b}$ in R^m that is not in the column space of A, and for any such $\mathbf{b}$ the system $A\mathbf{x} = \mathbf{b}$ is inconsistent.
(b) Assume that $m < n$. For each vector $\mathbf{b}$ in R^n there are two possibilities: either the system $A\mathbf{x} = \mathbf{b}$ is consistent or it is inconsistent. If it is inconsistent, then the proof is complete. If it is consistent, then the general solution has $n - r$ parameters, where $r = \text{rank}(A)$. But we know from Example 2 that $\text{rank}(A)$ is at most the smaller of m and n, so

$$n - r \geq n - m > 0.$$

This means that the general solution has at least one parameter and hence there are infinitely many solutions. □

Example 7.

(a) What can you say about the solutions of an overdetermined system $A\mathbf{x} = \mathbf{b}$ of 7 equations in 5 unknowns in which A has rank $r = 4$?

(b) What can you say about the solutions of an underdetermined system $A\mathbf{x} = \mathbf{b}$ of 5 equations in 7 unknowns in which A has rank $r = 4$?

Example 8. Under what conditions is the linear system

$$\begin{aligned} x_1 - 2x_2 &= b_1 \\ x_1 - x_2 &= b_2 \\ x_1 + x_2 &= b_3 \\ x_1 + 2x_2 &= b_4 \\ x_1 + 3x_2 &= b_5 \end{aligned}$$

consistent?

Chapter 5

Eigenvalues and Eigenvectors

5.1 Eigenvalues and Eigenvectors

Definition 5.1.1. If A is an $n \times n$ matrix, then a nonzero vector $\mathbf{x}$ in R^n is called an eigenvector of A (or of the matrix operator T_A) if $A\mathbf{x}$ is a scalar multiple of $\mathbf{x}$; that is,

$$A\mathbf{x} = \lambda\mathbf{x}$$

for some scalar λ. The scalar λ is called an eigenvalue of A (or of T_A), and $\mathbf{x}$ is said to be an eigenvector corresponding to λ.

Example 1. Determine whether the vector $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is an eigenvector of

$$A = \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}.$$

Theorem 5.1.1. *If A is an $n \times n$ matrix, then λ is an eigenvalue of A if and only if it satisfies the equation*

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

This is called the characteristic equation of A.

Example 2. In Example 1 we observed that $\lambda = 3$ is an eigenvalue of the matrix

$$A = \begin{bmatrix} 3 & 0 \\ 8 & -1 \end{bmatrix}$$

but we did not explain how we found it. Use the characteristic equation to find all eigenvalues of this matrix.

Remark 1. When the determinant $\det(\lambda I - A)$ is expanded, the characteristic equation of A takes the form

$$\lambda^n + c_1 \lambda^{n-1} + \cdots + c_n = 0$$

where the left side of this equation is a polynomial of degree n in which the coefficient of λ^n is 1. The polynomial

$$p(\lambda) = \lambda^n + c_1 \lambda^{n-1} + \cdots + c_n$$

is called the characteristic polynomial of A.

Example 3. Find the eigenvalues of

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 4 & -17 & 8 \end{bmatrix}.$$

Example 4. Find the eigenvalues of the upper triangular matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ 0 & a_{22} & a_{23} & a_{24} \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & 0 & a_{44} \end{bmatrix}.$$

Theorem 5.1.2. *If A is an $n \times n$ triangular matrix (upper triangular, lower triangular, or diagonal), then the eigenvalues of A are the entries on the main diagonal of A.*

Example 5. Find the eigenvalues of the lower triangular matrix

$$A = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ -1 & \frac{2}{3} & 0 \\ 5 & -8 & -\frac{1}{4} \end{bmatrix}.$$

Theorem 5.1.3. *If A is an $n \times n$ matrix, the following statements are equivalent.*

(*a*) *λ is an eigenvalue of A.*
(*b*) *λ is a solution of the characteristic equation* $\det(\lambda I - A) = 0$.
(*c*) *The system of equations $(\lambda I - A)\mathbf{x} = \mathbf{0}$ has nontrivial solutions.*
(*d*) *There is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$.*

Remark 2. By definition, the eigenvectors of A corresponding to an eigenvalue λ are the nonzero vectors that satisfy

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

Thus, we can find the eigenvectors of A corresponding to λ by finding the nonzero vectors in the solution space of this linear system. This solution space, which is called the eigenspace of A corresponding to λ, can also be viewed as:

1. the null space of the matrix $\lambda I - A$
2. the kernel of the matrix operator $T_{\lambda I - A} : R^n \to R^n$
3. the set of vectors for which $A\mathbf{x} = \lambda\mathbf{x}$

Example 6. Find bases for the eigenspaces of the matrix

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

Example 7. Find bases for the eigenspaces of the matrix

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

Theorem 5.1.4. *A square matrix A is invertible if and only if $\lambda = 0$ is not an eigenvalue of A.*

Proof. Assume that A is an $n \times n$ matrix and observe first that $\lambda = 0$ is a solution of the characteristic equation

$$\lambda^n + c_1\lambda^{n-1} + \cdots + c_n = 0$$

if and only if the constant term c_n is zero. Thus, it suffices to prove that A is invertible if and only if $c_n \neq 0$. But

$$\det(\lambda I - A) = \lambda^n + c_1\lambda^{n-1} + \cdots + c_n$$

or, on setting $\lambda = 0$,

$$\det(-A) = c_n \quad \text{or} \quad (-1)^n \det(A) = c_n.$$

It follows from the last equation that $\det(A) = 0$ if and only if $c_n = 0$, and this in turn implies that A is invertible if and only if $c_n \neq 0$. $\square$

Example 8. Verify Theorem 5.1.4 for the matrix A in Example 7.

Theorem 5.1.5 (Equivalent Statements)**.** *If A is an $n \times n$ matrix, then the following statements are equivalent.*

(a) *A is invertible.*
(b) *$A\mathbf{x} = \mathbf{0}$ has only the trivial solution.*
(c) *The reduced row echelon form of A is I_n.*
(d) *A is expressible as a product of elementary matrices.*
(e) *$A\mathbf{x} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.*
(f) *$A\mathbf{x} = \mathbf{b}$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.*
(g) $\det(A) \neq 0$.
(h) *The column vectors of A are linearly independent.*
(i) *The row vectors of A are linearly independent.*
(j) *The column vectors of A span R^n.*
(k) *The row vectors of A span R^n.*
(l) *The column vectors of A form a basis for R^n.*
(m) *The row vectors of A form a basis for R^n.*
(n) *A has rank n.*
(o) *A has nullity 0.*
(p) *The orthogonal complement of the null space of A is R^n.*
(q) *The orthogonal complement of the row space of A is $\{\mathbf{0}\}$.*
(r) *$\lambda = 0$ is not an eigenvalue of A.*

5.2 Diagonalization

Remark 1. Products of the form $P^{-1}AP$ in which A and P are $n \times n$ matrices and P is invertible can be viewed as transformations

$$A \to P^{-1}AP$$

in which the matrix A is mapped into the matrix $P^{-1}AP$. These are called similarity transformations. In general, any property that is preserved by a similarity transformation is called a similarity invariant and is said to be invariant under similarity.

Table 1 Similarity Invariants

Property	Description
Determinant	A and $P^{-1}AP$ have the same determinant.
Invertibility	A is invertible if and only if $P^{-1}AP$ is invertible.
Rank	A and $P^{-1}AP$ have the same rank.
Nullity	A and $P^{-1}AP$ have the same nullity.
Trace	A and $P^{-1}AP$ have the same trace.
Characteristic polynomial	A and $P^{-1}AP$ have the same characteristic polynomial.
Eigenvalues	A and $P^{-1}AP$ have the same eigenvalues.
Eigenspace dimension	If λ is an eigenvalue of A (and hence $P^{-1}AP$) then the eigenspace of A corresponding to λ and the eigenspace of $P^{-1}AP$ corresponding to λ have the same dimension.

Definition 5.2.1. If A and B are square matrices, then we say that B is similar to A if there is an invertible matrix P such that $B = P^{-1}AP$.

Remark 2. Note that if B is similar to A, then it is also true that A is similar to B since we can express A as $A = Q^{-1}BQ$ by taking $Q = P^{-1}$. This being the case, we will usually say that A and B are similar matrices if either is similar to the other.

Definition 5.2.2. A square matrix A is said to be diagonalizable if it is similar to some diagonal matrix; that is, if there exists an invertible matrix P such that $P^{-1}AP$ is diagonal. In this case the matrix P is said to diagonalize A.

Theorem 5.2.1. *If A is an $n \times n$ matrix, the following statements are equivalent.*
(a) *A is diagonalizable.*
(b) *A has n linearly independent eigenvectors.*

Proof. (a) $\Rightarrow$ (b) Since A is assumed to be diagonalizable, it follows that there exist an invertible matrix P and a diagonal matrix D such that $P^{-1}AP = D$ or, equivalently,

$$AP = PD.$$

If we denote the column vectors of P by $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$, and if we assume that the diagonal entries of D are $\lambda_1, \lambda_2, \ldots, \lambda_n$, then the left side of this equation can be expressed as

$$AP = A\begin{bmatrix}\mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_n\end{bmatrix} = \begin{bmatrix}A\mathbf{p}_1 & A\mathbf{p}_2 & \cdots & A\mathbf{p}_n\end{bmatrix}$$

and the right side can be expressed as

$$PD = \begin{bmatrix}\lambda_1\mathbf{p}_1 & \lambda_2\mathbf{p}_2 & \cdots & \lambda_n\mathbf{p}_n\end{bmatrix}.$$

Thus, it follows that

$$A\mathbf{p}_1 = \lambda_1\mathbf{p}_1, \quad A\mathbf{p}_2 = \lambda_2\mathbf{p}_2, \ldots, \quad A\mathbf{p}_n = \lambda_n\mathbf{p}_n.$$

Since P is invertible, we know that its column vectors $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ are linearly independent (and hence nonzero). Thus, it follows that these n column vectors are eigenvectors of A.
(b) $\Rightarrow$ (a) Assume that A has n linearly independent eigenvectors, $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$, and that $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the corresponding eigenvalues. If we let

$$P = \begin{bmatrix}\mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_n\end{bmatrix}$$

and if we let D be the diagonal matrix that has $\lambda_1, \lambda_2, \ldots, \lambda_n$ as it successive diagonal entries, then

$$\begin{aligned} AP &= A\begin{bmatrix}\mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_n\end{bmatrix} = \begin{bmatrix}A\mathbf{p}_1 & A\mathbf{p}_2 & \cdots & A\mathbf{p}_n\end{bmatrix} \\ &= \begin{bmatrix}\lambda_1\mathbf{p}_1 & \lambda_2\mathbf{p}_2 & \cdots & \lambda_n\mathbf{p}_n\end{bmatrix} = PD. \end{aligned}$$

Since the column vectors of P are linearly independent, it follows that P is invertible, so that this last equation can be rewritten as $P^{-1}AP = D$, which shows that A is diagonalizable. □

Theorem 5.2.2.

(a) *If $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues of a matrix A, and if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are corresponding eigenvectors, then $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is a linearly independent set.*
(b) *An $n \times n$ matrix with n distinct eigenvalues is diagonalizable.*

Example 1. Find a matrix P that diagonalizes

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

Example 2. Show that the following matrix is not diagonalizable:

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

Example 3. Show that the matrix

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 4 & -17 & 8 \end{bmatrix}$$

is diagonalizable.

Example 4. Show that the matrix

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

is diagonalizable.

Theorem 5.2.3. *If k is a positive integer, λ is an eigenvalue of a matrix A, and $\mathbf{x}$ is a corresponding eigenvector, then λ^k is an eigenvalue of A^k and $\mathbf{x}$ is a corresponding eigenvector.*

Example 5. In example 2 we found the eigenvalues and corresponding eigenvectors of the matrix

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

Do the same for A^7.

Remark 3. Suppose that A is a diagonalizable $n \times n$ matrix, that P diagonalizes A, and that

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = D.$$

If k is a positive integer, then

$$A^k = PD^kP^{-1} = P \begin{bmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n^k \end{bmatrix} P^{-1}.$$

Example 6. Use Remark 3 to find A^{13}, where

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

Example 7. Use the matrices

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad J = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

to show that the converse of Theorem 5.2.2(b) is false.

Remark 4. If λ_0 is an eigenvalue of an $n \times n$ matrix A, then the dimension of the eigenspace corresponding to λ_0 is called the geometric multiplicity of λ_0, and the number of times that $\lambda - \lambda_0$ appears as a factor in the characteristic polynomial of A is called the algebraic multiplicity of λ_0.

Theorem 5.2.4 (Geometric and Algebraic Multiplicity)**.** *If A is a square matrix, then:*

- (a) *For every eigenvalue of A, the geometric multiplicity is less than or equal to the algebraic multiplicity.*
- (b) *A is diagonalizable if and only if the geometric multiplicity of every eigenvalue is equal to the algebraic multiplicity.*

5.3 Complex Vector Spaces

Definition 5.3.1. If n is a positive integer, then a complex n-tuple is a sequence of n complex numbers $(v_1, v_2, \ldots, v_n)$. The set of all complex n-tuples is called complex n-space and is denoted by C^n. Scalars are complex numbers, and the operations of addition, subtraction, and scalar multiplication are performed componentwise.

Example 1. Let

$$\mathbf{v} = (3+i, -2i, 5) \quad \text{and} \quad A = \begin{bmatrix} 1+i & -i \\ 4 & 6-2i \end{bmatrix}.$$

Find $\overline{\mathbf{v}}$, $\mathrm{Re}(\mathbf{v})$, $\mathrm{Im}(\mathbf{v})$, $\overline{A}$, $\mathrm{Re}(A)$, $\mathrm{Im}(A)$, and $\det(A)$.

Theorem 5.3.1. *If $\mathbf{u}$ and $\mathbf{v}$ are vectors in C^n, and if k is a scalar, then:*
(*a*) $\overline{\overline{\mathbf{u}}} = \mathbf{u}$
(*b*) $\overline{k\mathbf{u}} = \overline{k}\overline{\mathbf{u}}$
(*c*) $\overline{\mathbf{u}+\mathbf{v}} = \overline{\mathbf{u}} + \overline{\mathbf{v}}$
(*d*) $\overline{\mathbf{u}-\mathbf{v}} = \overline{\mathbf{u}} - \overline{\mathbf{v}}$

Theorem 5.3.2. *If A is an $m \times k$ complex matrix and B is a $k \times n$ complex matrix, then:*
(*a*) $\overline{\overline{A}} = A$
(*b*) $(\overline{A^T}) = (\overline{A})^T$
(*c*) $\overline{AB} = \overline{A}\,\overline{B}$

Definition 5.3.2. If $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ are vectors in C^n, then the complex Euclidean inner product of $\mathbf{u}$ and $\mathbf{v}$ (also called the complex dot product) is denoted by $\mathbf{u} \cdot \mathbf{v}$ and is defined as

$$\mathbf{u} \cdot \mathbf{v} = u_1\overline{v}_1 + u_2\overline{v}_2 + \cdots + u_n\overline{v}_n.$$

We also define the Euclidean norm on C^n to be

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{|v_1|^2 + |v_2|^2 + \cdots + |v_n|^2}.$$

Example 2. Find $\mathbf{u} \cdot \mathbf{v}$, $\mathbf{v} \cdot \mathbf{u}$, $\|\mathbf{u}\|$, and $\|\mathbf{v}\|$ for the vectors

$$\mathbf{u} = (1+i, i, 3-i) \quad \text{and} \quad \mathbf{v} = (1+i, 2, 4i).$$

Theorem 5.3.3. *If $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ are vectors in C^n, and if k is a scalar, then the complex Euclidean inner product has the following properties:*

(a) $\mathbf{u} \cdot \mathbf{v} = \overline{\mathbf{v} \cdot \mathbf{u}}$
(b) $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$
(c) $k(\mathbf{u} \cdot \mathbf{v}) = (k\mathbf{u}) \cdot \mathbf{v}$
(d) $\mathbf{u} \cdot k\mathbf{v} = \overline{k}(\mathbf{u} \cdot \mathbf{v})$
(e) $\mathbf{v} \cdot \mathbf{v} \geq 0$ *and* $\mathbf{v} \cdot \mathbf{v} = 0$ *if and only if* $\mathbf{v} = \mathbf{0}$.

Theorem 5.3.4. *If λ is an eigenvalue of a real $n \times n$ matrix A, and if $\mathbf{x}$ is a corresponding eigenvector, then $\overline{\lambda}$ is also an eigenvalue of A, $\overline{\mathbf{x}}$ is a corresponding eigenvector.*

Proof. Since λ is an eigenvalue of A and $\mathbf{x}$ is a corresponding eigenvector, we have

$$\overline{A\mathbf{x}} = \overline{\lambda\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}.$$

However, $\overline{A} = A$, since A has real entries, so it follows that

$$\overline{A\mathbf{x}} = \overline{A}\overline{\mathbf{x}} = A\overline{\mathbf{x}}.$$

Therefore

$$A\overline{\mathbf{x}} = \overline{A\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}$$

in which $\overline{x} \neq \mathbf{0}$; this tells us that $\overline{\lambda}$ is an eigenvalue of A and $\overline{\mathbf{x}}$ is a corresponding eigenvector. $\square$

Example 3. Find the eigenvalues and bases for the eigenspaces of

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

Theorem 5.3.5. *If A is a 2×2 matrix with real entries, then the characteristic equation of A is $\lambda^2 - \operatorname{tr}(A)\lambda + \det(A) = 0$ and*
(a) *A has two distinct real eigenvalues if $\operatorname{tr}(A)^2 - 4\det(A) > 0$;*
(b) *A has one repeated real eigenvalue if $\operatorname{tr}(A)^2 - 4\det(A) = 0$;*
(c) *A has two complex conjugate eigenvalues if $\operatorname{tr}(A)^2 - 4\det(A) < 0$.*

Example 4. In each part, use the characteristic equation to find the eigenvalues of

(a) $A = \begin{bmatrix} 2 & 2 \\ -1 & 5 \end{bmatrix}$

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

(c) $A = \begin{bmatrix} 2 & 3 \\ -3 & 2 \end{bmatrix}$

Theorem 5.3.6. *If A is a real symmetric matrix, then A has real eigenvalues.*

Proof. Suppose that λ is an eigenvalue of A and $\mathbf{x}$ is a corresponding eigenvector, where we allow for the possibility that λ is complex and $\mathbf{x}$ is in C^n. Thus,

$$A\mathbf{x} = \lambda\mathbf{x}$$

where $\mathbf{x} \neq \mathbf{0}$. If we multiply both sides of this equation by $\overline{\mathbf{x}}^T$ and use the fact that

$$\overline{\mathbf{x}}^T A\mathbf{x} = \overline{\mathbf{x}}^T(\lambda\mathbf{x}) = \lambda(\overline{\mathbf{x}}^T\mathbf{x}) = \lambda(\mathbf{x} \cdot \mathbf{x}) = \lambda\|\mathbf{x}\|^2$$

then we obtain

$$\lambda = \frac{\overline{x}^T A\mathbf{x}}{\|\mathbf{x}\|^2}.$$

Since the denominator in this expression is real, we can prove that λ is real by showing that

$$\overline{\overline{\mathbf{x}}^T A\mathbf{x}} = \overline{\mathbf{x}}^T A\mathbf{x}.$$

But A is symmetric and has real entries, so it follows that

$$\overline{\overline{\mathbf{x}}^T A\mathbf{x}} = \overline{\overline{\mathbf{x}}}^T \overline{A\mathbf{x}} = \mathbf{x}^T \overline{A\mathbf{x}} = (\overline{A\mathbf{x}})^T\mathbf{x} = (\overline{A}\overline{\mathbf{x}})^T\mathbf{x} = (A\overline{\mathbf{x}})^T\mathbf{x} = \overline{\mathbf{x}}^T A^T\mathbf{x} = \overline{\mathbf{x}}^T A\mathbf{x}.$$

$\square$

Theorem 5.3.7. *The eigenvalues of the real matrix*

$$C = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

are $\lambda = a \pm bi$. If a and b are not both zero, then this matrix can be factored as

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$$

where ϕ is the angle from the positive x-axis to the ray that joins the origin to the point (a, b).

Proof. The characteristic equation of C is $(\lambda - a)^2 + b^2 = 0$, from which it follows that the eigenvalues of C are $\lambda = a \pm bi$. Assuming that a and b are not both zero, let ϕ be the angle from the positive x-axis to the ray that joins the origin to the point (a, b). The angle ϕ is an argument of the eigenvalue $\lambda = a + bi$, so

$$a = |\lambda| \cos\phi \quad \text{and} \quad b = |\lambda| \sin\phi.$$

It follows from this that the matrix C can be written as

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} = \begin{bmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{bmatrix} \begin{bmatrix} \dfrac{a}{|\lambda|} & -\dfrac{b}{|\lambda|} \\ \dfrac{b}{|\lambda|} & \dfrac{a}{|\lambda|} \end{bmatrix} = \begin{bmatrix} |\lambda| & 0 \\ 0 & |\lambda| \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \qquad \square$$

Theorem 5.3.8. *Let A be a real 2×2 matrix with complex eigenvalues $\lambda = a \pm bi$ (where $b \neq 0$). If $\mathbf{x}$ is an eigenvector of A corresponding to $\lambda = a - bi$, then the matrix $P = \begin{bmatrix} \mathrm{Re}(\mathbf{x}) & \mathrm{Im}(\mathbf{x}) \end{bmatrix}$ is invertible and*

$$A = P \begin{bmatrix} a & -b \\ b & a \end{bmatrix} P^{-1}.$$

Example 5. Factor the matrix in Example 3 into the form given in Theorem 5.3.8 using the eigenvalue $\lambda = -i$ and the corresponding eigenvector previously obtained.

5.4 Differential Equations

Remark 1. A differential equation is an equation involving unknown functions and their derivatives. The order of a differential equation is the order of the highest derivative it contains. The simplest differential equations are the first-order differential equations of the form

$$y' = ay$$

where $y = f(x)$ is an unknown differentiable function to be determined, $y' = dy/dx$ is its derivative, and a is a constant. As with most differential equations, this equation has infinitely many solutions; they are the functions of the form

$$y = ce^{ax}$$

where c is an arbitrary constant. That every function of this form is a solution follows from the computation

$$y' = cae^{ax} = ay.$$

Accordingly, we call $y = ce^{ax}$ the general solution of $y' = ay$.

A condition which specifies the value of the general solution at a point is called an initial condition, and the problem of solving a differential equation subject to an initial condition is called an initial-value problem.

Remark 2. The system of differential equations

$$\begin{array}{ccccccccc} y_1' & = & a_{11}y_1 & + & a_{12}y_2 & + & \cdots & + & a_{1n}y_n \\ y_2' & = & a_{21}y_1 & + & a_{22}y_2 & + & \cdots & + & a_{2n}y_n \\ \vdots & & \vdots & & \vdots & & & & \vdots \\ y_n' & = & a_{n1}y_1 & + & a_{n2}y_2 & + & \cdots & + & a_{nn}y_n. \end{array}$$

where $y_1 = f_1(x), y_2 = f_2(x), \ldots, y_n = f_n(x)$ are functions to be determined, and the a_{ij}'s are constants, is called a constant coefficient first-order homogeneous linear system. The solution

$$y_1 = y_2 = \cdots = y_n = 0$$

is called the trivial solution.

Example 1.

(a) Write the following system in matrix form:

$$\begin{aligned} y_1' &= 3y_1 \\ y_2' &= -2y_2 \\ y_3' &= 5y_3 \end{aligned}$$

(b) Solve the system.

(c) Find a solution of the system that satisfies the initial conditions $y_1(0) = 1$, $y_2(0) = 4$, and $y_3(0) = -2$.

Example 2.

(a) Solve the system

$$\begin{aligned} y_1' &= y_1 + y_2 \\ y_2' &= 4y_1 - 2y_2. \end{aligned}$$

(b) Find the solution that satisfies the initial conditions $y_1(0) = 1$, $y_2(0) = 6$.

5.5 Dynamical Systems and Markov Chains

Remark 1. A dynamical system is a finite set of variables whose values change with time. The value of a variable at a point in time is called the state of the variable at that time, and the vector formed from these states is called the state vector of the dynamical system at that time.

Example 1. Suppose that two competing television channels, channel 1 and channel 2, each have 50% of the viewer market at some initial point in time. Assume that over each one-year period channel 1 captures 10% of channel 2's share, and channel 2 captures 20% of channel 1's share. What is each channel's market share after one year?

Example 2. Track the market shares of channels 1 and 2 in Example 1 over a five-year period.

Remark 2. In many dynamical systems the states of the variables are not known with certainty but can be expressed as probabilities; such dynamical systems are called stochastic processes. Stated informally, the probability that an experiment or observation will have a certain outcome is the fraction of time that the outcome would occur if the experiment could be repeated indefinitely under constant conditions—the greater the number of actual repetitions, the more accurately the probability describes the fraction of time that the outcome occurs.

Example 3. Interpret the entries in the state vector in Example 1 as probabilities.

Remark 3. A square matrix, each of whose columns is a probability vector, is called a stochastic matrix.

Definition 5.5.1. A Markov chain is a dynamical system whose state vectors at a succession of equally spaced times are probability vectors and for which the state vectors at successive times are related by an equation of the form

$$\mathbf{x}(k+1) = P\mathbf{x}(k)$$

in which $P = [p_{ij}]$ is a stochastic matrix and p_{ij} is the probability that the system will be in state i at time $t = k+1$ if it is in state j at time $t = k$. The matrix P is called the transition matrix for the system.

Example 4. Suppose that a tagged lion can migrate over three adjacent game reserves in search of food: Reserve 1, Reserve 2, and Reserve 3. Based on data about the food resources, researchers conclude that the monthly migration pattern of the lion can be modeled by a Markov chain with transition matrix

$$\begin{array}{c} \textbf{Reserve at time } \boldsymbol{t = k} \\ \begin{array}{ccc} \textbf{1} & \textbf{2} & \textbf{3} \end{array} \end{array}$$

$$P = \begin{bmatrix} 0.5 & 0.4 & 0.6 \\ 0.2 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.1 \end{bmatrix} \begin{array}{c} \textbf{1} \\ \textbf{2} \\ \textbf{3} \end{array} \quad \textbf{Reserve at time } \boldsymbol{t = k+1}$$

That is,

$p_{11} = 0.5 =$ probability that the lion will stay in Reserve 1 when it is in Reserve 1
$p_{12} = 0.4 =$ probability that the lion will move from Reserve 2 to Reserve 1
$p_{13} = 0.6 =$ probability that the lion will move from Reserve 3 to Reserve 1
$p_{21} = 0.2 =$ probability that the lion will move from Reserve 1 to Reserve 2
$p_{22} = 0.2 =$ probability that the lion will stay in Reserve 2 when it is in Reserve 2
$p_{23} = 0.3 =$ probability that the lion will move from Reserve 3 to Reserve 2
$p_{31} = 0.3 =$ probability that the lion will move from Reserve 1 to Reserve 3
$p_{32} = 0.4 =$ probability that the lion will move from Reserve 2 to Reserve 3
$p_{33} = 0.1 =$ probability that the lion will stay in Reserve 3 when it is in Reserve 3.

Assuming that t is in months and the lion is released in Reserve 2 at time $t = 0$, track its probable locations over a six-month period, and find the reserve in which it is most likely to be at the end of that period.

Remark 4. In a Markov chain with an initial state of $\mathbf{x}(0)$, the successive state vectors are

$$\mathbf{x}(1) = P\mathbf{x}(0), \quad \mathbf{x}(2) = P\mathbf{x}(1), \quad \mathbf{x}(3) = P\mathbf{x}(2), \quad \mathbf{x}(4) = P\mathbf{x}(3), \ldots.$$

For brevity, it is common to denote $\mathbf{x}(k)$ by $\mathbf{x}_k$, which allows us to write the successive state vectors more briefly as

$$\mathbf{x}_1 = P\mathbf{x}_0, \quad \mathbf{x}_2 = P\mathbf{x}_1, \quad \mathbf{x}_3 = P\mathbf{x}_2, \quad \mathbf{x}_4 = P\mathbf{x}_3, \ldots.$$

Alternatively, these state vectors can be expressed in terms of the initial state vector $\mathbf{x}_0$ as

$$\mathbf{x}_1 = P\mathbf{x}_0, \quad \mathbf{x}_2 = P(P\mathbf{x}_0) = P^2\mathbf{x}_0, \quad \mathbf{x}_3 = P(P^2\mathbf{x}_0) = P^3\mathbf{x}_0, \quad \mathbf{x}_4 = P(P^3\mathbf{x}_0) = P^4\mathbf{x}_0, \ldots$$

from which it follows that

$$\mathbf{x}_k = P^k\mathbf{x}_0.$$

Example 5. Use Remark 4 to find the state vector $\mathbf{x}(3)$ in Example 2.

Example 6. The matrix

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

is stochastic and hence can be regarded as the transition matrix for a Markov chain. Find the successive states in the Markov chain with initial vector $\mathbf{x}_0$.

Remark 5. We say that a sequence of vectors

$$\mathbf{x}_1, \quad \mathbf{x}_2, \ldots, \quad \mathbf{x}_k, \ldots$$

approaches a limit $\mathbf{q}$ or that it converges to $\mathbf{q}$ if all entries in $\mathbf{x}_k$ can be made as close as we like to the corresponding entries in the vector $\mathbf{q}$ by taking k to be sufficiently large. We denote this by writing $\mathbf{x}_k \to \mathbf{q}$ as $k \to \infty$. Similarly, we say that a sequence of matrices

$$P_1, P_2, P_3, \ldots, P_k, \ldots$$

converges to a matrix Q, written $P_k \to Q$ as $k \to \infty$, if each entry of P_k can be made as close as we like to the corresponding entry of Q by taking k to be sufficiently large.

Definition 5.5.2. A stochastic matrix P is said to be regular if P or some positive power of P has all positive entries, and a Markov chain whose transition matrix is regular is said to be a regular Markov chain.

Example 7. Which transition matrices in Examples 2, 4, and 6 are regular?

Theorem 5.5.1. *If P is the transition matrix for a regular Markov chain, then:*

(*a*) *There is a unique probability vector $\mathbf{q}$ with positive entries such that $P\mathbf{q} = \mathbf{q}$.*

(*b*) *For any initial probability vector $\mathbf{x}_0$, the sequence of state vectors*

$$\mathbf{x}_0, \quad P\mathbf{x}_0, \ldots, \quad P^k\mathbf{x}_0, \ldots$$

converges to $\mathbf{q}$.

(*c*) *The sequence $P, P^2, P^3, \ldots, P^k, \ldots$ converges to the matrix Q each of whose column vectors is $\mathbf{q}$.*

Remark 6. The vector $\mathbf{q}$ in Theorem 5.5.1 is called the steady-state vector of the Markov chain.

Example 8. Find the steady-state vector of the Markov chain in Example 2.

Example 9. Find the steady-state vector of the Markov chain in Example 4.

Chapter 6

Inner Product Spaces

6.1 Inner Products

Definition 6.1.1. An inner product on a real vector space V is a function that associates a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ with each pair of vectors in V in such a way that the following axioms are satisfied for all vectors $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V and all scalars k.

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
3. $\langle k\mathbf{u}, \mathbf{v} \rangle = k\langle \mathbf{u}, \mathbf{v} \rangle$
4. $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ and $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$

A real vector space with an inner product is called a real inner product space.

Remark 1. The inner product

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$

of two vectors $\mathbf{u}$ and $\mathbf{v}$ in R^n is called the Euclidean inner product (or the standard inner product) on R^n to distinguish it from other possible inner products that might be defined on R^n. We call R^n with the Euclidean inner product Euclidean n-space.

Definition 6.1.2. If V is a real inner product space, then the norm (or length) of a vector $\mathbf{v}$ in V is denoted by $\|\mathbf{v}\|$ and is defined by

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$$

and the distance between two vectors is denoted by $d(\mathbf{u}, \mathbf{v})$ and is defined by

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\| = \sqrt{\langle \mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v} \rangle}.$$

A vector of norm 1 is called a unit vector.

Theorem 6.1.1. *If* $\mathbf{u}$ *and* $\mathbf{v}$ *are vectors in a real inner product space* V*, and if* k *is a scalar, then:*
(a) $\|\mathbf{v}\| \geq 0$ *with equality if and only if* $\mathbf{v} = \mathbf{0}$.
(b) $\|k\mathbf{v}\| = |k|\|\mathbf{v}\|$.
(c) $d(\mathbf{u}, \mathbf{v}) = d(\mathbf{v}, \mathbf{u})$.
(d) $d(\mathbf{u}, \mathbf{v}) \geq 0$ *with equality if and only if* $\mathbf{u} = \mathbf{v}$.

Remark 2. If

$$w_1, w_2, \ldots, w_n$$

are *positive* real numbers, which we will call weights, and if $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ are vectors in R^n, then it can be shown that the formula

$$\langle \mathbf{u}, \mathbf{v} \rangle = w_1 u_1 v_1 + w_2 u_2 v_2 + \cdots + w_n u_n v_n$$

defines an inner product on R^n that we call the weighted Euclidean inner product with weights $w_1, w_2, \ldots, w_n$.

Example 1. Let $\mathbf{u} = (u_1, u_2)$ and $\mathbf{v} = (v_1, v_2)$ be vectors in R^2. Verify that the weighted Euclidean inner product

$$\langle \mathbf{u}, \mathbf{v} \rangle = 3u_1 v_1 + 2u_2 v_2$$

satisfies the four inner product axioms.

Example 2. Calculate $\|\mathbf{u}\|$ and $d(\mathbf{u}, \mathbf{v})$ for the vectors $\mathbf{u} = (1, 0)$ and $\mathbf{v} = (0, 1)$ in R^2 with the Euclidean inner product and with the weighted Euclidean inner product from Example 1.

Definition 6.1.3. If V is an inner product space, then the set of points in V that satisfy

$$\|\mathbf{u}\| = 1$$

is called the <u>unit sphere</u> or sometimes the <u>unit circle</u> in V.

Example 3.

(a) Sketch the unit circle in an xy-coordinate system in R^2 using the Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_1 + u_2 v_2$.

(b) Sketch the unit circle in an xy-coordinate system in R^2 using the weighted Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = \frac{1}{9} u_1 v_1 + \frac{1}{4} u_2 v_2$.

Remark 3. The Euclidean inner product and the weighted Euclidean inner products are special cases of a general class of inner products on R^n called matrix inner products. To define this class of inner products, let $\mathbf{u}$ and $\mathbf{v}$ be vectors in R^n that are expressed in *column form*, and let A be an *invertible* $n \times n$ matrix. It can be shown that if $\mathbf{u} \cdot \mathbf{v}$ is the Euclidean inner product on R^n, then the formula

$$\langle \mathbf{u}, \mathbf{v} \rangle = A\mathbf{u} \cdot A\mathbf{v}$$

also defines an inner product; it is called the inner product on R^n generated by A.

Example 4. Show that the standard Euclidean and weighted Euclidean inner products are special cases of matrix inner products.

Example 5. The weighted Euclidean inner product discussed in Example 1 is the inner product on R^2 generated by what matrix?

Example 6. If $\mathbf{u} = U$ and $\mathbf{v} = V$ are matrices in the vector space M_{nn}, then show that the formula

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathrm{tr}(U^T V)$$

defines an inner product on M_{nn} called the standard inner product on that space.

Example 7. If

$$\mathbf{p} = a_0 + a_1 x + \cdots + a_n x^n \quad \text{and} \quad \mathbf{q} = b_0 + b_1 x + \cdots + b_n x^n$$

are polynomials in P_n, then show that the following formula defines an inner product on P_n that we call the standard inner product on this space:

$$\langle \mathbf{p}, \mathbf{q} \rangle = a_0 b_0 + a_1 b_1 + \cdots + a_n b_n.$$

Example 8. If

$$\mathbf{p} = a_0 + a_1 x + \cdots + a_n x^n \quad \text{and} \quad \mathbf{q} = b_0 + b_1 x + \cdots + b_n x^n$$

are polynomials in P_n, and if $x_0, x_1, \ldots, x_n$ are distinct real numbers (called sample points), then show that the formula

$$\langle \mathbf{p}, \mathbf{q} \rangle = p(x_0)q(x_0) + p(x_1)q(x_1) + \cdots + p(x_n)q(x_n).$$

defines an inner product on P_n called the evaluation inner product at $x_0, x_1, \ldots, x_n$.

Example 9. Let P_2 have the evaluation inner product at the points

$$x_0 = -2, \quad x_1 = 0, \quad \text{and} \quad x_2 = 2.$$

Compute $\langle \mathbf{p}, \mathbf{q} \rangle$ and $\|\mathbf{p}\|$ for the polynomials $\mathbf{p} = p(x) = x^2$ and $\mathbf{q} = q(x) = 1 + x$.

Example 10. Let $\mathbf{f} = f(x)$ and $\mathbf{g} = g(x)$ be two functions in $C[a, b]$ and define

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_a^b f(x)g(x)\,dx.$$

Show that this formula defines an inner product on $C[a, b]$.

Example 11. If $C[a,b]$ has the inner product that was defined in Example 10, then what is the norm of a function $\mathbf{f} = f(x)$ relative to this inner product?

Theorem 6.1.2. *If* $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *are vectors in a real inner product space* V*, and if* k *is a scalar, then:*

(a) $\langle \mathbf{0}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{0} \rangle = 0$
(b) $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$
(c) $\langle \mathbf{u}, \mathbf{v} - \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle - \langle \mathbf{u}, \mathbf{w} \rangle$
(d) $\langle \mathbf{u} - \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle - \langle \mathbf{v}, \mathbf{w} \rangle$
(e) $k\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, k\mathbf{v} \rangle$

Example 12. Compute

$$\langle \mathbf{u} - 2\mathbf{v}, 3\mathbf{u} + 4\mathbf{v} \rangle$$

in terms of $\|\mathbf{u}\|$, $\|\mathbf{v}\|$, and $\langle \mathbf{u}, \mathbf{v} \rangle$.

6.2 Angle and Orthogonality in Inner Product Spaces

Theorem 6.2.1 (Cauchy-Schwarz Inequality)**.** *If* $\mathbf{u}$ *and* $\mathbf{v}$ *are vectors in a real inner product space* V*, then*

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\|.$$

Proof. In the case where $\mathbf{u} = \mathbf{0}$ the two sides of the inequality are equal since $\langle \mathbf{u}, \mathbf{v} \rangle$ and $\|\mathbf{u}\|$ are both zero. Thus, we need only consider the case where $\mathbf{u} \neq \mathbf{0}$. Making this assumption, let

$$a = \langle \mathbf{u}, \mathbf{u} \rangle, \quad b = 2\langle \mathbf{u}, \mathbf{v} \rangle, \quad c = \langle \mathbf{v}, \mathbf{v} \rangle$$

and let t be any real number. Since the positivity axiom states that the inner product of any vector with itself is nonnegative, it follows that

$$\begin{aligned} 0 \leq \langle t\mathbf{u} + \mathbf{v}, t\mathbf{u} + \mathbf{v} \rangle &= \langle \mathbf{u}, \mathbf{u} \rangle t^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle t + \langle \mathbf{v}, \mathbf{v} \rangle \\ &= at^2 + bt + c. \end{aligned}$$

This inequality implies that the quadratic polynomial $at^2 + bt + c$ has either no real roots or a repeated real root. Therefore, its discriminant must satisfy the inequality $b^2 - 4ac \leq 0$. Expressing the coefficients a, b, and c in terms of the vectors $\mathbf{u}$ and $\mathbf{v}$ gives $4\langle \mathbf{u}, \mathbf{v} \rangle^2 - 4\langle \mathbf{u}, \mathbf{u} \rangle \langle \mathbf{v}, \mathbf{v} \rangle \leq 0$ or, equivalently,

$$\langle \mathbf{u}, \mathbf{v} \rangle^2 \leq \langle \mathbf{u}, \mathbf{u} \rangle \langle \mathbf{v}, \mathbf{v} \rangle.$$

Taking square roots of both sides and using the fact that $\langle \mathbf{u}, \mathbf{u} \rangle$ and $\langle \mathbf{v}, \mathbf{v} \rangle$ are nonnegative yields

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \langle \mathbf{u}, \mathbf{u} \rangle^{1/2} \langle \mathbf{v}, \mathbf{v} \rangle^{1/2} \quad \text{or equivalently} \quad |\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\|. \qquad \square$$

Remark 1. If $\mathbf{u}$ and $\mathbf{v}$ are vectors in a real inner product space V, then the angle θ between $\mathbf{u}$ and $\mathbf{v}$ is defined to be

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

Example 1. Let M_{22} have the standard inner product. Find the cosine of the angle between the vectors

$$\mathbf{u} = U = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{v} = V = \begin{bmatrix} -1 & 0 \\ 3 & 2 \end{bmatrix}.$$

Theorem 6.2.2. *If* $\mathbf{u}$, $\mathbf{v}$, *and* $\mathbf{w}$ *are vectors in a real inner product space* V, *and if* k *is a scalar, then:*
(*a*) $\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$
(*b*) $d(\mathbf{u}, \mathbf{v}) \leq d(\mathbf{u}, \mathbf{w}) + d(\mathbf{w}, \mathbf{v})$

Definition 6.2.1. Two vectors $\mathbf{u}$ and $\mathbf{v}$ in an inner product space V are called orthogonal if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

Example 2. Are the vectors $\mathbf{u} = (1, 1)$ and $\mathbf{v} = (1, -1)$ orthogonal with respect to the Euclidean inner product on R^2? What about with respect to the weighted Euclidean inner product $\langle \mathbf{u}, \mathbf{v} \rangle = 3u_1v_1 + 2u_2v_2$?

Example 3. If M_{22} has the inner product of Example 6 in the preceding section, then are the matrices

$$U = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}$$

orthogonal?

Example 4. Let P_2 have the inner product

$$\langle \mathbf{p}, \mathbf{q} \rangle = \int_{-1}^{1} p(x)q(x)\,dx$$

and let $\mathbf{p} = x$ and $\mathbf{q} = x^2$. Find $\|\mathbf{p}\|$ and $\|\mathbf{q}\|$ and show that $\mathbf{p}$ and $\mathbf{q}$ are orthogonal relative to the given inner product.

Theorem 6.2.3 (Generalized Theorem of Pythagoras)**.** *If* $\mathbf{u}$ *and* $\mathbf{v}$ *are orthogonal vectors in a real inner product space, then*

$$\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2.$$

Proof. The orthogonality of $\mathbf{u}$ and $\mathbf{v}$ implies that $\langle \mathbf{u}, \mathbf{v} \rangle = 0$, so

$$\begin{aligned} \|\mathbf{u} + \mathbf{v}\|^2 &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle = \|\mathbf{u}\|^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle + \|\mathbf{v}\|^2 \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2. \end{aligned}$$ □

Example 5. Verify Theorem 6.2.3 for the vectors $\mathbf{p}$ and $\mathbf{q}$ and inner product discussed in Example 4.

Definition 6.2.2. If W is a subspace of a real inner product space V, then the set of all vectors in V that are orthogonal to every vector in W is called the orthogonal complement of W and is denoted by the symbol $W^{\perp}$.

Theorem 6.2.4. *If W is a subspace of a real inner product space V, then:*
(*a*) *$W^\perp$ is a subspace of V.*
(*b*) *$W \cap W^\perp = \{\mathbf{0}\}$.*

Proof. (a) The set $W^\perp$ contains at least the zero vector, since $\langle \mathbf{0}, \mathbf{w} \rangle = 0$ for every vector $\mathbf{w}$ in W. Thus, it remains to show that $W^\perp$ is closed under addition and scalar multiplication. To do this, suppose that $\mathbf{u}$ and $\mathbf{v}$ are vectors in $W^\perp$, so that for every vector $\mathbf{w}$ in W we have $\langle \mathbf{u}, \mathbf{w} \rangle = 0$ and $\langle \mathbf{v}, \mathbf{w} \rangle = 0$. It follows from the additivity and homogeneity axioms of inner products that

$$\begin{aligned} \langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle &= \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle = 0 + 0 = 0 \\ \langle k\mathbf{u}, \mathbf{w} \rangle &= k\langle \mathbf{u}, \mathbf{w} \rangle = k(0) = 0 \end{aligned}$$

which proves that $\mathbf{u} + \mathbf{v}$ and $k\mathbf{u}$ are in $W^\perp$.
(b) If $\mathbf{v}$ is any vector in both W and $W^\perp$, then $\mathbf{v}$ is orthogonal to itself; that is, $\langle \mathbf{v}, \mathbf{v} \rangle = 0$. It follows from the positivity axiom for inner products that $\mathbf{v} = \mathbf{0}$. □

Theorem 6.2.5. *If W is a subspace of a real finite-dimensional inner product space V, then the orthogonal complement of $W^\perp$ is W; that is,*

$$(W^\perp)^\perp = W.$$

Example 6. Let W be the subspace of R^6 spanned by the vectors

$$\begin{aligned} \mathbf{w}_1 &= (1, 3, -2, 0, 2, 0), \qquad & \mathbf{w}_2 &= (2, 6, -5, -2, 4-3), \\ \mathbf{w}_3 &= (0, 0, 5, 10, 0, 15), \qquad & \mathbf{w}_4 &= (2, 6, 0, 8, 4, 18). \end{aligned}$$

Find a basis for the orthogonal complement of W.

6.3 Gram-Schmidt Process; QR-Decomposition

Definition 6.3.1. A set of two or more vectors in a real inner product space is said to be orthogonal if all pairs of distinct vectors in the set are orthogonal. An orthogonal set in which each vector has norm 1 is said to be orthonormal.

Example 1. Let

$$\mathbf{v}_1 = (0, 1, 0), \quad \mathbf{v}_2 = (1, 0, 1), \quad \mathbf{v}_3 = (1, 0, -1)$$

and assume that R^3 has the Euclidean inner product. Is the set $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ orthogonal?

Remark 1. The process of multiplying a vector $\mathbf{v}$ by the reciprocal of its length is called normalizing $\mathbf{v}$.

Example 2. Normalize the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ in Example 1.

Theorem 6.3.1. *If $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is an orthogonal set of nonzero vectors in an inner product space, then S is linearly independent.*

Proof. Assume that

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_n\mathbf{v}_n = \mathbf{0}.$$

To demonstrate that $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ is linearly independent, we must prove that $k_1 = k_2 = \cdots = k_n = 0$.

For each $\mathbf{v}_i$ in S, it follows that

$$\langle k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_n\mathbf{v}_n, \mathbf{v}_i\rangle = \langle \mathbf{0}, \mathbf{v}_i\rangle = 0$$

or, equivalently,

$$k_1\langle \mathbf{v}_1, \mathbf{v}_i\rangle + k_2\langle \mathbf{v}_2, \mathbf{v}_i\rangle + \cdots + k_n\langle \mathbf{v}_n, \mathbf{v}_i\rangle = 0.$$

From the orthogonality of S it follows that $\langle \mathbf{v}_j, \mathbf{v}_i\rangle = 0$ when $j \neq i$, so this equation reduces to

$$k_i\langle \mathbf{v}_i, \mathbf{v}_i\rangle = 0.$$

Since the vectors in S are assumed to be nonzero, it follows from the positivity axiom for inner products that $\langle \mathbf{v}_i, \mathbf{v}_i\rangle \neq 0$. Thus, the preceding equation implies that each k_i is zero, which is what we wanted to prove. □

Remark 2. In an inner product space, a basis consisting of orthonormal vectors is called an orthonormal basis, and a basis consisting of orthogonal vectors is called an orthogonal basis.

Example 3. Show that the standard basis is orthonormal with respect to the standard inner product for P_n.

Example 4. Show that the vectors $\mathbf{u}_1$, $\mathbf{u}_2$, and $\mathbf{u}_3$ from Example 2 form an orthonormal basis for R^3.

Theorem 6.3.2.

(*a*) *If* $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ *is an orthogonal basis for an inner product space* V*, and if* $\mathbf{u}$ *is any vector in* V*, then*

$$\mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{u}, \mathbf{v}_2\rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{u}, \mathbf{v}_n\rangle}{\|\mathbf{v}_n\|^2}\mathbf{v}_n.$$

(*b*) *If* $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ *is an orthonormal basis for an inner product space* V*, and if* $\mathbf{u}$ *is any vector in* V*, then*

$$\mathbf{u} = \langle \mathbf{u}, \mathbf{v}_1\rangle\mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2\rangle\mathbf{v}_2 + \cdots + \langle \mathbf{u}, \mathbf{v}_n\rangle\mathbf{v}_n.$$

Proof. (a) Since $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for V, every vector $\mathbf{u}$ in V can be expressed in the form

$$\mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n.$$

We will complete the proof by showing that

$$c_i = \frac{\langle \mathbf{u}, \mathbf{v}_i \rangle}{\|\mathbf{v}_i\|^2}$$

for $i = 1, 2, \ldots, n$. To do this, observe first that

$$\begin{aligned}\langle \mathbf{u}, \mathbf{v}_i \rangle &= \langle c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n, \mathbf{v}_i \rangle \\ &= c_1\langle \mathbf{v}_1, \mathbf{v}_i \rangle + c_2\langle \mathbf{v}_2, \mathbf{v}_i \rangle + \cdots + c_n\langle \mathbf{v}_n, \mathbf{v}_i \rangle.\end{aligned}$$

Since S is an orthogonal set, all of the inner products in the last equality are zero except the ith, so we have

$$\langle \mathbf{u}, \mathbf{v}_i \rangle = c_i\langle \mathbf{v}_i, \mathbf{v}_i \rangle = c_i\|\mathbf{v}_i\|^2.$$

Solving this equation for c_i yields the desired result.
(b) Here $\|\mathbf{v}_1\| = \|\mathbf{v}_2\| = \cdots = \|\mathbf{v}_n\| = 1$, so part (a) simplifies to part (b). □

Remark 3. The coordinate vector of a vector $\mathbf{u}$ in V relative to an orthogonal basis $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is

$$(\mathbf{u})_S = \left(\frac{\langle \mathbf{u}, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}, \frac{\langle \mathbf{u}, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}, \ldots, \frac{\langle \mathbf{u}, \mathbf{v}_n \rangle}{\|\mathbf{v}_n\|^2} \right)$$

and relative to an orthonormal basis $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is

$$(\mathbf{u})_S = (\langle \mathbf{u}, \mathbf{v}_1 \rangle, \langle \mathbf{u}, \mathbf{v}_2 \rangle, \ldots, \langle \mathbf{u}, \mathbf{v}_n \rangle).$$

Example 5. Let

$$\mathbf{v}_1 = (0, 1, 0), \quad \mathbf{v}_2 = \left(-\tfrac{4}{5}, 0, \tfrac{3}{5}\right), \quad \mathbf{v}_3 = \left(\tfrac{3}{5}, 0, \tfrac{4}{5}\right).$$

It is easy to check that $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal basis for R^3 with the Euclidean inner product. Express the vector $\mathbf{u} = (1, 1, 1)$ as a linear combination of the vectors in S and find the coordinate vector $(\mathbf{u})_S$.

Example 6.

(a) Show that the vectors

$$\mathbf{w}_1 = (0, 2, 0), \quad \mathbf{w}_2 = (3, 0, 3), \quad \mathbf{w}_3 = (-4, 0, 4)$$

form an orthogonal basis for R^3 with the Euclidean inner product, and use that basis to find an orthonormal basis by normalizing each vector.

(b) Express the vector $\mathbf{u} = (1, 2, 4)$ as a linear combination of the orthonormal basis vectors obtained in part (a).

Theorem 6.3.3 (Projection Theorem)**.** *If W is a finite-dimensional subspace of an inner product space V, then every vector $\mathbf{u}$ in V can be expressed in exactly one way as*

$$\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2$$

where $\mathbf{w}_1$ is in W and $\mathbf{w}_2$ is in $W^\perp$.

Remark 4. The vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ in Theorem 6.3.3 are commonly denoted

$$\mathbf{w}_1 = \operatorname{proj}_W \mathbf{u} \quad \text{and} \quad \mathbf{w}_2 = \operatorname{proj}_{W^\perp} \mathbf{u}.$$

These are called the orthogonal projection of $\mathbf{u}$ on W and the orthogonal projection of $\mathbf{u}$ on $W^\perp$, respectively. The vector $\mathbf{w}_2$ is also called the component of $\mathbf{u}$ orthogonal to W. Using this notation, we can write

$$\mathbf{u} = \operatorname{proj}_W \mathbf{u} + \operatorname{proj}_{W\perp} \mathbf{u} = \operatorname{proj}_W \mathbf{u} + (\mathbf{u} - \operatorname{proj}_W \mathbf{u}).$$

Theorem 6.3.4. *Let W be a finite-dimensional subspace of an inner product space V.*

(*a*) *If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is an orthogonal basis for W, and $\mathbf{u}$ is any vector in V, then*

$$\mathrm{proj}_W\,\mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{u}, \mathbf{v}_2\rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{u}, \mathbf{v}_r\rangle}{\|\mathbf{v}_r\|^2}\mathbf{v}_r.$$

(*b*) *If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is an orthonormal basis for W, and $\mathbf{u}$ is any vector in V, then*

$$\mathrm{proj}_W\,\mathbf{u} = \langle \mathbf{u}, \mathbf{v}_1\rangle\mathbf{v}_1 + \langle \mathbf{u}, \mathbf{v}_2\rangle\mathbf{v}_1 + \cdots + \langle \mathbf{u}, \mathbf{v}_r\rangle\mathbf{v}_r.$$

Proof. (a) It follows from Theorem 6.3.3 that the vector $\mathbf{u}$ can be expressed in the form $\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{w}_1 = \mathrm{proj}_W\,\mathbf{u}$ is in W and $\mathbf{w}_2$ is in $W^\perp$; and it follows from Theorem 6.3.2 that the component $\mathrm{proj}_W\,\mathbf{u} = \mathbf{w}_1$ can be expressed in terms of the basis vectors for W as

$$\mathrm{proj}_W\,\mathbf{u} = \mathbf{w}_1 = \frac{\langle \mathbf{w}_1, \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{w}_1, \mathbf{v}_2\rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{w}_1, \mathbf{v}_r\rangle}{\|\mathbf{v}_r\|^2}\mathbf{v}_r.$$

Since $\mathbf{w}_2$ is orthogonal to W, it follows that

$$\langle \mathbf{w}_2, \mathbf{v}_1\rangle = \langle \mathbf{w}_2, \mathbf{v}_2\rangle = \cdots = \langle \mathbf{w}_2, \mathbf{v}_r\rangle = 0,$$

so we can write

$$\mathrm{proj}_W\,\mathbf{u} = \mathbf{w}_1 = \frac{\langle \mathbf{w}_1 + \mathbf{w}_2, \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{w}_1 + \mathbf{w}_2, \mathbf{v}_2\rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{w}_1 + \mathbf{w}_2, \mathbf{v}_r\rangle}{\|\mathbf{v}_r\|^2}\mathbf{v}_r$$

or, equivalently, as

$$\mathrm{proj}_W\,\mathbf{u} = \mathbf{w}_1 = \frac{\langle \mathbf{u}, \mathbf{v}_1\rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 + \frac{\langle \mathbf{u}, \mathbf{v}_2\rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 + \cdots + \frac{\langle \mathbf{u}, \mathbf{v}_r\rangle}{\|\mathbf{v}_r\|^2}\mathbf{v}_r.$$

(b) Here $\|\mathbf{v}_1\| = \|\mathbf{v}_2\| = \cdots = \|\mathbf{v}_r\| = 1$, so part (a) simplifies to part (b). □

Example 7. Let R^3 have the Euclidean inner product, and let W be the subspace spanned by the orthonormal vectors $\mathbf{v}_1 = (0, 1, 0)$ and $\mathbf{v}_2 = (-\frac{4}{5}, 0, \frac{3}{5})$. Find the orthogonal projection of $\mathbf{u} = (1, 1, 1)$ on W and the component of $\mathbf{u}$ orthogonal to W.

Theorem 6.3.5. *Every nonzero finite-dimensional inner product space has an orthonormal basis.*

Proof. Let W be any nonzero finite-dimensional subspace of an inner product space, and suppose that $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ is any basis for W. It suffices to show that W has an orthogonal basis since the vectors in that basis can be normalized to obtain an orthonormal basis. The following sequence of steps will produce an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ for W:

Step 1. Let $\mathbf{v}_1 = \mathbf{u}_1$.

Step 2. We can obtain a vector $\mathbf{v}_2$ that is orthogonal to $\mathbf{v}_1$ by computing the component of $\mathbf{u}_2$ that is orthogonal to the space W_1 spanned by $\mathbf{v}_1$. Using Theorem 6.3.4,

$$\mathbf{v}_2 = \mathbf{u}_2 - \operatorname{proj}_{W_1} \mathbf{u}_2 = \mathbf{u}_2 - \frac{\langle \mathbf{u}_2, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1.$$

Of course, if $\mathbf{v}_2 = \mathbf{0}$, then $\mathbf{v}_2$ is not a basis vector. But this cannot happen, since it would then follow from the preceding formula for $\mathbf{v}_2$ that

$$\mathbf{u}_2 = \frac{\langle \mathbf{u}_2, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 = \frac{\langle \mathbf{u}_2, \mathbf{v}_1 \rangle}{\|\mathbf{u}_1\|^2}\mathbf{u}_1$$

which implies that $\mathbf{u}_2$ is a multiple of $\mathbf{u}_1$, contradicting the linear independence of the basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$.

Step 3. To construct a vector $\mathbf{v}_3$ that is orthogonal to both $\mathbf{v}_1$ and $\mathbf{v}_2$, we compute the component of $\mathbf{u}_3$ orthogonal to the space W_2 spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$. Using Theorem 6.3.4,

$$\mathbf{v}_3 = \mathbf{u}_3 - \operatorname{proj}_{W_2} \mathbf{u}_3 = \mathbf{u}_3 - \frac{\langle \mathbf{u}_3, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\langle \mathbf{u}_3, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2.$$

As in Step 2, the linear independence of $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\}$ ensures that $\mathbf{v}_3 \neq \mathbf{0}$.

Step 4. To determine a vector $\mathbf{v}_4$ that is orthogonal to $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$, we compute the component of $\mathbf{u}_4$ orthogonal to the space W_3 spanned by $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$. Using Theorem 6.3.4,

$$\mathbf{v}_4 = \mathbf{u}_4 - \operatorname{proj}_{W_3} \mathbf{u}_4 = \mathbf{u}_4 - \frac{\langle \mathbf{u}_4, \mathbf{v}_1 \rangle}{\|\mathbf{v}_1\|^2}\mathbf{v}_1 - \frac{\langle \mathbf{u}_4, \mathbf{v}_2 \rangle}{\|\mathbf{v}_2\|^2}\mathbf{v}_2 - \frac{\langle \mathbf{u}_4, \mathbf{v}_3 \rangle}{\|\mathbf{v}_3\|^2}\mathbf{v}_3.$$

Continuing in this way we will produce after r steps an orthogonal set of nonzero vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$. Since such sets are linearly independent, we will have produced an orthogonal basis for the r-dimensional space W. By normalizing these basis vectors we can obtain an orthonormal basis. $\square$

Example 8. Assume that the vector space R^3 has the Euclidean inner product. Apply the Gram-Schmidt process to transform the basis vectors

$$\mathbf{u}_1 = (1, 1, 1), \quad \mathbf{u}_2 = (0, 1, 1), \quad \mathbf{u}_3 = (0, 0, 1)$$

into an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$, and then normalize the orthogonal basis vectors to obtain an orthonormal basis $\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\}$.

Example 9. Let the vector space P_2 have the inner product

$$\langle \mathbf{p}, \mathbf{q} \rangle = \int_{-1}^{1} p(x)q(x)\,dx.$$

Apply the Gram-Schmidt process to transform the standard basis $\{1, x, x^2\}$ for P_2 into an orthogonal basis $\{\phi_1(x), \phi_2(x), \phi_3(x)\}$.

Theorem 6.3.6. *If W is a finite-dimensional inner product space, then:*

(a) *Every orthogonal set of nonzero vectors in W can be enlarged to an orthogonal basis for W.*

(b) *Every orthonormal set in W can be enlarged to an orthonormal basis for W.*

Theorem 6.3.7 (QR-Decomposition)**.** *If A is an $m \times n$ matrix with linearly independent column vectors, then A can be factored as*

$$A = QR$$

where Q is an $m \times n$ matrix with orthonormal column vectors, and R is an $n \times n$ invertible upper triangular matrix.

Example 10. Find a QR-decomposition of

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

6.4 Best Approximation; Least Squares

Theorem 6.4.1 (Best Approximation Theorem)**.** *If W is a finite-dimensional subspace of an inner product space V, and if $\mathbf{b}$ is a vector in V, then $\operatorname{proj}_W \mathbf{b}$ is the best approximation to $\mathbf{b}$ from W in the sense that*

$$\|\mathbf{b} - \operatorname{proj}_W \mathbf{b}\| < \|\mathbf{b} - \mathbf{w}\|$$

for every vector $\mathbf{w}$ in W that is different from $\operatorname{proj}_W \mathbf{b}$.

Proof. For every vector $\mathbf{w}$ in W, we can write

$$\mathbf{b} - \mathbf{w} = (\mathbf{b} - \operatorname{proj}_W \mathbf{b}) + (\operatorname{proj}_W \mathbf{b} - \mathbf{w}).$$

But $\operatorname{proj}_W \mathbf{b} - \mathbf{w}$, being a difference of vectors in W, is itself in W; and since $\mathbf{b} - \operatorname{proj}_W \mathbf{b}$ is orthogonal to W, the two terms on the right side of the equation are orthogonal. Thus, it follows from the Theorem of Pythagoras that

$$\|\mathbf{b} - \mathbf{w}\|^2 = \|\mathbf{b} - \operatorname{proj}_W \mathbf{b}\|^2 + \|\operatorname{proj}_W \mathbf{b} - \mathbf{w}\|^2.$$

If $\mathbf{w} \neq \operatorname{proj}_W \mathbf{b}$, it follows that the second term in this sum is positive, and hence that

$$\|\mathbf{b} - \operatorname{proj}_W \mathbf{b}\|^2 < \|\mathbf{b} - \mathbf{w}\|^2.$$

Since norms are nonnegative, it follows that

$$\|\mathbf{b} - \operatorname{proj}_W \mathbf{b}\| < \|\mathbf{b} - \mathbf{w}\|. \qquad \square$$

Theorem 6.4.2. *For every linear system $A\mathbf{x} = \mathbf{b}$, the associated normal system*

$$A^T A\mathbf{x} = A^T\mathbf{b}$$

is consistent, and all solutions are least squares solutions of $A\mathbf{x} = \mathbf{b}$. Moreover, if W is the column space of A, and $\mathbf{x}$ is any least squares solution of $A\mathbf{x} = \mathbf{b}$, then the orthogonal projection of $\mathbf{b}$ on W is

$$\operatorname{proj}_W \mathbf{b} = A\mathbf{x}.$$

Example 1. Find the least squares solution, the least squares error vector, and the least squares error of the linear system

$$\begin{aligned} x_1 - x_2 &= 4 \\ 3x_1 + 2x_2 &= 1 \\ -2x_1 + 4x_2 &= 3. \end{aligned}$$

Example 2. Find the least squares solutions, the least squares error vector, and the least squares error of the linear system

$$\begin{aligned} 3x_1 + \ \ 2x_2 - \ \ x_3 &= \ \ 2 \\ x_1 - \ \ 4x_2 + 3x_3 &= -2 \\ x_1 + 10x_2 - 7x_3 &= \ \ 1. \end{aligned}$$

Theorem 6.4.3. *If A is an $m \times n$ matrix, then the following are equivalent.*

(*a*) *The column vectors of A are linearly independent.*
(*b*) *$A^T A$ is invertible.*

Theorem 6.4.4. *If A is an $m \times n$ matrix with linearly independent column vectors, then for every $m \times 1$ matrix $\mathbf{b}$, the linear system $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution. This solution is given by*

$$\mathbf{x} = (A^T A)^{-1} A^T \mathbf{b}.$$

Moreover, if W is the column space of A, then the orthogonal projection of $\mathbf{b}$ on W is

$$\operatorname{proj}_W \mathbf{b} = A\mathbf{x} = A(A^T A)^{-1} A^T \mathbf{b}.$$

Example 3. Use Theorem 6.4.4 to find the least squares solution of the linear system in Example 1.

Example 4. We showed in Section 3.3 that the standard matrix for the orthogonal projection onto the line W through the origin of R^2 that makes an angle θ with the positive x-axis is

$$P_\theta = \begin{bmatrix} \cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{bmatrix}.$$

Derive this result using Theorem 6.4.4.

Theorem 6.4.5 (Equivalent Statements)**.** *If A is an $n \times n$ matrix, then the following statements are equivalent.*

(a) *A is invertible.*
(b) *$A\mathbf{x} = \mathbf{0}$ has only the trivial solution.*
(c) *The reduced row echelon form of A is I_n.*
(d) *A is expressible as a product of elementary matrices.*
(e) *$A\mathbf{x} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.*
(f) *$A\mathbf{x} = \mathbf{b}$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.*
(g) $\det(A) \neq 0$.
(h) *The column vectors of A are linearly independent.*
(i) *The row vectors of A are linearly independent.*
(j) *The column vectors of A span R^n.*
(k) *The row vectors of A span R^n.*
(l) *The column vectors of A form a basis for R^n.*
(m) *The row vectors of A form a basis for R^n.*
(n) *A has rank n.*
(o) *A has nullity 0.*
(p) *The orthogonal complement of the null space of A is R^n.*
(q) *The orthogonal complement of the row space of A is $\{\mathbf{0}\}$.*
(r) *$\lambda = 0$ is not an eigenvalue of A.*
(s) *$A^T A$ is invertible.*

Theorem 6.4.6. *If A is an $m \times n$ matrix with linearly independent column vectors, and if $A = QR$ is a QR-decomposition of A, then for each $\mathbf{b}$ in R^m the system $A\mathbf{x} = \mathbf{b}$ has a unique least squares solution given by*

$$\mathbf{x} = R^{-1}Q^T\mathbf{b}.$$

6.5 Mathematical Modeling Using Least Squares

Theorem 6.5.1 (Uniqueness of the Least Squares Solution). *Let* $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ *be a set of two or more data points, not all lying on a vertical line, and let*

$$M = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad \textit{and} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Then there is a unique least squares straight line fit

$$y = a^* + b^* x$$

to the data points. Moreover,

$$\mathbf{v}^* = \begin{bmatrix} a^* \\ b^* \end{bmatrix}$$

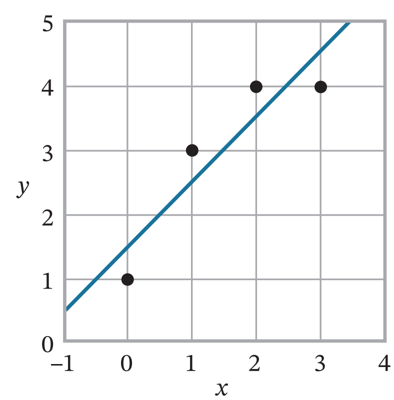
is given by the formula

$$\mathbf{v}^* = (M^T M)^{-1} M^T \mathbf{y}$$

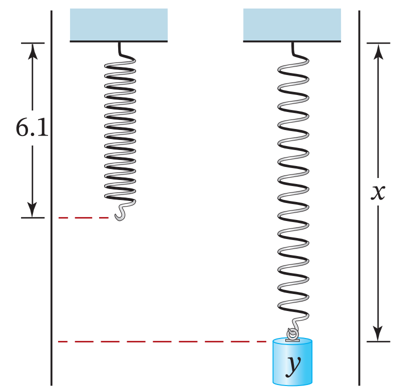
which expresses the fact that $\mathbf{v} = \mathbf{v}^*$ *is the unique solution of the normal equation*

$$M^T M \mathbf{v} = M^T \mathbf{y}.$$

Example 1. Find the least squares straight line fit to the four points $(0, 1)$, $(1, 3)$, $(2, 4)$, and $(3, 4)$. (See the figure.)



Example 2. Hooke's law in physics states that the length x of a uniform spring is a linear function of the force y applied to it. If we express this relationship as $y = a + bx$, then the coefficient b is called the spring constant. Suppose a particular unstretched spring has a measured length of 6.1 inches (i.e., $x = 6.1$ when $y = 0$). Suppose further that, as illustrated in the figure, various weights are attached to the end of the spring and the following table of resulting spring lengths is recorded. Find the least squares straight line fit to the data and use it to approximate the spring constant.



Weight y (lb)	0	2	4	6
Length x (in)	6.1	7.6	8.7	10.4

Example 3. According to Newton's second law of motion, a body near the Earth's surface falls vertically downward in accordance with the equation

$$s = s_0 + v_0 t + \frac{1}{2} g t^2$$

where

$$\begin{aligned} s &= \text{vertical displacement downward relative to some reference point} \\ s_0 &= \text{displacement from the reference point at time } t = 0 \\ v_0 &= \text{velocity at time } t = 0 \\ g &= \text{acceleration of gravity at the Earth's surface.} \end{aligned}$$

Suppose that a laboratory experiment is performed to approximate g by measuring the displacement s relative to a fixed reference point of a falling weight at various times. Use the experimental results shown in the following table to approximate g.

Time t (sec)	.1	.2	.3	.4	.5
Displacement s (ft)	-0.18	0.31	1.03	2.48	3.73

6.6 Function Approximation; Fourier Series

Theorem 6.6.1. *If* $\mathbf{f}$ *is a continuous function on* $[a, b]$*, and* W *is a finite-dimensional subspace of* $C[a, b]$*, then the function* $\mathbf{g}$ *in* W *that minimizes the* *mean square error*

$$\int_a^b [f(x) - g(x)]^2 \, dx$$

is $\mathbf{g} = \operatorname{proj}_W \mathbf{f}$*, where the orthogonal projection is relative to the inner product*

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_a^b f(x) g(x) \, dx.$$

The function $\mathbf{g} = \operatorname{proj}_W \mathbf{f}$ *is called the* *least squares approximation* *to* $\mathbf{f}$ *from* W.

Remark 1. A function of the form

$$\begin{aligned} T(x) = c_0 + c_1 \cos x + c_2 \cos 2x + \cdots + c_n \cos nx \\ + d_1 \sin x + d_2 \sin 2x + \cdots + d_n \sin nx \end{aligned}$$

is called a trigonometric polynomial; if c_n and d_n are not both zero, then $T(x)$ is said to have order n.

Remark 2. To find the least squares approximation of a continuous function $f(x)$ over the interval $[0, 2\pi]$ by a trigonometric polynomial of order n or less we use

$$\operatorname{proj}_W \mathbf{f} = \frac{a_0}{2} + [a_1 \cos x + \cdots + a_n \cos nx] + [b_1 \sin x + \cdots + b_n \sin nx]$$

where

$$a_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos kx \, dx, \quad b_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \sin kx \, dx.$$

The numbers $a_0, a_1, \ldots, a_n, b_1, \ldots, b_n$ are called the Fourier coefficients of $\mathbf{f}$.

Example 1. Find the least squares approximation of $f(x) = x$ on $[0, 2\pi]$ by

(a) a trigonometric polynomial of order 2 or less;

(b) a trigonometric polynomial of order n or less.

Chapter 7

Diagonalization and Quadratic Forms

7.1 Orthogonal Matrices

Definition 7.1.1. A square matrix A is said to be orthogonal if its transpose is the same as its inverse, that is, if

$$A^{-1} = A^T$$

or, equivalently, if

$$AA^T = A^T A = I.$$

Example 1. Determine whether the matrix

$$A = \begin{bmatrix} \frac{3}{7} & \frac{2}{7} & \frac{6}{7} \\ -\frac{6}{7} & \frac{3}{7} & \frac{2}{7} \\ \frac{2}{7} & \frac{6}{7} & -\frac{3}{7} \end{bmatrix}$$

is orthogonal.

Example 2. Recall from Table 5 of Section 1.8 that the standard matrix for the counterclockwise rotation of R^2 through an angle θ is

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

Verify that this matrix is orthogonal, along with the reflection matrices in Tables 1 and 2 of Section 1.8.

Theorem 7.1.1. *The following are equivalent for an $n \times n$ matrix A.*

(a) *A is orthogonal.*
(b) *The row vectors of A form an orthonormal set in R^n with the Euclidean inner product.*
(c) *The column vectors of A form an orthonormal set in R^n with the Euclidean inner product.*

Theorem 7.1.2.

(a) *The transpose of an orthogonal matrix is orthogonal.*
(b) *The inverse of an orthogonal matrix is orthogonal.*
(c) *A product of orthogonal matrices is orthogonal.*
(d) *If A is orthogonal, then* $\det(A) = 1$ *or* $\det(A) = -1$.

Example 3. Verify Theorem 7.1.2 (d) for the matrix

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

Theorem 7.1.3. *If A is an $n \times n$ matrix, then the following are equivalent.*
(a) *A is orthogonal.*
(b) *$\|A\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$ in R^n.*
(c) *$A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ for all $\mathbf{x}$ and $\mathbf{y}$ in R^n.*

Proof. (a) $\Rightarrow$ (b) Assume that A is orthogonal, so that $A^T A = I$. It follows that

$$\|A\mathbf{x}\| = (A\mathbf{x} \cdot A\mathbf{x})^{1/2} = (\mathbf{x} \cdot A^T A\mathbf{x})^{1/2} = (\mathbf{x} \cdot \mathbf{x})^{1/2} = \|\mathbf{x}\|.$$

(b) $\Rightarrow$ (c) Assume that $\|A\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$ in R^n. Then we have

$$\begin{aligned} A\mathbf{x} \cdot A\mathbf{y} &= \tfrac{1}{4}\|A\mathbf{x} + A\mathbf{y}\|^2 - \tfrac{1}{4}\|A\mathbf{x} - A\mathbf{y}\|^2 = \tfrac{1}{4}\|A(\mathbf{x} + \mathbf{y})\|^2 - \tfrac{1}{4}\|A(\mathbf{x} - \mathbf{y})\|^2 \\ &= \tfrac{1}{4}\|\mathbf{x} + \mathbf{y}\|^2 - \tfrac{1}{4}\|\mathbf{x} - \mathbf{y}\|^2 = \mathbf{x} \cdot \mathbf{y}. \end{aligned}$$

(c) $\Rightarrow$ (a) Assume that $A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ for all $\mathbf{x}$ and $\mathbf{y}$ in R^n. It follows that

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

which can be rewritten as $\mathbf{x} \cdot (A^T A\mathbf{y} - \mathbf{y}) = 0$ or as

$$\mathbf{x} \cdot (A^T A - I)\mathbf{y} = 0.$$

Since this equation holds for all $\mathbf{x}$ in R^n, it holds in particular if $\mathbf{x} = (A^T A - I)\mathbf{y}$, so

$$(A^T A - I)\mathbf{y} \cdot (A^T A - I)\mathbf{y} = 0.$$

Thus, it follows from the positivity axiom for inner products that

$$(A^T A - I)\mathbf{y} = \mathbf{0}.$$

Since this equation is satisfied by every vector $\mathbf{y}$ in R^n, it must be that $A^T A - I$ is the zero matrix and hence $A^T A = I$. Thus, A is orthogonal. $\square$

Theorem 7.1.4. *If S is an orthonormal basis for an n-dimensional inner product space V, and if*

$$(\mathbf{u})_S = (u_1, u_2, \ldots, u_n) \quad \textit{and} \quad (\mathbf{v})_S = (v_1, v_2, \ldots, v_n)$$

then:
(a) $\|\mathbf{u}\| = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}$
(b) $d(\mathbf{u}, \mathbf{v}) = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_n - v_n)^2}$
(c) $\langle \mathbf{u}, \mathbf{v} \rangle = u_1v_1 + u_2v_2 + \cdots + u_nv_n$

Theorem 7.1.5. *Let V be a finite-dimensional inner product space. If P is the transition matrix from one orthonormal basis for V to another orthonormal basis for V, then P is an orthogonal matrix.*

Example 4. Let the $x'y'$-coordinate system be the system obtained by rotating a rectangular xy-coordinate system counterclockwise about the origin through an angle θ. Write the coordinates (x', y') in terms of the coordinates (x, y).

Example 5. Use the rotation equations for R^2 to find the new coordinates of the point $Q(2, 1)$ if the coordinate axes of a rectangular coordinate system are rotated through an angle of $\theta = \pi/4$.

Example 6. Let the $x'y'z'$-coordinate system be the system obtained by rotating a rectangular xyz-coordinate system around its z-axis counterclockwise through an angle θ. Write the coordinates (x', y', z') in terms of the coordinates (x, y, z).

7.2 Orthogonal Diagonalization

Definition 7.2.1. If A and B are square matrices, then we say that B is orthogonally similar to A if there is an orthogonal matrix P such that $B = P^T AP$.

Remark 1. If A is orthogonally similar to some diagonal matrix, say

$$P^T AP = D$$

then we say that A is orthogonally diagonalizable and that P orthogonally diagonalizes A.

Theorem 7.2.1. *If A is an $n \times n$ matrix with real entries, then the following are equivalent.*

(*a*) *A is orthogonally diagonalizable.*
(*b*) *A has an orthonormal set of n eigenvectors.*
(*c*) *A is symmetric.*

Theorem 7.2.2. *If A is a symmetric matrix with real entries, then:*

(*a*) *The eigenvalues of A are all real numbers.*
(*b*) *Eigenvectors from different eigenspaces are orthogonal.*

Example 1. Find an orthogonal matrix P that diagonalizes

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

Remark 2. If A is a symmetric matrix that is orthogonally diagonalized by

$$P = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix}$$

and if $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A corresponding to the unit eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$, then

$$A = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T + \lambda_2 \mathbf{u}_2 \mathbf{u}_2^T + \cdots + \lambda_n \mathbf{u}_n \mathbf{u}_n^T,$$

which is called a spectral decomposition of A.

Example 2. Find a spectral decomposition of the matrix

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

Theorem 7.2.3 (Schur's Theorem)**.** *If A is an $n \times n$ matrix with real entries and real eigenvalues, then there is an orthogonal matrix P such that P^TAP is an upper triangular matrix of the form*

$$P^TAP = \begin{bmatrix} \lambda_1 & \times & \times & \cdots & \times \\ 0 & \lambda_2 & \times & \cdots & \times \\ 0 & 0 & \lambda_3 & \cdots & \times \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

in which $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A repeated according to multiplicity.

Theorem 7.2.4 (Hessenberg's Theorem)**.** *If A is an $n \times n$ matrix with real entries, then there is an orthogonal matrix P such that P^TAP is a matrix of the form*

$$P^TAP = \begin{bmatrix} \times & \times & \cdots & \times & \times & \times \\ \times & \times & \cdots & \times & \times & \times \\ 0 & \times & \ddots & \times & \times & \times \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \times & \times & \times \\ 0 & 0 & \cdots & 0 & \times & \times \end{bmatrix}.$$

7.3 Quadratic Forms

Remark 1. If $a_1, a_2, \ldots, a_n$ are treated as fixed constants, then the expression

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

is a real-valued function of the n variables $x_1, x_2, \ldots, x_n$ and is called a linear form on R^n. A quadratic form on R^n is a function of the form

$$a_1x_1^2 + a_2x_2^2 + \cdots + a_nx_n^2 + (\text{all possible terms } a_kx_ix_j \text{ in which } i \neq j).$$

The terms of the form $a_kx_ix_j$ are called cross product terms.

Remark 2. If A is a symmetric $n \times n$ matrix and $\mathbf{x}$ is an $n \times 1$ column vector of variables, then we call the function

$$Q_A(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$$

the quadratic form associated with A. When convenient, this function can be expressed in dot product notation as

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

Example 1. In each part, express the quadratic form in the matrix notation $\mathbf{x}^T A\mathbf{x}$, where A is symmetric.

(a) $2x^2 + 6xy - 5y^2$

(b) $x_1^2 + 7x_2^2 - 3x_3^2 + 4x_1x_2 - 2x_1x_3 + 8x_2x_3$

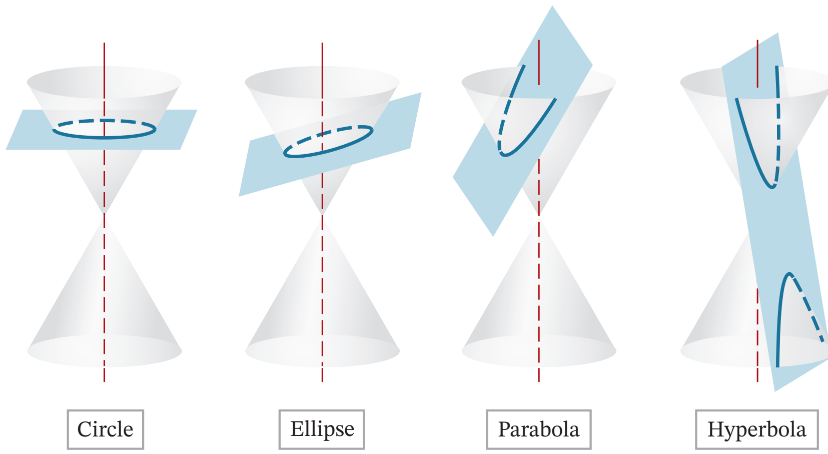
Theorem 7.3.1 (The Principal Axes Theorem). *If A is a symmetric $n \times n$ matrix, then there is an orthogonal change of variable that transforms the quadratic form $\mathbf{x}^T A\mathbf{x}$ into a quadratic form $\mathbf{y}^T D\mathbf{y}$ with no cross product terms. Specifically, if P orthogonally diagonalizes A, then making the change of variable $\mathbf{x} = P\mathbf{y}$ in the quadratic form $\mathbf{x}^T A\mathbf{x}$ yields the quadratic form*

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

in which $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A corresponding to the eigenvectors that form the successive columns of P.

Example 2. Find an orthogonal change of variable that eliminates the cross product terms in the quadratic form $Q = x_1^2 - x_3^2 - 4x_1x_2 + 4x_2x_3$, and express Q in terms of the new variables.

Remark 3. A conic section is a curve that results by cutting a double-napped cone with a plane (see the figure).



If the cutting plane passes through the vertex, then the resulting intersection is called a degenerate conic. An equation of the form

$$ax^2 + 2bxy + cy^2 + dx + ey + f = 0$$

in which a, b, and c are not all zero, represents a conic section. If $d = e = 0$, the equation becomes

$$ax^2 + 2bxy + cy^2 + f = 0$$

and is said to represent a central conic. Furthermore, if $b = 0$, the equation becomes

$$ax^2 + cy^2 + f = 0$$

and is said to represent a central conic in standard position. If we take the constant f in these equations to the right side and let $k = -f$, then we can rewrite these equations in matrix form as

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = k \quad \text{and} \quad \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & 0 \\ 0 & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = k.$$

The three-dimensional analogs of these equations are

$$\begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} a & d & e \\ d & b & f \\ e & f & c \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = k \quad \text{and} \quad \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = k.$$

If a, b, and c are not all zero, the the graphs in R^3 of these equations are called central quadrics; the graph of the second of these equations, which is a special case of the first, is called a central quadric in standard position.

Example 3.

(a) Identify the conic whose equation is $5x^2 - 4xy + 8y^2 - 36 = 0$ by rotating the xy-axes to put the conic in standard position.

(b) Find the angle θ through which you rotated the xy-axes in part (a).

Definition 7.3.1. A quadratic form $\mathbf{x}^T A\mathbf{x}$ is said to be
positive definite if $\mathbf{x}^T A\mathbf{x} > 0$ for $\mathbf{x} \neq \mathbf{0}$;
negative definite if $\mathbf{x}^T A\mathbf{x} < 0$ for $\mathbf{x} \neq \mathbf{0}$;
indefinite if $\mathbf{x}^T A\mathbf{x}$ has both positive and negative values.

Theorem 7.3.2. *If A is a symmetric matrix, then:*

(a) $\mathbf{x}^T A\mathbf{x}$ *is positive definite if and only if all eigenvalues of A are positive.*
(b) $\mathbf{x}^T A\mathbf{x}$ *is negative definite if and only if all eigenvalues of A are negative.*
(c) $\mathbf{x}^T A\mathbf{x}$ *is indefinite if and only if A has at least one positive eigenvalue and at least one negative eigenvalue.*

Example 4. Determine whether the matrix

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

is positive definite, negative definite, indefinite, or none of these.

Remark 4. The kth principal submatrix of an $n \times n$ matrix A is the $k \times k$ submatrix consisting of the first k rows and columns of A.

Theorem 7.3.3. *If A is a symmetric 2×2 matrix, then:*

(a) $\mathbf{x}^T A\mathbf{x} = 1$ *represents an ellipse if A is positive definite.*
(b) $\mathbf{x}^T A\mathbf{x} = 1$ *has no graph if A is negative definite.*
(c) $\mathbf{x}^T A\mathbf{x} = 1$ *represents a hyperbola if A is indefinite.*

Theorem 7.3.4. *If A is a symmetric matrix, then:*

(a) *A is positive definite if and only if the determinant of every principal submatrix is positive.*
(b) *A is negative definite if and only if the determinants of the principal submatrices alternate between negative and positive values starting with a negative value for the determinant of the first principal submatrix.*
(c) *A is indefinite if and only if it is neither positive definite nor negative definite and at least one principal submatrix has a positive determinant and at least one has a negative determinant.*

Example 5. Determine whether the matrix

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

is positive definite, negative definite, indefinite, or none of these.

7.4 Optimization Using Quadratic Forms

Theorem 7.4.1 (Constrained Extremum Theorem)**.** *Let A be a symmetric $n \times n$ matrix whose eigenvalues in order of decreasing size are $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Then:*

(a) *The quadratic form $\mathbf{x}^T A\mathbf{x}$ attains a maximum value and a minimum value on the set of vectors for which $\|\mathbf{x}\| = 1$.*

(b) *The maximum value attained in part (a) occurs at a vector corresponding to the eigenvalue λ_1.*

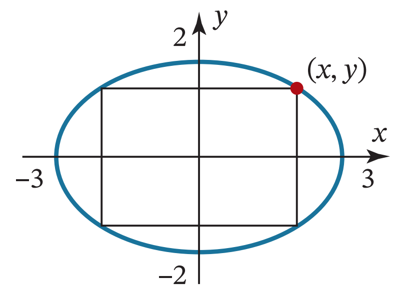
(c) *The minimum value attained in part (a) occurs at a vector corresponding to the eigenvalue λ_n.*

Example 1. Find the maximum and minimum values of the quadratic form

$$z = 5x^2 + 5y^2 + 4xy$$

subject to the constraint $x^2 + y^2 = 1$.

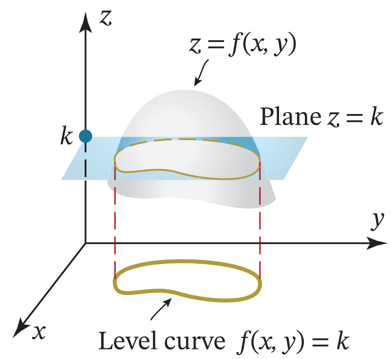
Example 2. A rectangle is to be inscribed in the ellipse $4x^2 + 9y^2 = 36$, as shown in the figure. Use eigenvalue methods to find nonnegative values of x and y that produce the inscribed rectangle with maximum area.



Remark 1. The curves in the xy-plane for which the function $f(x, y)$ is constant have equations of the form

$$f(x, y) = k$$

and are called the level curves of f (see the figure).



Example 3. Geometrically interpret the level curves of the quadratic form

$$z = 5x^2 + 5y^2 + 4xy$$

subject to the constraint $x^2 + y^2 = 1$.

Remark 2. If a function $f(x, y)$ has first-order partial derivatives, then its relative maxima and minima, if any, occur at points where the conditions

$$f_x(x, y) = 0 \quad \text{and} \quad f_y(x, y) = 0$$

are both true. These are called critical points of f. The specific behavior of f at a critical point (x_0, y_0) is determined by the sign of

$$D(x, y) = f(x, y) - f(x_0, y_0)$$

at points (x, y) that are close to, but different from, (x_0, y_0):

- If $D(x, y) > 0$ at points (x, y) that are sufficiently close to, but different from, (x_0, y_0), then $f(x_0, y_0) < f(x, y)$ at such points and f is said to have a relative minimum at (x_0, y_0).
- If $D(x, y) < 0$ at points (x, y) that are sufficiently close to, but different from, (x_0, y_0), then $f(x_0, y_0) > f(x, y)$ at such points and f is said to have a relative maximum at (x_0, y_0).
- If $D(x, y)$ has both positive and negative values inside *every* circle centered at (x_0, y_0), then are points (x, y) that are arbitrarily close to (x_0, y_0) at which $f(x_0, y_0) < f(x, y)$ and points (x, y) that are arbitrarily close to (x_0, y_0) at which $f(x_0, y_0) > f(x, y)$. In this case we say that f has a saddle point at (x_0, y_0).

Theorem 7.4.2 (Second Derivative Test)**.** *Suppose that (x_0, y_0) is a critical point of $f(x, y)$ and that f has continuous second-order partial derivatives in some circular region centered at (x_0, y_0). Then:*

(*a*) *f has a relative minimum at (x_0, y_0) if*

$$f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) > 0 \quad \textit{and} \quad f_{xx}(x_0, y_0) > 0$$

(*b*) *f has a relative maximum at (x_0, y_0) if*

$$f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) > 0 \quad \textit{and} \quad f_{xx}(x_0, y_0) < 0$$

(*c*) *f has a saddle point at (x_0, y_0) if*

$$f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) < 0$$

(*d*) *The test is inconclusive if*

$$f_{xx}(x_0, y_0)f_{yy}(x_0, y_0) - f_{xy}^2(x_0, y_0) = 0$$

Remark 3. The symmetric matrix

$$H(x, y) = \begin{bmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{xy}(x, y) & f_{yy}(x, y) \end{bmatrix}$$

is called the Hessian or Hessian matrix of f.

Theorem 7.4.3 (Hessian Form of the Second Derivative Test)**.** *Suppose that (x_0, y_0) is a critical point of $f(x, y)$ and that f has continuous second-order partial derivatives in some circular region centered at (x_0, y_0). If $H(x_0, y_0)$ is the Hessian of f at $x_0, y_0)$, then:*

(*a*) *f has a relative minimum at (x_0, y_0) if $H(x_0, y_0)$ is positive definite.*
(*b*) *f has a relative maximum at (x_0, y_0) if $H(x_0, y_0)$ is negative definite.*
(*c*) *f has a saddle point at (x_0, y_0) if $H(x_0, y_0)$ is indefinite.*
(*d*) *The test is inconclusive otherwise.*

Example 4. Find the critical points of the function

$$f(x,y) = \frac{1}{3}x^3 + xy^2 - 8xy + 3$$

and use the eigenvalues of the Hessian matrix at those points to determine which of them, if any, are relative maxima, relative minima, or saddle points.

7.5 Hermitian, Unitary, and Normal Matrices

Definition 7.5.1. If A is a complex matrix, then the conjugate transpose of A, denoted by A^*, is defined by

$$A^* = \overline{A}^T.$$

Example 1. Find the conjugate transpose A^* of the matrix

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

Theorem 7.5.1. *If k is a complex scalar, and if A and B are complex matrices whose sizes are such that the stated operations can be performed, then:*
(a) $(A^*)^* = A$
(b) $(A+B)^* = A^* + B^*$
(c) $(A-B)^* = A^* - B^*$
(d) $(kA)^* = \overline{k}A^*$
(e) $(AB)^* = B^*A^*$

Definition 7.5.2. A square matrix A is said to be unitary if

$$AA^* = A^*A = I$$

or, equivalently, if

$$A^* = A^{-1}$$

and it is said to be Hermitian if

$$A^* = A.$$

Example 2. Determine whether the matrix

$$A = \begin{bmatrix} 1 & i & 1+i \\ -i & -5 & 2-i \\ 1-i & 2+i & 3 \end{bmatrix}$$

is Hermitian.

Example 3. Determine whether the matrix

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}i \\ -\frac{1}{\sqrt{2}}i & \frac{1}{\sqrt{2}} \end{bmatrix}$$

is unitary.

Theorem 7.5.2. *If A is a Hermitian matrix, then:*

(a) *The eigenvalues of A are all real numbers.*

(b) *Eigenvalues from different eigenspaces are orthogonal.*

Example 4. Confirm that the Hermitian matrix

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

has real eigenvalues and that eigenvectors from different eigenspaces are orthogonal.

Theorem 7.5.3. *If A is an $n \times n$ matrix with complex entries, then the following are equivalent.*

(a) *A is unitary.*

(b) *$\|A\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x}$ in C^n.*

(c) *$A\mathbf{x} \cdot A\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ for all $\mathbf{x}$ and $\mathbf{y}$ in C^n.*

(d) *The column vectors of A form an orthonormal set in C^n with respect to the complex Euclidean inner product.*

(e) *The row vectors of A form an orthonormal set in C^n with respect to the complex Euclidean inner product.*

Example 5. Use Theorem 7.5.3 to show that

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

is unitary, and then find A^{-1}.

Definition 7.5.3. A square complex matrix A is said to be unitarily diagonalizable if there is a unitary matrix P such that $P^*AP = D$ is a complex diagonal matrix. Any such matrix P is said to unitarily diagonalize A.

Theorem 7.5.4. *Every $n \times n$ Hermitian matrix A has an orthonormal set of n eigenvectors and is unitarily diagonalized by any $n \times n$ matrix P whose column vectors form an orthonormal set of eigenvectors of A.*

Example 6. Find a matrix P that unitarily diagonalizes the Hermitian matrix

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

Remark 1. A square *real* matrix A is said to be skew-symmetric if $A^T = -A$, and a square *complex* matrix A is said to be skew-Hermitian if $A^* = -A$. Matrices with the property

$$AA^* = A^*A$$

are said to be normal.

Chapter 8

General Linear Transformations

8.1 General Linear Transformations

Definition 8.1.1. If $T : V \to W$ is a mapping from a vector space V to a vector space W, then T is called a linear transformation from V to W if the following two properties hold for all vectors $\mathbf{u}$ and $\mathbf{v}$ in V and for all scalars k:

(i) $T(k\mathbf{u}) = kT(\mathbf{u})$
(ii) $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$

In the special case where $V = W$, the linear transformation T is called a linear operator on the vector space V.

Theorem 8.1.1. *If $T : V \to W$ is a linear transformation, then:*
(a) $T(\mathbf{0}) = \mathbf{0}$
(b) $T(\mathbf{u}-\mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v})$ *for all* $\mathbf{u}$ *and* $\mathbf{v}$ *in* V.

Proof. Let $\mathbf{u}$ be any vector in V. Since $0\mathbf{u} = \mathbf{0}$, it follows that

$$T(\mathbf{0}) = T(0\mathbf{u}) = 0T(\mathbf{u}) = \mathbf{0}$$

which proves (a).

We can prove part (b) by rewriting $T(\mathbf{u}-\mathbf{v})$ as

$$\begin{aligned} T(\mathbf{u}-\mathbf{v}) &= T(\mathbf{u}+(-1)\mathbf{v}) \\ &= T(\mathbf{u}) + (-1)T(\mathbf{v}) \\ &= T(\mathbf{u}) - T(\mathbf{v}). \end{aligned}$$

□

Example 1. Verify that every matrix transformation $T_A : R^n \to R^m$ is also a linear transformation.

Example 2. Let V and W be any two vector spaces. Verify that the mapping $T : V \to W$ such that $T(\mathbf{v}) = \mathbf{0}$ for every $\mathbf{v}$ is a linear transformation, called the zero transformation.

Example 3. Let V be any vector space. Verify that the mapping $I : V \to V$ such that $I(\mathbf{v}) = \mathbf{v}$ is a linear transformation, called the identity operator.

Example 4. If V is a vector space and c is any scalar, then verify that the mapping $T : V \to V$ given by $T(\mathbf{x}) = c\mathbf{x}$ is a linear operator on V. If $0 < c < 1$, then T is called the contraction of V with factor c, and if $c > 1$, it is called the dilation of V with factor c.

Example 5. Let $\mathbf{p} = p(x) = c_0 + c_1x + \cdots + c_nx^n$ be a polynomial in P_n, and define the transformation $T : P_n \to P_{n+1}$ by

$$T(\mathbf{p}) = T(p(x)) = xp(x) = c_0x + c_1x^2 + \cdots + c_nx^{n+1}.$$

Verify that T is linear.

Example 6. Let $\mathbf{v}_0$ be any fixed vector in a real inner product space V, and let $T : V \to R$ be the transformation

$$T(\mathbf{x}) = \langle \mathbf{x}, \mathbf{v}_0 \rangle$$

that maps a vector $\mathbf{x}$ to its inner product with $\mathbf{v}_0$. Verify this transformation is linear.

Example 7. Let M_{nn} be the vector space of $n \times n$ matrices. In each part determine whether the transformation is linear.

(a) $T_1(A) = A^T$

(b) $T_2(A) = \det(A)$

Example 8. If $\mathbf{x}_0$ is a fixed nonzero vector in a real inner product space V, determine whether the transformation

$$T(\mathbf{x}) = \mathbf{x} + \mathbf{x}_0$$

is linear.

Example 9. Let V be a subspace of $F(-\infty, \infty)$, let

$$x_1, x_2, \ldots, x_n$$

be a sequence of distinct real numbers, and let $T : V \to R^n$ be the transformation

$$T(f) = (f(x_1), f(x_2), \ldots, f(x_n))$$

that associates with f the n-tuple of function values at $x_1, x_2, \ldots, x_n$. We call this the evaluation transformation on V at $x_1, x_2, \ldots, x_n$. Verify that the evaluation transformation is linear.

Theorem 8.1.2. *Let $T : V \to W$ be a linear transformation, where V is finite-dimensional. If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis for V, then the image of any vector $\mathbf{v}$ in V can be expressed as*

$$T(\mathbf{v}) = c_1 T(\mathbf{v}_1) + c_2 T(\mathbf{v}_2) + \cdots + c_n T(\mathbf{v}_n)$$

where $c_1, c_2, \ldots, c_n$ are the coefficients required to express $\mathbf{v}$ as a linear combination of the vectors in the basis S.

Example 10. Consider the basis $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ for R^3, where

$$\mathbf{v}_1 = (1, 1, 1), \quad \mathbf{v}_2 = (1, 1, 0), \quad \mathbf{v}_3 = (1, 0, 0).$$

Let $T : R^3 \to R^2$ be the linear transformation for which

$$T(\mathbf{v}_1) = (1, 0), \quad T(\mathbf{v}_2) = (2, -1), \quad T(\mathbf{v}_3) = (4, 3).$$

Find a formula for $T(x_1, x_2, x_3)$, and then use that formula to compute $T(2, -3, 5)$.

Example 11. Let $V = C^1(-\infty, \infty)$ be the vector space of functions with continuous first derivatives on $(-\infty, \infty)$, and let $W = F(-\infty, \infty)$ be the vector space of all real-valued functions defined on $(-\infty, \infty)$. Let $D : V \to W$ be the transformation that maps a function $\mathbf{f} = f(x)$ into its derivative—that is,

$$D(\mathbf{f}) = f'(x).$$

Verify that D is a linear transformation.

Example 12. Let $V = C(-\infty, \infty)$ be the vector space of continuous functions on the interval $(-\infty, \infty)$, let $W = C^1(-\infty, \infty)$ be the vector space of functions with continuous first derivatives on $(-\infty, \infty)$, and let $J : V \to W$ be the transformation that maps a function f in V into

$$J(f) = \int_0^x f(t)\,dt.$$

Verify that J is a linear transformation.

Definition 8.1.2. If $T : V \to W$ is a linear transformation, then the set of vectors in V that maps into $\mathbf{0}$ is called the kernel of T and is denoted by $\ker(T)$. The set of all vectors in W that are images under T of at least one vector in V is called the range of T and is denoted by $R(T)$.

Example 13. If $T_A : R^n \to R^m$ is multiplication by the $m \times n$ matrix A, then what are the kernel and range of T_A?

Example 14. Let $T : V \to W$ be the zero transformation. What are the kernel and range of T?

Example 15. Let $I : V \to V$ be the identity operator. What are the kernel and range of I?

Example 16. Let $T : R^3 \to R^3$ be the orthogonal projection onto the xy-plane. What are the kernel and range of T?

Example 17. Let $T : R^2 \to R^2$ be the linear operator that rotates each vector in the xy-plane through some angle θ. What are the kernel and range of T?

Example 18. Let $V = C^1(-\infty, \infty)$ be the vector space of functions with continuous first derivatives on $(-\infty, \infty)$, let $W = F(-\infty, \infty)$ be the vector space of all real-valued functions defined on $(-\infty, \infty)$, and let $D : V \to W$ be the differentiation transformation $D(\mathbf{f}) = f'(x)$. What is the kernel of D?

Theorem 8.1.3. *If $T : V \to W$ is a linear transformation, then:*
(a) *The kernel of T is a subspace of V.*
(b) *The range of T is a subspace of W.*

Proof. (a) To show that $\ker(T)$ is a subspace, we must show that it contains at least one vector and is closed under addition and scalar multiplication. By part (a) of Theorem 8.1.1, the vector $\mathbf{0}$ is in $\ker(T)$, so the kernel contains at least one vector. Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be vectors in $\ker(T)$, and let k be any scalar. Then

$$T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2) = \mathbf{0} + \mathbf{0} = \mathbf{0},$$

so $\mathbf{v}_1 + \mathbf{v}_2$ is in $\ker(T)$. Also,

$$T(k\mathbf{v}_1) = kT(\mathbf{v}_1) = k\mathbf{0} = \mathbf{0},$$

so $k\mathbf{v}_1$ is in $\ker(T)$.
(b) To show that $R(T)$ is a subspace of W, we must show that it contains at least one vector and is closed under addition and scalar multiplication. However, it contains at least the zero vector of W since $T(\mathbf{0}) = (\mathbf{0})$. To prove that it is closed under addition and scalar multiplication, we must show that if $\mathbf{w}_1$ and $\mathbf{w}_2$ are vectors in $R(T)$, and if k is any scalar, then there exist vectors $\mathbf{a}$ and $\mathbf{b}$ in V for which

$$T(\mathbf{a}) = \mathbf{w}_1 + \mathbf{w}_2 \quad \text{and} \quad T(\mathbf{b}) = k\mathbf{w}_1.$$

But the fact that $\mathbf{w}_1$ and $\mathbf{w}_2$ are in $R(T)$ tells us there exist vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ in V such that

$$T(\mathbf{v}_1) = \mathbf{w}_1 \quad T(\mathbf{v}_2) = \mathbf{w}_2.$$

The following computations complete the proof by showing that the vectors $\mathbf{a} = \mathbf{v}_1 + \mathbf{v}_2$ and $\mathbf{b} = k\mathbf{v}_1$ satisfy the desired equations:

$$T(\mathbf{a}) = T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2) = \mathbf{w}_1 + \mathbf{w}_2$$
$$T(\mathbf{b}) = T(k\mathbf{v}_1) = kT(\mathbf{v}_1) = k\mathbf{w}_1.$$

□

Example 19. Differential equations of the form

$$y'' + \omega^2 y = 0 \quad (\omega \text{ a positive constant})$$

arise in the study of vibrations. Confirm that

$$y_1 = \cos \omega x \quad \text{and} \quad y_2 = \sin \omega x$$

are solutions of these differential equations, and use them to find a general solution.

Definition 8.1.3. Let $T : V \to W$ be a linear transformation. If the range of T is finite-dimensional, then its dimension is called the rank of T; and if the kernel of T is finite-dimensional, then its dimension is called the nullity of T. The rank of T is denoted by $\text{rank}(T)$ and the nullity of T by $\text{nullity}(T)$.

Theorem 8.1.4 (Dimension Theorem for Linear Transformations)**.** *If $T : V \to W$ is a linear transformation from a finite-dimensional vector space V to a vector space W, then the range of T is finite-dimensional, and*

$$\text{rank}(T) + \text{nullity}(T) = \dim(V).$$

8.2 Compositions and Inverse Transformations

Definition 8.2.1. If $T : V \to W$ is a linear transformation from a vector space V to a vector space W, then T is said to be one-to-one if T maps distinct vectors in V into distinct vectors in W.

Definition 8.2.2. If $T : V \to W$ is a linear transformation from a vector space V to a vector space W, then T is said to be onto (or onto W) if every vector in W is the image of at least one vector in V.

Theorem 8.2.1. *If $T : V \to W$ is a linear transformation, then the following statements are equivalent.*
(a) *T is one-to-one.*
(b) $\ker(T) = \{\mathbf{0}\}$.

Proof. (a) $\Rightarrow$ (b) Since T is linear, we know that $T(\mathbf{0}) = \mathbf{0}$. Since T is one-to-one, there can be no other vectors in V that map into $\mathbf{0}$, so $\ker(T) = \{\mathbf{0}\}$.
(b) $\Rightarrow$ (a) Assume that $\ker(T) = \{\mathbf{0}\}$. If $\mathbf{u}$ and $\mathbf{v}$ are distinct vectors in V, then $\mathbf{u} - \mathbf{v} \neq \mathbf{0}$. This implies that $T(\mathbf{u} - \mathbf{v}) \neq \mathbf{0}$, for otherwise $\ker(T)$ would contain a nonzero vector. Since T is linear, it follows that

$$T(\mathbf{u}) - T(\mathbf{v}) = T(\mathbf{u} - \mathbf{v}) \neq \mathbf{0},$$

so T maps distinct vectors in V into distinct vectors in W and hence is one-to-one. $\square$

Example 1. Let $T : R^2 \to R^2$ be the linear operator that rotates each vector in the plane about the origin through an angle θ. Is T one-to-one? Is T onto?

Example 2. Let $T : R^2 \to R^2$ be the linear operator that maps points orthogonally on to the x-axis in R^2. Is T one-to-one? Is T onto?

Example 3. Verify that the linear transformations $T_1 : P_3 \to R^4$ and $T_2 : M_{22} \to R^4$ defined by

$$T_1(a + bx + cx^2 + dx^3) = (a, b, c, d)$$
$$T_2\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = (a, b, c, d)$$

are both one-to-one and onto.

Example 4. Let $T : P_n \to P_{n+1}$ be the linear transformation

$$T(\mathbf{p}) = T(p(x)) = xp(x)$$

discussed in Example 5 of Section 8.1. Is T one-to-one? Is T onto?

Example 5. Let $V = R^\infty$ be the sequence space discussed in Example 3 of Section 4.1, and consider the linear "shifting operators" on V defined by

$$T_1(u_1, u_2, \ldots, u_n, \ldots) = (0, u_1, u_2, \ldots, u_n, \ldots)$$
$$T_2(u_1, u_2, \ldots, u_n, \ldots) = (u_2, u_3, \ldots, u_n, \ldots).$$

(a) Show that T_1 is one-to-one but not onto.

(b) Show that T_2 is onto but not one-to-one.

Example 6. Let

$$D : C^1(-\infty, \infty) \to F(-\infty, \infty)$$

be the differentiation transformation discussed in Example 11 of Section 8.1. Is D one-to-one?

Theorem 8.2.2. *If V and W are finite-dimensional vector spaces with the same dimension, and if $T : V \to W$ is a linear transformation, then the following statements are equivalent.*

(a) *T is one-to-one.*
(b) $\ker(T) = \{\mathbf{0}\}$.
(c) *T is onto [i.e., $R(T) = W$].*

Example 7. If $T_A : R^n \to R^m$ is multiplication by an $m \times n$ matrix A, then when is T_A one-to-one and when is T_A onto?

Theorem 8.2.3. *If T_A is a matrix transformation, then*

(a) *T_A is one-to-one if and only if the columns of A are linearly independent.*
(b) *T_A is onto if and only if the columns of A span R^m.*

Proof. (a) It follows from Theorem 8.2.1 that T_A is one-to-one if and only if A has nullity 0, which is equivalent to saying that A has rank m, which is equivalent to saying that the m column vectors of A are linearly independent. (b) To say that T_A is onto is equivalent to saying that the system $A\mathbf{x} = \mathbf{b}$ has a solution for every vector $\mathbf{b}$ in R^m. But this is so if and only if the columns of A span R^m. $\square$

Theorem 8.2.4 (Equivalent Statements)**.** *If A is an $n \times n$ matrix, then the following statements are equivalent.*

(a) *A is invertible.*
(b) *$A\mathbf{x} = \mathbf{0}$ has only the trivial solution.*
(c) *The reduced row echelon form of A is I_n.*
(d) *A is expressible as a product of elementary matrices.*
(e) *$A\mathbf{x} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.*
(f) *$A\mathbf{x} = \mathbf{b}$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.*
(g) $\det(A) \neq 0$.
(h) *The column vectors of A are linearly independent.*
(i) *The row vectors of A are linearly independent.*
(j) *The column vectors of A span R^n.*
(k) *The row vectors of A span R^n.*
(l) *The column vectors of A form a basis for R^n.*
(m) *The row vectors of A form a basis for R^n.*
(n) *A has rank n.*
(o) *A has nullity 0.*
(p) *The orthogonal complement of the null space of A is R^n.*
(q) *The orthogonal complement of the row space of A is $\{\mathbf{0}\}$.*
(r) *$\lambda = 0$ is not an eigenvalue of A.*
(s) *A^TA is invertible.*
(t) *The kernel of T_A is $\{\mathbf{0}\}$.*
(u) *The range of T_A is R^n.*
(v) *T_A is one-to-one.*

Example 8. Let $T : R^3 \to R^3$ be the linear operator defined by the formula

$$T(x_1, x_2, x_3) = (3x_1 + x_2, -2x_1 - 4x_2 + 3x_3, 5x_1 + 4x_2 - 2x_3).$$

Determine whether T is one-to-one; if so, find $T^{-1}(x_1, x_2, x_3)$.

Remark 1. If $T : V \to W$ is a one-to-one linear transformation with range $R(T)$, and if $\mathbf{w}$ is any vector in $R(T)$, then the fact that T is one-to-one means that there is *exactly one* vector $\mathbf{v}$ in V for which $T(\mathbf{v}) = \mathbf{w}$. This fact allows us to define a new function, called the inverse of T (and denoted by T^{-1}), that is defined on the range of T and that maps $\mathbf{w}$ back into $\mathbf{v}$.

Example 9. Find the inverse of the linear transformation $T : P_n \to P_{n+1}$ given by

$$T(\mathbf{p}) = T(p(x)) = xp(x).$$

Definition 8.2.3. If $T_1 : U \to V$ and $T_2 : V \to W$ are linear transformations, then the composition of T_2 with T_1, denoted by $T_2 \circ T_1$ (which is read "T_2 circle T_1"), is the function defined by the formula

$$(T_2 \circ T_1)(\mathbf{u}) = T_2(T_1(\mathbf{u}))$$

where $\mathbf{u}$ is a vector in U.

Theorem 8.2.5. *If $T_1 : U \to V$ and $T_2 : V \to W$ are linear transformations, then $(T_2 \circ T_1) : U \to W$ is also a linear transformation.*

Proof. If $\mathbf{u}$ and $\mathbf{v}$ are vectors in U and c is a scalar, then it follows from the linearity of T_1 and T_2 that

$$\begin{aligned}(T_2 \circ T_1)(\mathbf{u}+\mathbf{v}) &= T_2(T_1(\mathbf{u}+\mathbf{v})) = T_2(T_1(\mathbf{u}) + T_1(\mathbf{v})) \\ &= T_2(T_1(\mathbf{u})) + T_2(T_1(\mathbf{v})) \\ &= (T_2 \circ T_1)(\mathbf{u}) + (T_2 \circ T_1)(\mathbf{v})\end{aligned}$$

and

$$\begin{aligned}(T_2 \circ T_1)(c\mathbf{u}) &= T_2(T_1(c\mathbf{u})) = T_2(cT_1(\mathbf{u})) \\ &= cT_2(T_1(\mathbf{u})) = c(T_2 \circ T_1)(\mathbf{u}).\end{aligned}$$

Thus, $T_2 \circ T_1$ satisfies the two requirements of a linear transformation. □

Example 10. Let $T_1 : P_1 \to P_2$ and $T_2 : P_2 \to P_2$ be the linear transformations given by the formulas

$$T_1(p(x)) = xp(x) \quad \text{and} \quad T_2(p(x)) = p(2x + 4).$$

Find the composition $(T_2 \circ T_1) : P_1 \to P_2$ if $p(x) = c_0 + c_1 x$.

Example 11. If $T : V \to V$ is any linear operator, and if $I : V \to V$ is the identity operator, then show that for all vectors $\mathbf{v}$ in V that $T \circ I$ and $I \circ T$ are the same as T.

Remark 2. Compositions can be defined for more than two linear transformations. For example, if

$$T_1 : U \to V, \quad T_2 : V \to W, \quad \text{and} \quad T_3 : W \to Y$$

are linear transformations, then the composition $T_3 \circ T_2 \circ T_1$ is defined by

$$(T_3 \circ T_2 \circ T_1)(\mathbf{u}) = T_3(T_2(T_1(\mathbf{u}))).$$

Theorem 8.2.6. *If $T_1 : U \to V$ and $T_2 : V \to W$ are one-to-one linear transformations, then:*
(*a*) $T_2 \circ T_1$ *is one-to-one.*
(*b*) $(T_2 \circ T_1)^{-1} = T_1^{-1} \circ T_2^{-1}$.

Proof. (a) We want to show that $T_2 \circ T_1$ maps distinct vectors in U into distinct vectors in W. But if $\mathbf{u}$ and $\mathbf{v}$ are distinct vectors in U, then $T_1(\mathbf{u})$ and $T_1(\mathbf{v})$ are distinct vectors in V since T_1 is one-to-one. This and the fact that T_2 is one-to-one imply that

$$T_2(T_1(\mathbf{u})) \quad \text{and} \quad T_2(T_1(\mathbf{v}))$$

are also distinct vectors. But these expressions can also be written as

$$(T_2 \circ T_1)(\mathbf{u}) \quad \text{and} \quad (T_2 \circ T_1)(\mathbf{v}),$$

so $T_2 \circ T_1$ maps $\mathbf{u}$ and $\mathbf{v}$ into distinct vectors in W.
(b) We want to show that

$$(T_2 \circ T_1)^{-1}(\mathbf{w}) = (T_1^{-1} \circ T_2^{-1})(\mathbf{w})$$

for every vector $\mathbf{w}$ in the range of $T_2 \circ T_1$. For this purpose, let

$$\mathbf{u} = (T_2 \circ T_1)^{-1}(\mathbf{w}),$$

so our goal is to show that

$$\mathbf{u} = (T_1^{-1} \circ T_2^{-1})(\mathbf{w}).$$

But it follows from $\mathbf{u} = (T_2 \circ T_1)^{-1}(\mathbf{w})$ that

$$(T_2 \circ T_1)(\mathbf{u}) = \mathbf{w},$$

or, equivalently,

$$T_2(T_1(\mathbf{u})) = \mathbf{w}.$$

Now, taking T_2^{-1} of each side of this equation, then taking T_1^{-1} of each side of the result yields

$$\mathbf{u} = T_1^{-1}(T_2^{-1}(\mathbf{w})),$$

or, equivalently,

$$\mathbf{u} = (T_1^{-1} \circ T_2^{-1})(\mathbf{w}). \qquad \square$$

8.3 Isomorphism

Definition 8.3.1. A linear transformation $T : V \to W$ that is both one-to-one and onto is said to be an isomorphism, and W is said to be isomorphic to V.

Theorem 8.3.1. *Every real n-dimensional vector space is isomorphic to R^n.*

Proof. Let V be a real n-dimensional vector space. To prove that V is isomorphic to R^n we must find a linear transformation $T : V \to R^n$ that is one-to-one and onto. For this purpose, let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be any basis for V, let

$$\mathbf{u} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_n\mathbf{v}_n$$

be the representation of a vector $\mathbf{u}$ in V as a linear combination of the basis vectors, and let $T : V \to R^n$ be the coordinate map

$$T(\mathbf{u}) = (\mathbf{u})_S = (k_1, k_2, \ldots, k_n).$$

We will show that T is an isomorphism. To prove the linearity, let $\mathbf{u}$ and $\mathbf{v}$ be vectors in V, let c be a scalar, and let

$$\mathbf{u} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_n\mathbf{v}_n \quad \text{and} \quad \mathbf{v} = d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + \cdots + d_n\mathbf{v}_n$$

be the representations of $\mathbf{u}$ and $\mathbf{v}$ as linear combinations of the basis vectors. Then it follows that

$$\begin{aligned} T(c\mathbf{u}) &= T(ck_1\mathbf{v}_1 + ck_2\mathbf{v}_2 + \cdots + ck_n\mathbf{v}_n) \\ &= (ck_1, ck_2, \ldots, ck_n) \\ &= c(k_1, k_2, \ldots, k_n) = cT(\mathbf{u}) \end{aligned}$$

and that

$$\begin{aligned} T(\mathbf{u}+\mathbf{v}) &= T\big((k_1+d_1)\mathbf{v}_1 + (k_2+d_2)\mathbf{v}_2 + \cdots + (k_n+d_n)\mathbf{v}_n\big) \\ &= (k_1+d_1, k_2+d_2, \ldots, k_n+d_n) \\ &= (k_1, k_2, \ldots, k_n) + (d_1, d_2, \ldots, d_n) \\ &= T(\mathbf{u}) + T(\mathbf{v}), \end{aligned}$$

which shows that T is linear. To show that T is one-to-one, we must show that if $\mathbf{u}$ and $\mathbf{v}$ are distinct vectors in V, then so are their images in R^n. But if $\mathbf{u} \neq \mathbf{v}$, and if the representations of these vectors in terms of the basis vectors are as above, then we must have $k_i \neq d_i$ for at least one i. Thus,

$$T(\mathbf{u}) = (k_1, k_2, \ldots, k_n) \neq (d_1, d_2, \ldots, d_n) = T(\mathbf{v}),$$

which shows that $\mathbf{u}$ and $\mathbf{v}$ have distinct images under T. Finally, the transformation T is onto, for if $\mathbf{w} = (k_1, k_2, \ldots, k_n)$ is any vector in R^n, then it follows that $\mathbf{w}$ is the image under T of the vector $\mathbf{u} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + \cdots + k_n\mathbf{v}_n$. □

Theorem 8.3.2. *If S is an ordered basis for a vector space V, then the coordinate map*

$$\mathbf{u} \xrightarrow{T} (\mathbf{u})_S$$

is an isomorphism between V and R^n.

Example 1. Find an isomorphism between P_{n-1} and R^n.

Example 2. Find an isomorphism between M_{22} and R^4.

Example 3. Use isomorphisms to calculate the derivative

$$\frac{d}{dx}(2 + x + 4x^2 - x^3) = 1 + 8x - 3x^2$$

as a matrix product.

Example 4. Use the natural isomorphism between P_5 and R^6 to determine whether the following polynomials are linearly independent.

$$\begin{aligned}
\mathbf{p}_1 &= 1 + 2x - 3x^2 + 4x^3 + x^5 \\
\mathbf{p}_2 &= 1 + 3x - 4x^2 + 6x^3 + 5x^4 + 4x^5 \\
\mathbf{p}_3 &= 3 + 8x - 11x^2 - 16x^3 + 10x^4 + 9x^5
\end{aligned}$$

Remark 1. If V and W are inner product spaces, then we call an isomorphism $T : V \to W$ an <u>inner product space isomorphism</u> if

$$\langle T(\mathbf{u}), T(\mathbf{v})\rangle = \langle \mathbf{u}, \mathbf{v}\rangle \quad \text{for all } \mathbf{u} \text{ and } \mathbf{v} \text{ in } V.$$

Theorem 8.3.3. *If $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an ordered orthonormal basis for a real vector space V, then the coordinate map*

$$\mathbf{u} \xrightarrow{T} (\mathbf{u})_S$$

is an inner product space isomorphism between V and the vector space R^n with the Euclidean inner product.

Example 5. Show that the isomorphism in Example 1 is an inner product space isomorphism.

Example 6. Find an inner product isomorphism between R^n and M_n, the vector space of real $n \times 1$ matrices.

8.4 Matrices for General Linear Transformations

Remark 1. Suppose that V is an n-dimensional vector space, that W is an m-dimensional vector space, and that $T : V \to W$ is a linear transformation. Suppose further that B is a basis for V, that B' is a basis for W, and that for each vector $\mathbf{x}$ in V, the coordinate matrices for $\mathbf{x}$ and $T(\mathbf{x})$ are $[\mathbf{x}]_B$ and $[T(\mathbf{x})]_{B'}$, respectively. Then the matrix for T relative to the bases B and B' is denoted by the symbol $[T]_{B',B}$ and given by

$$[T]_{B',B} = \big[[T(\mathbf{u}_1)]_{B'} \mid [T(\mathbf{u}_2)]_{B'} \mid \cdots \mid [T(\mathbf{u}_n)]_{B'}\big]$$

and has the property

$$[T]_{B',B}[\mathbf{x}]_B = [T(\mathbf{x})]_{B'}.$$

Example 1. Let $T : P_1 \to P_2$ be the linear transformation defined by

$$T(p(x)) = xp(x).$$

Find the matrix for T with respect to the standard bases

$$B = \{\mathbf{u}_1, \mathbf{u}_2\} \quad \text{and} \quad B' = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$$

where

$$\mathbf{u}_1 = 1, \quad \mathbf{u}_2 = x; \quad \mathbf{v}_1 = 1, \quad \mathbf{v}_2 = x, \quad \mathbf{v}_3 = x^2.$$

Example 2. Let $T : P_1 \to P_2$ be the linear transformation in Example 1, and use the three-step procedure illustrated in the following figure to perform the computation

$$T(a + bx) = x(a + bx) = ax + bx^2.$$

$$\begin{tikzcd}
\mathbf{x} \arrow[rr, "\text{Direct}", "\text{computation}"'] \arrow[d, "(1)"'] & & T(\mathbf{x}) \\
{[\mathbf{x}]_B} \arrow[rr, "\text{Multiply by } {[T]_{B',B}}", "(2)"'] & & {[T(\mathbf{x})]_{B'}} \arrow[u, "(3)"']
\end{tikzcd}$$

Example 3. Let $T : R^2 \to R^3$ be the linear transformation defined by

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_2 \\ -5x_1 + 13x_2 \\ -7x_1 + 16x_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -5 & 13 \\ -7 & 16 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Find the matrix for the transformation T with respect to the bases $B = \{\mathbf{u}_1, \mathbf{u}_2\}$ for R^2 and $B' = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ for R^3, where

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 5 \\ 2 \end{bmatrix}; \quad \mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}.$$

Remark 2. In the special case where $V = W$ (so that $T : V \to V$ is a linear operator), it is usual to take $B = B'$ when constructing a matrix for T. In this case the resulting matrix is called the matrix for T relative to the basis B and is usually denoted by $[T]_B$ rather than $[T]_{B,B}$. If $B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$, then

$$[T]_B = \big[[T(\mathbf{u}_1)]_B \mid [T(\mathbf{u}_2)]_B \mid \cdots \mid [T(\mathbf{u}_n)]_B\big]$$

and has the property

$$[T]_B[\mathbf{x}]_B = [T(\mathbf{x})]_B.$$

Example 4. If $B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is a basis for an n-dimensional vector space V, and if $I : V \to V$ is the identity operator on V, then find $[I]_B$.

Example 5. Let $T : P_2 \to P_2$ be the linear operator defined by

$$T(p(x)) = p(3x - 5),$$

that is, $T(c_0 + c_1x + c_2x^2) = c_0 + c_1(3x - 5) + c_2(3x - 5)^2$.

(a) Find $[T]_B$ relative to the basis $B = \{1, x, x^2\}$.

(b) Use the indirect procedure to compute $T(1 + 2x + 3x^2)$.

(c) Check the result in (b) by computing $T(1 + 2x + 3x^2)$ directly.

Theorem 8.4.1. *If $T_1 : U \to V$ and $T_2 : V \to W$ are linear transformations, and if B, B'', and B' are bases for U, V, and W respectively, then*

$$[T_2 \circ T_1]_{B',B} = [T_2]_{B',B''}[T_1]_{B'',B}.$$

Theorem 8.4.2. *If $T : V \to V$ is a linear operator, and if B is a basis for V, then the following are equivalent.*
(a) *T is one-to-one.*
(b) *$[T]_B$ is invertible.*
Moreover, when these equivalent conditions hold,

$$[T^{-1}]_B = [T]_B^{-1}.$$

Example 6. Let $T_1 : P_1 \to P_2$ be the linear transformation defined by

$$T_1(p(x)) = xp(x)$$

and let $T_2 : P_2 \to P_2$ be the linear operator defined by

$$T_2(p(x)) = p(3x - 5).$$

Find $[T_2 \circ T_1]_{B',B}$ relative to the bases $B = \{1, x\}$ and $B' = \{1, x, x^2\}$.

8.5 Similarity

Theorem 8.5.1. *If B and B' are bases for a finite-dimensional vector space V, and if $I : V \to V$ is the identity operator on V, then*

$$P_{B\to B'} = [I]_{B',B} \quad \textit{and} \quad P_{B'\to B} = [I]_{B,B'}.$$

Proof. Suppose that $B = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ and $B' = \{\mathbf{u}'_1, \mathbf{u}'_2, \dots, \mathbf{u}'_n\}$ are bases for V. Using the fact that $I(\mathbf{v}) = \mathbf{v}$ for all $\mathbf{v}$ in V, it follows that

$$\begin{aligned}[I]_{B',B} &= \big[[I(\mathbf{u}_1)]_{B'} \mid [I(\mathbf{u}_2)]_{B'} \mid \cdots \mid [I(\mathbf{u}_n)]_{B'}\big] \\ &= \big[[\mathbf{u}_1]_{B'} \mid [\mathbf{u}_2]_{B'} \mid \cdots \mid [\mathbf{u}_n]_{B'}\big] \\ &= P_{B\to B'}.\end{aligned}$$

The proof that $[I]_{B,B'} = P_{B'\to B}$ is similar. □

Theorem 8.5.2. *Let $T : V \to V$ be a linear operator on a finite-dimensional vector space V, and let B and B' be bases for V. Then*

$$[T]_{B'} = P^{-1}[T]_B P$$

where $P = P_{B'\to B}$ and $P^{-1} = P_{B\to B'}$.

Theorem 8.5.3. *If V is a finite-dimensional vector space, then two matrices A and B represent the same linear operator (but possibly with respect to different bases) if and only if they are similar. Moreover, if $B = P^{-1}AP$, then $P$is the transition matrix from the bases used for B to the basis used for A.*

Example 1. Show that the matrices

$$C = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

represent the same linear operator $T : R^2 \to R^2$ where C is the matrix relative to the basis $B = \{\mathbf{e}_1, \mathbf{e}_2\}$ and D is the matrix relative to the basis $B' = \{\mathbf{u}'_1, \mathbf{u}'_2\}$ in which

$$\mathbf{u}'_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}'_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

Verify that these matrices are similar by finding a matrix P for which $D = P^{-1}CP$.

Remark 1. We define the determinant of the linear operator T to be

$$\det(T) = \det[T]_B$$

where B is *any* basis for V.

Table 1 Similarity Invariants

Property	Similarity
Determinant	$[T]_B$ and $P^{-1}[T]_BP$ have the same determinant.
Invertibility	$[T]_B$ is invertible if and only if $P^{-1}[T]_BP$ is invertible.
Rank	$[T]_B$ and $P^{-1}[T]_BP$ have the same rank.
Nullity	$[T]_B$ and $P^{-1}[T]_BP$ have the same nullity.
Trace	$[T]_B$ and $P^{-1}[T]_BP$ have the same trace.
Characteristic polynomial	$[T]_B$ and $P^{-1}[T]_BP$ have the same characteristic polynomial.
Eigenvalues	$[T]_B$ and $P^{-1}[T]_BP$ have the same eigenvalues.
Eigenspace dimension	If λ is an eigenvalue of $[T]_B$ and $P^{-1}[T]_BP$, then the eigenspace of $[T]_B$ corresponding to λ and the eigenspace of $P^{-1}[T]_BP$ corresponding to λ have the same dimension.

Example 2. Find $\det[T]$ and $\det[T]_{B'}$ for

$$[T] = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix} \quad \text{and} \quad [T]_{B'} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}.$$

Example 3. Find the eigenvalues of the linear operator $T : P_2 \to P_2$ defined by

$$T(a + bx + cx^2) = -2c + (a + 2b + c)x + (a + 3c)x^2.$$

8.6 Geometry of Matrix Operators

Theorem 8.6.1. *If $T : R^2 \to R^2$ is multiplication by an invertible matrix, then:*

(a) *The image of a straight line is a straight line.*
(b) *The image of a line through the origin is a line through the origin.*
(c) *The images of parallel lines are parallel lines.*
(d) *The image of the line segment joining points P and Q is the line segment joining the images of P and Q.*
(e) *The images of three points lie on a line if and only if the points themselves lie on a line.*

Example 1. According to Theorem 8.6.1, the invertible matrix

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

maps the line $y = 2x + 1$ into another line. Find its equation.

Example 2. Sketch the image of the unit square under multiplication by the invertible matrix

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

Label the vertices of the image with their coordinates, and number the edges of the unit square and their corresponding images.

Table 1

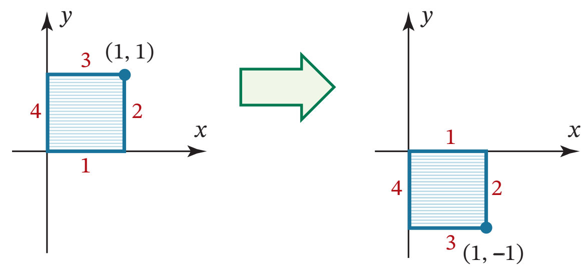
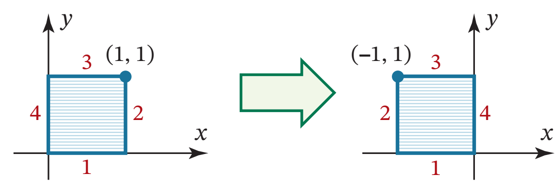
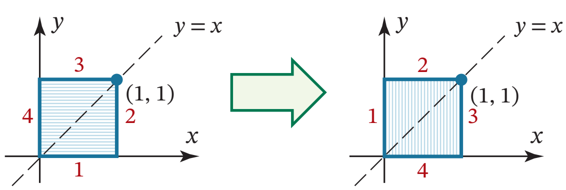
Operator	Standard Matrix	Effect on the Unit Square
Reflection about the x-axis	$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$	
Reflection about the y-axis	$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$	
Reflection about the line $y = x$	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$	

Table 2

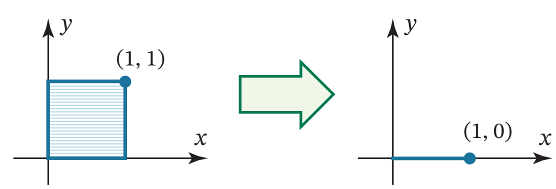
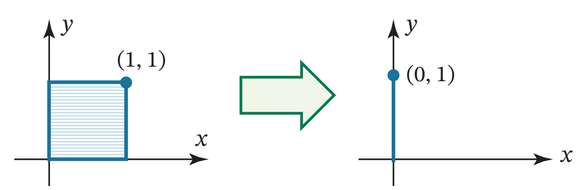
Operator	Standard Matrix	Effect on the Unit Square
Orthogonal projection onto the x-axis	$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$	
Orthogonal projection onto the y-axis	$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$	

Table 3

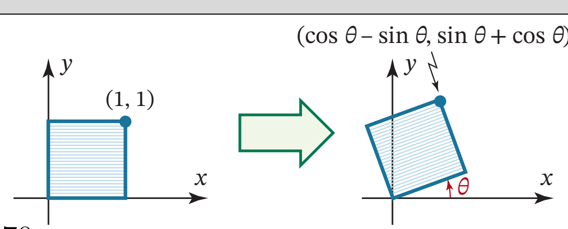
Operator	Standard Matrix	Effect on the Unit Square
Rotation about the origin through a positive angle θ	$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$	

Table 4

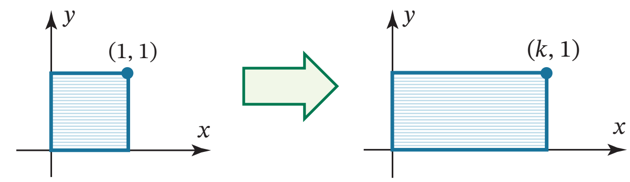
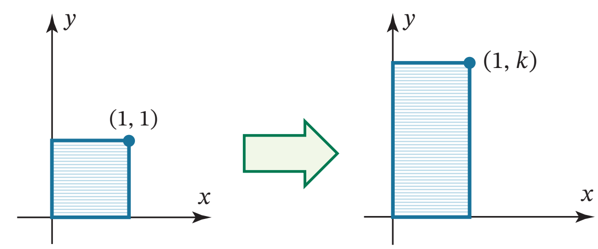
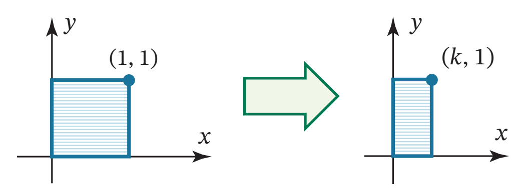
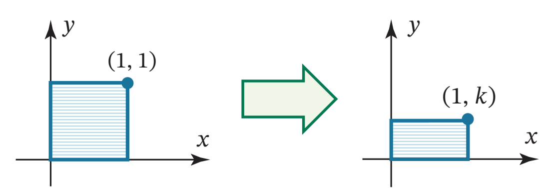
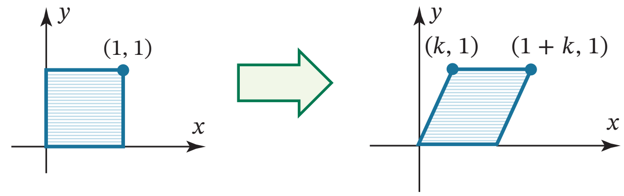
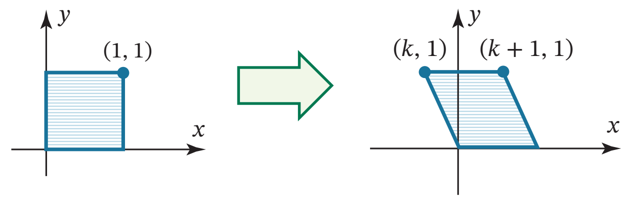
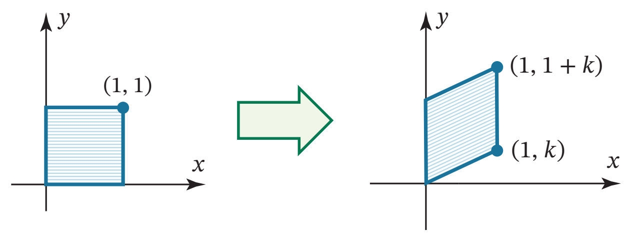
Operator	Standard Matrix	Effect on the Unit Square
Expansion in the x-direction with factor k $(k > 1)$	$\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$	
Expansion in the y-direction with factor k $(k > 1)$	$\begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$	
Compression in the x-direction with factor k $(0 < k < 1)$	$\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix}$	
Compression in the y-direction with factor k $(0 < k < 1)$	$\begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$	

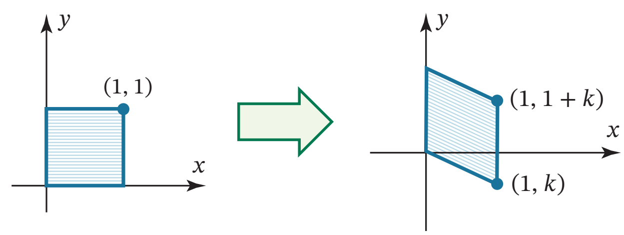
Table 5

<table>
<tr><th>Operator</th><th>Standard Matrix</th><th>Effect on the Unit Square</th></tr>
<tr><td>Shear in the positive x-direction by a factor k

$(k > 0)$</td><td>$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$</td><td>
</td></tr>
<tr><td>Shear in the negative x-direction by a factor k

$(k < 0)$</td><td>$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$</td><td>
</td></tr>
<tr><td>Shear in the positive y-direction by a factor k

$(k > 0)$</td><td>$\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$</td><td>
</td></tr>
<tr><td>Shear in the negative y-direction by a factor k

$(k < 0)$</td><td>$\begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix}$</td><td>
</td></tr>
</table>

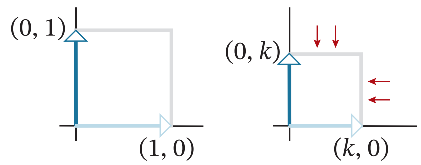
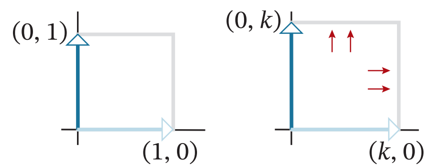
Example 3.

(a) Find the standard matrix for the operator on R^2 that first shears by a factor of 2 in the x-direction and then reflects the result about the line $y = x$. Sketch the image of the unit square under this operator.

(b) Find the standard matrix for the operator on R^2 that first reflects about $y = x$ and then shears by a factor of 2 in the x-direction. Sketch the image of the unit square under this operator.

(c) Confirm that the shear and the reflection in parts (a) and (b) do not commute.

Table 6

Operator	Effect on the Unit Square	Standard Matrix
Contraction with factor k in R^2 $(0 \le k < 1)$		$\begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix}$
Dilation with factor k in R^2 $(k > 1)$		

Example 4. Discuss the geometric effect on the unit square of multiplication by a diagonal matrix

$$A = \begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix}$$

in which the entries k_1 and k_2 are positive real numbers ($\neq 1$).

Example 5. Discuss the geometric effect on the unit square of multiplication by the matrix

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

Example 6. Discuss the geometric effect on the unit square of multiplication by the matrix

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

Theorem 8.6.2. *If E is an elementary matrix, then $T_E : R^2 \to R^2$ is one of the following:*

(a) *A shear along a coordinate axis.*
(b) *A reflection about $y = x$.*
(c) *A compression along a coordinate axis.*
(d) *An expansion along a coordinate axis.*
(e) *A reflection about a coordinate axis.*
(f) *A compression or expansion along a coordinate axis followed by a reflection about a coordinate axis.*

Proof. Because a 2×2 elementary matrix results from performing a single elementary row operation on the 2×2 identity matrix, such a matrix must have one of the following forms:

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

The first two matrices represent shears along coordinate axes, and the third represents a reflection about $y = x$. If $k > 0$, the last two matrices represent compressions or expansions along coordinate axes, depending on whether $0 \leq k < 1$ or $k > 1$. If $k < 0$, and if we express k in the form $k = -k_1$ where $k_1 > 0$, then the last two matrices can be written as

$$\begin{bmatrix} k & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -k_1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} k_1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -k_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & k_1 \end{bmatrix}$$

Since $k_1 > 0$, the first product represents a compression or expansion along the x-axis followed by a reflection about the y-axis, and the second product represents a compression or expansion along the y-axis followed by a reflection about the x-axis. In the case where $k = -1$, these transformations are simply reflections about the y-axis and x-axis, respectively. $\square$

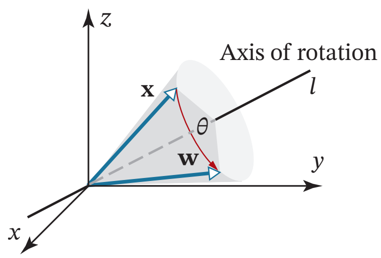
Theorem 8.6.3. *If $T_A : R^2 \to R^2$ is multiplication by an invertible matrix A, then the geometric effect of T_A is the same as an appropriate succession of shears, compressions, expansions, and reflections.*

Example 7. In Example 2 we illustrated the effect on the unit square of multiplication by

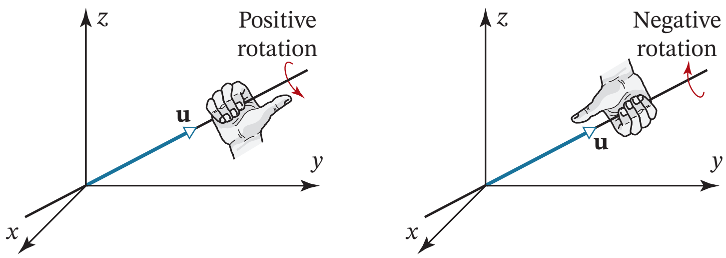
$$A = \begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix}.$$

Express this matrix as a product of elementary matrices, and then describe the effect of multiplication by A in terms of shears, compressions, expansions, and reflections.

Remark 1. The right-hand rule can be used to establish a sign for an angle of rotation about a unit vector **u** by cupping the fingers of your right hand so they curl in the direction of rotation. If your thumb points in the direction of **u**, then the angle of rotation is regarded to be positive relative to **u**, and if it points in the direction opposite to **u**, then it is regarded to be be negative to **u**.

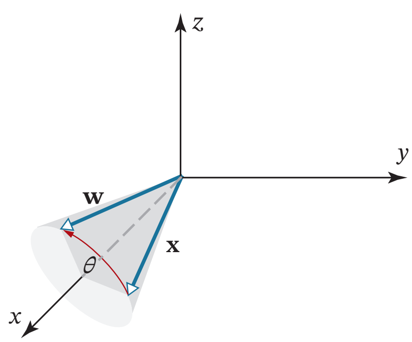
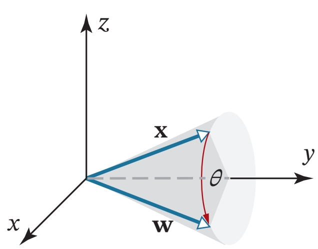
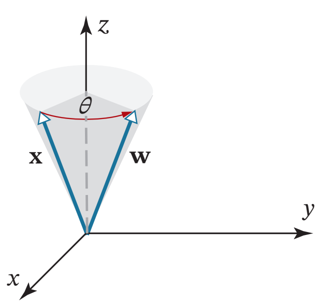


(*a*) Angle of rotation



(*b*) Right-hand rule

Table 6

Operator	Illustration	Rotation Equations	Standard Matrix
Counterclockwise rotation about the positive x-axis through an angle θ		$w_1 = x$ $w_2 = y\cos\theta - z\sin\theta$ $w_3 = y\sin\theta + z\cos\theta$	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}$
Counterclockwise rotation about the positive y-axis through an angle θ		$w_1 = x\cos\theta + z\sin\theta$ $w_2 = y$ $w_3 = -x\sin\theta + z\cos\theta$	$\begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}$
Counterclockwise rotation about the positive z-axis through an angle θ		$w_1 = x\cos\theta - y\sin\theta$ $w_2 = x\sin\theta + y\cos\theta$ $w_3 = z$	$\begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Chapter 9

Numerical Methods

9.1 LU-Decompositions

Definition 9.1.1. A factorization of square matrix A as

$$A = LU$$

where L is lower triangular and U is upper triangular, is called an <u>LU-decomposition</u> (or <u>LU-factorization</u>) of A.

Example 1. Use the factorization

$$\begin{bmatrix} 2 & 6 & 2 \\ -3 & -8 & 0 \\ 4 & 9 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ -3 & 1 & 0 \\ 4 & -3 & 7 \end{bmatrix} \begin{bmatrix} 1 & 3 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}$$

to solve the linear system

$$\begin{bmatrix} 2 & 6 & 2 \\ -3 & -8 & 0 \\ 4 & 9 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}.$$

Theorem 9.1.1. *If A is a square matrix that can be reduced to a row echelon form U by Gaussian elimination without row interchanges, then A can be factored as $A = LU$, where L is a lower triangular matrix.*

Example 2. Find an LU-decomposition of

$$A = \begin{bmatrix} 2 & 6 & 2 \\ -3 & -8 & 0 \\ 4 & 9 & 2 \end{bmatrix}.$$

Example 3. Find an LU-decomposition of

$$A = \begin{bmatrix} 6 & -2 & 0 \\ 9 & -1 & 1 \\ 3 & 7 & 5 \end{bmatrix}.$$

9.2 The Power Method

Remark 1. There are many applications in which some vector $\mathbf{x}_0$ in R^n is multiplied repeatedly by an $n \times n$ matrix A to produce a sequence

$$\mathbf{x}_0, \quad A\mathbf{x}_0, \quad A^2\mathbf{x}_0, \ldots, \quad A^k\mathbf{x}_0, \ldots.$$

We call a sequence of this form a power sequence generated by A.

Definition 9.2.1. If the *distinct* eigenvalues of a matrix A are $\lambda_1, \lambda_2, \ldots, \lambda_k$, and if $|\lambda_1|$ is larger than $|\lambda_2|, \ldots, |\lambda_k|$, then λ_1 is called a dominant eigenvalue of A. Any eigenvector corresponding to a dominant eigenvalue is called a dominant eigenvector of A.

Example 1. Find the dominant eigenvalues, if any, of a matrix with distinct eigenvalues

$$\lambda_1 = -4, \quad \lambda_2 = -2, \quad \lambda_3 = 1, \quad \lambda_4 = 3$$

and of a matrix with distinct eigenvalues

$$\lambda_1 = 7, \quad \lambda_2 = -7, \quad \lambda_3 = -2, \quad \lambda_4 = 5.$$

Theorem 9.2.1. *Let A be a symmetric $n \times n$ matrix that has a positive dominant eigenvalue λ. If $\mathbf{x}_0$ is a unit vector in R^n that is not orthogonal to the eigenspace corresponding to λ, then the normalized power sequence*

$$\mathbf{x}_0, \quad \mathbf{x}_1 = \frac{A\mathbf{x}_0}{\|A\mathbf{x}_0\|}, \quad \mathbf{x}_2 = \frac{A\mathbf{x}_1}{\|A\mathbf{x}_1\|}, \ldots, \quad \mathbf{x}_k = \frac{A\mathbf{x}_{k-1}}{\|A\mathbf{x}_{k-1}\|}, \ldots$$

converges to a unit dominant eigenvector, and the sequence

$$A\mathbf{x}_1 \cdot \mathbf{x}_1, \quad A\mathbf{x}_2 \cdot \mathbf{x}_2, \quad A\mathbf{x}_3 \cdot \mathbf{x}_3, \ldots, \quad A\mathbf{x}_k \cdot \mathbf{x}_k, \ldots$$

converges to the dominant eigenvalue λ.

Remark 2. Theorem 9.2.1 provides us with an algorithm for approximating the dominant eigenvalue and a corresponding unit eigenvector of a symmetric matrix A, provided that the dominant eigenvalue is positive. This algorithm is called the power method with Euclidean scaling.

Example 2. Apply the power method with Euclidean scaling to

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} \quad \text{with} \quad \mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Stop at $\mathbf{x}_5$ and compare the resulting approximations to the exact values of the dominant eigenvalue and eigenvector.

Theorem 9.2.2. *Let A be a symmetric $n \times n$ matrix that has a positive dominant eigenvalue λ. If $\mathbf{x}_0$ is a nonzero vector in R^n that is not orthogonal to the eigenspace corresponding to λ, then the sequence*

$$\mathbf{x}_0, \quad \mathbf{x}_1 = \frac{A\mathbf{x}_0}{\max(A\mathbf{x}_0)}, \quad \mathbf{x}_2 = \frac{A\mathbf{x}_1}{\max(A\mathbf{x}_1)}, \ldots, \quad \mathbf{x}_k = \frac{A\mathbf{x}_{k-1}}{\max(A\mathbf{x}_{k-1})}, \ldots$$

converges to an eigenvector corresponding to λ, and the sequence

$$\frac{A\mathbf{x}_0 \cdot \mathbf{x}_0}{\mathbf{x}_0 \cdot \mathbf{x}_0}, \quad \frac{A\mathbf{x}_1 \cdot \mathbf{x}_1}{\mathbf{x}_1 \cdot \mathbf{x}_1}, \ldots, \quad \frac{A\mathbf{x}_k \cdot \mathbf{x}_k}{\mathbf{x}_k \cdot \mathbf{x}_k}, \ldots$$

converges to λ.

Remark 3. The algorithm provided by Theorem 9.2.2 is called the <u>power method with maximum entry scaling</u>, where $\max(\mathbf{x})$ denotes the maximum *absolute value* of the entries in a vector $\mathbf{x}$.

Example 3. Apply the power method with maximum entry scaling to

$$A = \begin{bmatrix} 3 & 2 \\ 2 & 2 \end{bmatrix} \quad \text{with} \quad \mathbf{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Stop at $\mathbf{x}_5$ and compare the resulting approximations to the exact values and to the approximations obtained in Example 2.

Remark 4. If λ is the exact value of the dominant eigenvalue, and if a power method produces the approximation $\lambda^{(k)}$ at the kth iteration, then we call

$$\left|\frac{\lambda - \lambda^{(k)}}{\lambda}\right|$$

the relative error in $\lambda^{(k)}$. Expressed as a percentage it is called the percentage error in $\lambda^{(k)}$. It is usual to estimate λ by $\lambda^{(k)}$ and stop computations when

$$\left|\frac{\lambda^{(k)} - \lambda^{(k-1)}}{\lambda^{(k)}}\right| < E$$

for a known relative error E. The quantity on the left side is called the estimated relative error in $\lambda^{(k)}$ and its percentage form is called the estimated percentage error in $\lambda^{(k)}$.

Example 4. For the computations in Example 3, find the smallest value of k for which the estimated percentage error in $\lambda^{(k)}$ is less than 0.1%.

9.3 Comparison of Procedures for Solving Linear Systems

Remark 1. In computer jargon, an arithmetic operation $(+, -, *, \div)$ on two real numbers is called a flop, which is an acronym for "floating-point operation." The total number of flops required to solve a problem, which is called the cost of the solution, provides a convenient way of choosing between various algorithms for solving the problem.

Table 1 Approximate Cost for an $n \times n$ matrix A with Large n

Algorithm	Cost in Flops
Gauss-Jordan elimination (forward phase)	$\approx \frac{2}{3}n^3$
Gauss-Jordan elimination (backward phase)	$\approx n^2$
LU-decomposition of A	$\approx \frac{2}{3}n^3$
Forward substitution to solve $L\mathbf{y} = \mathbf{b}$	$\approx n^2$
Backward substitution to solve $U\mathbf{x} = \mathbf{y}$	$\approx n^2$
A^{-1} by reducing $[A \mid I]$ to $[I \mid A^{-1}]$	$\approx 2n^3$
Compute $A^{-1}\mathbf{b}$	$\approx 2n^3$

Example 1. Approximate the time required to execute the forward and backward phases of Gauss-Jordan elimination for a system of one million $(= 10^6)$ equations in one million unknowns using a computer that can execute 10 petaflops per second (1 petaflop $= 10^{15}$ flops).

9.4 Singular Value Decomposition

Theorem 9.4.1. *If A is an $m \times n$ matrix, then:*
(a) *A and A^TA have the same null space.*
(b) *A and A^TA have the same row space.*
(c) *A^T and A^TA have the same column space.*
(d) *A and A^TA have the same rank.*

Theorem 9.4.2. *If A is an $m \times n$ matrix, then:*
(a) *A^TA is orthogonally diagonalizable.*
(b) *The eigenvalues of A^TA are nonnegative real numbers.*

Proof. (a) The matrix A^TA, being symmetric, is orthogonally diagonalizable. (b) Since A^TA is orthogonally diagonalizable, there is an orthonormal basis for R^n consisting of eigenvectors of A^TA, say $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$. If we let $\lambda_1, \lambda_2, \dots, \lambda_n$ be the corresponding eigenvalues, then for $1 \le i \le n$ we have

$$\begin{aligned}\|A\mathbf{v}_i\|^2 &= A\mathbf{v}_i \cdot A\mathbf{v}_i = \mathbf{v}_i \cdot A^TA\mathbf{v}_i \\ &= \mathbf{v}_i \cdot \lambda_i\mathbf{v}_i = \lambda_i(\mathbf{v}_i \cdot \mathbf{v}_i) = \lambda_i\|\mathbf{v}_i\|^2 = \lambda_i.\end{aligned}$$

It follows from this relationship that $\lambda_i \ge 0$. □

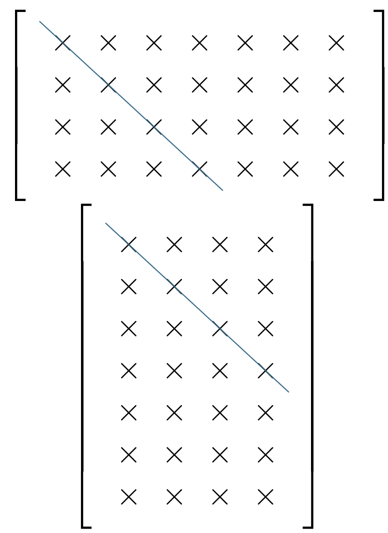
Definition 9.4.1. If A is an $m \times n$ matrix, and if $\lambda_1, \lambda_2, \dots, \lambda_n$ are the eigenvalues of A^TA, then the numbers

$$\sigma_1 = \sqrt{\lambda_1}, \quad \sigma_2 = \sqrt{\lambda_2}, \dots, \quad \sigma_n = \sqrt{\lambda_n}.$$

are called the <u>singular values</u> of A.

Example 1. Find the singular values of matrix

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



Remark 1. We define the <u>main diagonal</u> of an $m \times n$ matrix to be the line of entries shown in the figure—it starts at the upper left corner and extends diagonally as far as it can go. We will refer to the entries on the main diagonal as <u>diagonal entries</u>.

Theorem 9.4.3 (Singular Value Decomposition (Brief Form))**.** *If A is an $m \times n$ matrix of rank k, then A can be expressed in the form $A = U\Sigma V^T$, where Σ has size $m \times n$ and can be expressed in partitioned form as*

$$\Sigma = \left[\begin{array}{c|c} D & 0_{k\times(n-k)} \\ \hline 0_{(m-k)\times k} & 0_{(m-k)\times(n-k)} \end{array}\right]$$

in which D is a diagonal $k \times k$ matrix whose successive entries are the first k singular values of A in nonincreasing order, U is an $m \times n$ orthogonal matrix, and V is an $n \times n$ orthogonal matrix.

Theorem 9.4.4 (Singular Value Decomposition (Expanded Form))**.** *If A is an $m \times n$ matrix of rank k, then A can be factored as*

$$A = U\Sigma V^T = \left[\begin{array}{cccc|ccc} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k & \mathbf{u}_{k+1} & \cdots & \mathbf{u}_m \end{array}\right] \left[\begin{array}{cccc|c} \sigma_1 & 0 & \cdots & 0 & \\ 0 & \sigma_2 & \cdots & 0 & \\ \vdots & \vdots & \ddots & \vdots & 0_{k\times(n-k)} \\ 0 & 0 & \cdots & \sigma_k & \\ \hline & & 0_{(m-k)\times k} & & 0_{(m-k)\times(n-k)} \end{array}\right] \left[\begin{array}{c} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_k^T \\ \hline \mathbf{v}_{k+1}^T \\ \vdots \\ \mathbf{v}_n^T \end{array}\right]$$

in which U, Σ, and V have sizes $m \times m$, $m \times n$, and $n \times n$, respectively, and in which:

(*a*) $V = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix}$ *orthogonally diagonalizes A^TA.*

(*b*) *The nonzero diagonal entries of Σ are* $\sigma_1 = \sqrt{\lambda_1}, \sigma_2 = \sqrt{\lambda_2}, \ldots, \sigma_k = \sqrt{\lambda_k}$, *where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the nonzero eigenvalues of A^TA corresponding to the column vectors of V.*

(*c*) *The column vectors of V are ordered so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > 0$.*

(*d*) $\mathbf{u}_i = \dfrac{A\mathbf{v}_i}{\|A\mathbf{v}_i\|} = \dfrac{1}{\sigma_i}A\mathbf{v}_i \qquad (i = 1, 2, \ldots, k).$

(*e*) $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ *is an orthonormal basis for* $\text{col}(A)$.

(*f*) $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k, \mathbf{u}_{k+1}, \ldots, \mathbf{u}_m\}$ *is an extension of* $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ *to an orthonormal basis for R^m.*

Example 2. Find a singular value decomposition of the matrix

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

9.5 Data Compression Using Singular Value Decomposition

Remark 1. The zero rows and columns of the matrix Σ in Theorem 9.4.4 can be eliminated by multiplying out the expression $U\Sigma V^T$ using block multiplication and the partitioning shown in that formula. The products that involve zero blocks as factors drop out, leaving

$$\begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_k \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_k^T \end{bmatrix},$$

which is called a reduced singular value decomposition of A. We will denote the matrices on the right side by U_1, Σ_1, and V_1^T, respectively, and we will write this equation as

$$A = U_1 \Sigma_1 V_1^T.$$

Note that the sizes of U_1, Σ_1, and V_1^T, are $m \times k$, $k \times k$, and $k \times n$, respectively, and that the matrix Σ_1 is invertible since its diagonal entries are positive.
If we multiply out the right side of the equation using the column-row rule, then we obtain

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T,$$

which is called a reduced singular value expansion of A.

Example 1. Find a reduced singular value decomposition and a reduced singular value expansion of the matrix

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

Remark 2. If a matrix A has size $m \times n$, then one might store each of its mn entries individually. An alternative procedure is to compute the reduced singular value decomposition

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T$$

in which $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k$, and store the σ's, the $\mathbf{u}$'s, and $\mathbf{v}$'s. When needed, the matrix A can be reconstructed from this decomposition. Since each $\mathbf{u}_j$ has m entries and each $\mathbf{v}_j$ has n entries, this method requires storage space for

$$km + kn + k = k(m + n + 1)$$

numbers. Suppose, however, that the singular values $\sigma_{r+1}, \ldots, \sigma_k$ are sufficiently small that dropping the corresponding terms in the decomposition produces an acceptable approximation

$$A_r = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \sigma_2 \mathbf{u}_2 \mathbf{v}_2^T + \cdots + \sigma_r \mathbf{u}_r \mathbf{v}_r^T$$

to A and the image that it represents. We call this the rank r approximation of A. This matrix requires storage space for only

$$rm + rn + r = r(m + n + 1)$$

numbers, compared to mn numbers required for entry-by-entry storage of A.

Example 2. Suppose A is a 1000×1000 matrix. How many numbers must be stored in the rank 100 approximation of A? Compare this with the number of entries of A.

Chapter 10

Applications of Linear Algebra

10.1 Constructing Curves and Surfaces Through Specified Points

Theorem 10.1.1. *A homogeneous linear system with as many equations as unknowns has a nontrivial solution if and only if the determinant of the coefficient matrix is zero.*

Remark 1. The line with equation

$$c_1x + c_2y + c_3 = 0$$

that passes through two distinct points (x_1, y_1) and (x_2, y_2) is given by the determinant equation

$$\begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = 0.$$

Example 1. Find the equation of the line that passes through the two points $(2, 1)$ and $(3, 7)$.

Remark 2. The circle with equation

$$c_1(x^2 + y^2) + c_2 x + c_3 y + c_4 = 0$$

that passes through three noncollinear points (x_1, y_1), (x_2, y_2), and (x_3, y_3) is given by the determinant equation

$$\begin{vmatrix} x^2 + y^2 & x & y & 1 \\ x_1^2 + y_1^2 & x_1 & y_1 & 1 \\ x_2^2 + y_2^2 & x_2 & y_2 & 1 \\ x_3^2 + y_3^2 & x_3 & y_3 & 1 \end{vmatrix} = 0.$$

Example 2. Find the equation of the circle that passes through the three points $(1, 7)$, $(6, 2)$, and $(4, 6)$.

Remark 3. The conic section with equation

$$c_1 x^2 + c_2 xy + c_3 y^2 + c_4 x + c_5 y + c_6 = 0$$

that passes through five distinct points (x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), and (x_5, y_5) is given by the determinant equation

$$\begin{vmatrix} x^2 & xy & y^2 & x & y & 1 \\ x_1^2 & x_1 y_1 & y_1^2 & x_1 & y_1 & 1 \\ x_2^2 & x_2 y_2 & y_2^2 & x_2 & y_2 & 1 \\ x_3^2 & x_3 y_3 & y_3^2 & x_3 & y_3 & 1 \\ x_4^2 & x_4 y_4 & y_4^2 & x_4 & y_4 & 1 \\ x_5^2 & x_5 y_5 & y_5^2 & x_5 & y_5 & 1 \end{vmatrix} = 0.$$

Example 3. An astronomer who wants to determine the orbit of an asteroid about the Sun sets up a Cartesian coordinate system in the plane of the orbit with the Sun at the origin. Astronomical units of measurement are used along the axes (1 astronomical unit = mean distance of Earth to Sun = 93 million miles). By Kepler's first law, the orbit must be an ellipse, so the astronomer makes five observations of the asteroid at five different times and finds five points along the orbit to be

$(8.025, 8.310)$, $(10.170, 6.355)$, $(11.202, 3.212)$, $(10.736, 0.375)$, $(9.092, -2.267)$.

Find the equation of the orbit.

Remark 4. The plane in 3-space with equation

$$c_1x + c_2y + c_3z + c_4 = 0$$

that passes through three noncollinear points (x_1, y_1, z_1), (x_2, y_2, z_2), and (x_3, y_3, z_3) is given by the determinant equation

$$\begin{vmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix} = 0.$$

Example 4. Find the equation of the plane that passes through the three points $(1, 1, 0)$, $(2, 0, -1)$, and $(2, 9, 2)$.

Remark 5. The sphere in 3-space with equation

$$c_1(x^2 + y^2 + z^2) + c_2x + c_3y + c_4z + c_5 = 0$$

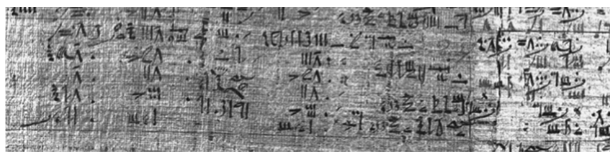
that passes through four noncoplanar points (x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3), and (x_4, y_4, z_4) is given by the determinant equation

$$\begin{vmatrix} x^2 + y^2 + z^2 & x & y & z & 1 \\ x_1^2 + y_1^2 + z_1^2 & x_1 & y_1 & z_1 & 1 \\ x_2^2 + y_2^2 + z_2^2 & x_2 & y_2 & z_2 & 1 \\ x_3^2 + y_3^2 + z_3^2 & x_3 & y_3 & z_3 & 1 \\ x_4^2 + y_4^2 + z_4^2 & x_4 & y_4 & z_4 & 1 \end{vmatrix} = 0.$$

Example 5. Find the equation of the sphere that passes through the four points $(0, 3, 2)$, $(1, -1, 1)$, $(2, 1, 0)$, and $(5, 1, 3)$.

10.2 The Earliest Applications of Linear Algebra

Example 1.

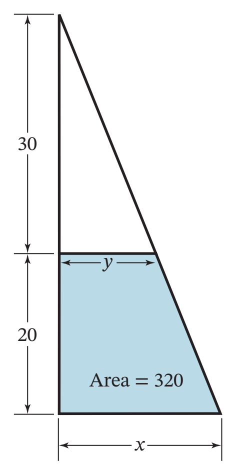


Problem 40 of the Ahmes Papyrus

The Ahmes (or Rhind) Papyrus is the source of most of our information about ancient Egyptian mathematics. This 5-meter-long papyrus contains 84 short mathematical problems, together with their solutions, and dates from about 1650 B.C. Problem 40 in this papyrus is the following:

> *Divide* 100 *hekats of barley among five men in arithmetic progression so that the sum of the two smallest is one-seventh the sum of the three largest.*

Example 2. The Old Babylonian Empire flourished in Mesopotamia between 1900 and 1600 B.C. Many clay tablets containing mathematical tables and problems survive from that period, one of which (designated Ca MLA 1950) contains the next problem. The statement of the problem is a bit muddled because of the condition of the tablet, but the diagram and solution on the tablet indicate that the problem is as follows:



A trapezoid with an area of 320 *square units is cut off from a right triangle by a line parallel to one of its sides. The other side has length* 50 *units, and the height of the trapezoid is* 20 *units. What are the upper and the lower widths of the trapezoid?*

Example 3. The most important treatise in the history of Chinese mathematics is the Chiu Chang Suan Shu, or "The Nine Chapters of the Mathematical Art." This treatise, which is a collection of 246 problems and their solutions, was assembled in its final form by Liu Hui in A.D. 263. Its contents, however, go back to at least the beginning of the Han dynasty in the second century B.C. The eighth of its nine chapters, entitled "The Way of Calculating by Arrays," contains 18 word problems that lead to linear systems in three to six unknowns. The general solution procedure described is almost identical to the Gaussian elimination technique developed in Europe in the nineteenth century by Carl Friedrich Gauss. The first problem in the eighth chapter is the following:

There are three classes of corn, of which three bundles of the first class, two of the second, and one of the third make 39 *measures. Two of the first, three of the second, and one of the third make* 34 *measures. And one of the first, two of the second, and three of the third make* 26 *measures. How many measures of grain are contained in one bundle of each class?*

Example 4. Perhaps the most famous system of linear equations from antiquity is the one associated with the first part of Archimedes' celebrated Cattle Problem. This problem supposedly was posed by Archimedes as a challenge to his colleague Eratosthenes. No solution has come down to us from ancient times, so that it is not known how, or even whether, either of these two geometers solved it.

> *If thou art diligent and wise, O stranger, compute the number of cattle of the Sun, who once upon a time grazed on the fields of the Thrinacian isle of Sicily, divided into four herds of different colors, one milk white, another glossy black, a third yellow, and the last dappled. In each herd were bulls, mighty in number according to these proportions: Understand, stranger, that the white bulls were equal to a half and a third of the black together with the whole of the yellow, while the black were equal to the fourth part of the dappled and a fifth, together with, once more, the whole of the yellow. Observe further that the remaining bulls, the dappled, were equal to a sixth part of the white and a seventh, together with all of the yellow. These were the proportions of the cows: The white were precisely equal to the third part and a fourth of the whole herd of the black; while the black were equal to the fourth part once more of the dappled and with it a fifth part, when all, including the bulls, went to pasture together. Now the dappled in four parts were equal in number to a fifth part and a sixth of the yellow herd. Finally the yellow were in number equal to a sixth part and a seventh of the white herd. If thou canst accurately tell, O stranger, the number of cattle of the Sun, giving separately the number of well-fed bulls and again the number of females according to each color, thou wouldst not be called unskilled or ignorant of numbers, but not yet shalt thou be numbered among the wise.*

Example 5. The Bakhshali Manuscript is an ancient work of Indian/Hindu mathematics dating from around the fourth century A.D., although some of its materials undoubtedly come from many centuries before. It consists of about 70 leaves or sheets of birch bark containing mathematical problems and their solutions. Many of its problems are so-called equalization problems that lead to systems of linear equations. One such problem on the fragment shown is the following:

> *One merchant has seven asava horses, a second has nine haya horses, and a third has ten camels. They are equally well off in the value of their animals if each gives two animals, one to each of the others. Find the price of each animal and the total value of the animals possessed by each merchant.*

10.3 Cubic Spline Interpolation

Remark 1. A curve that passes through a set of points in the plane is said to interpolate those points, and the curve is called an interpolating curve for those points.

Theorem 10.3.1 (Cubic Spline Interpolation)**.** *Given n points $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ with $x_{i+1} - x_i = h$, $i = 1, 2, \dots, n-1$, the cubic spline*

$$S(x) = \begin{cases} a_1(x-x_1)^3 + b_1(x-x_1)^2 + c_1(x-x_1) + d_1, & x_1 \le x \le x_2 \\ a_2(x-x_2)^3 + b_2(x-x_2)^2 + c_2(x-x_2) + d_2, & x_2 \le x \le x_3 \\ \qquad\qquad\vdots & \\ a_{n-1}(x-x_{n-1})^3 + b_{n-1}(x-x_{n-1})^2 & \\ \qquad + c_{n-1}(x-x_{n-1}) + d_{n-1}, & x_{n-1} \le x \le x_n \end{cases}$$

that interpolates these points is given by

$$\begin{aligned} a_i &= (M_{i+1} - M_i)/6h \\ b_i &= M_i/2 \\ c_i &= (y_{i+1} - y_i)/h - [(M_{i+1} + 2M_i)h/6] \\ d_i &= y_i \end{aligned}$$

for $i = 1, 2, \dots, n-1$, where $M_i = S''(x_i)$, $i = 1, 2, \dots, n$.

Table 1

Natural Spline	The second derivative of the spline is zero at the endpoints.	$M_1 = 0$ $M_n = 0$	$\begin{bmatrix} 4 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 4 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 4 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 4 \end{bmatrix} \begin{bmatrix} M_2 \\ M_3 \\ \vdots \\ M_{n-2} \\ M_{n-1} \end{bmatrix} = \frac{6}{h^2} \begin{bmatrix} y_1 - 2y_2 + y_3 \\ y_2 - 2y_3 + y_4 \\ \vdots \\ y_{n-2} - 2y_{n-1} + y_n \end{bmatrix}$
Parabolic Runout Spline	The spline reduces to a parabolic curve on the first and last intervals.	$M_1 = M_2$ $M_n = M_{n-1}$	$\begin{bmatrix} 5 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 4 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 4 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 5 \end{bmatrix} \begin{bmatrix} M_2 \\ M_3 \\ \vdots \\ M_{n-2} \\ M_{n-1} \end{bmatrix} = \frac{6}{h^2} \begin{bmatrix} y_1 - 2y_2 + y_3 \\ y_2 - 2y_3 + y_4 \\ \vdots \\ y_{n-2} - 2y_{n-1} + y_n \end{bmatrix}$
Cubic Runout Spline	The spline is a single cubic curve on the first two and last two intervals.	$M_1 = 2M_2 - M_3$ $M_n = 2M_{n-1} - M_{n-2}$	$\begin{bmatrix} 6 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 4 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 4 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 6 \end{bmatrix} \begin{bmatrix} M_2 \\ M_3 \\ \vdots \\ M_{n-2} \\ M_{n-1} \end{bmatrix} = \frac{6}{h^2} \begin{bmatrix} y_1 - 2y_2 + y_3 \\ y_2 - 2y_3 + y_4 \\ \vdots \\ y_{n-2} - 2y_{n-1} + y_n \end{bmatrix}$

Example 1. The density of water is well known to reach a maximum at a temperature slightly above freezing. Table 2, from the *Handbook of Chemistry and Physics* (CRC Press, 2009), gives the density of water in grams per cubic centimeter for five equally spaced temperatures from -10°C to 30°C. Interpolate these five temperature-density measurements with a parabolic runout spline and find the maximum density of water in this range by finding the maximum value on this cubic spline.

Table 2

Temperature (°C)	Density (g/cm^3)
-10	.99815
0	.99987
10	.99973
20	.99823
30	.99567

10.4 Markov Chains

Remark 1. Suppose a physical or mathematical system undergoes a process of change such that at any moment it can occupy one of a finite number of states. Suppose that such a system changes with time from one state to another and at scheduled times the state of the system is observed. If the state of the system at any observation cannot be predicted with certainty, but the probability that a given state occurs can be predicted by just knowing the state of the system at the preceding observation, then the process of change is called a Markov chain or Markov process.

Definition 10.4.1. If a Markov chain has k possible states, which we label as $1, 2, \ldots, k$, then the probability that the system is in state i at any observation after it was in state j at the preceding observation is denoted by p_{ij} and is called the transition probability from state j to state i. The matrix $P = [p_{ij}]$ is called the transition matrix of the Markov chain.

Example 1. A car rental agency has three rental locations, denoted by 1, 2, and 3. A customer may rent a car from any of the three locations and return the car to any of the three locations. The manager finds that customers return the cars to the various locations according to the following probabilities:

$$\begin{array}{cc} \text{Rented from Location} & \\ \begin{array}{ccc} 1 & 2 & 3 \end{array} & \\ \begin{bmatrix} .8 & .3 & .2 \\ .1 & .2 & .6 \\ .1 & .5 & .2 \end{bmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \end{array} \end{array} \quad \begin{array}{c} \text{Returned} \\ \text{to} \\ \text{Location} \end{array}$$

Find the probability that a car rented from location 3 will be returned to location 2, and the probability that a car rented from location 1 will be returned to location 1.

Example 2. By reviewing its donation records, the alumni office of a college finds that 80% of its alumni who contribute to the annual fund one year will also contribute the next year, and 30% of those who do not contribute one year will contribute the next. What is the transition matrix?

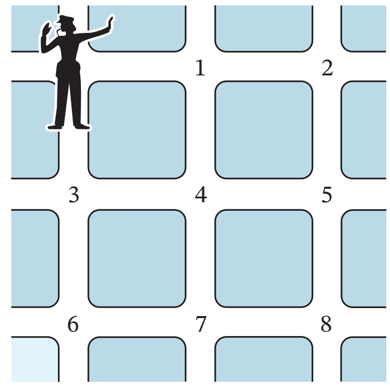
Definition 10.4.2. The state vector for an observation of a Markov chain with k states is a column vector $\mathbf{x}$ whose ith component x_i is the probability that the system is in the ith state at that time.

Theorem 10.4.1. *If P is the transition matrix of a Markov chain and $\mathbf{x}^{(n)}$ is the state vector at the nth observation, then $\mathbf{x}^{(n+1)} = P\mathbf{x}^{(n)}$.*

Example 3. Use the transition matrix from Example 2 to construct the probable future donation record of a new graduate who did not give a donation in the initial year after graduation.

Example 4. Determine whether the state vectors for Example 1 approach a fixed vector.

Example 5. A traffic officer is assigned to control the traffic at the eight intersections indicated in the figure. She is instructed to remain at each intersection for an hour and then to either remain at the same intersection or move to a neighboring intersection. To avoid establishing a pattern, she is told to choose her new intersection on a random basis, with each possible choice equally likely. For example, if she is at intersection 5, her next intersection can be 2, 4, 5, or 8, each with probability $\frac{1}{4}$. Every day she starts at the location where she stopped the day before. Find the transition matrix for this Markov chain and use it to determine whether the state vectors approach a fixed vector.



Example 6. Let

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{x}^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

Determine whether the state vectors approach a fixed vector.

Definition 10.4.3. A transition matrix is <u>regular</u> if some integer power of it has all positive entries.

Theorem 10.4.2 (Behavior of P^n as $n \to \infty$)**.** *If P is a regular transition matrix, then as $n \to \infty$,*

$$P^n \to \begin{bmatrix} q_1 & q_1 & \cdots & q_1 \\ q_2 & q_2 & \cdots & q_2 \\ \vdots & \vdots & & \vdots \\ q_k & q_k & \cdots & q_k \end{bmatrix}$$

where the q_i are positive numbers such that $q_1 + q_2 + \cdots + q_k = 1$.

Theorem 10.4.3 (Behavior of $P^n\mathbf{x}$ as $n \to \infty$)**.** *If P is a regular transition matrix and $\mathbf{x}$ is any probability vector, then as $n \to \infty$,*

$$P^n\mathbf{x} \to \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_k \end{bmatrix} = \mathbf{q}$$

where $\mathbf{q}$ is a fixed probability vector, independent of n, all of whose entries are positive.

Theorem 10.4.4 (Steady-State Vector)**.** *The steady-state vector $\mathbf{q}$ of a regular transition matrix P is the unique probability vector that satisfies the equation $P\mathbf{q} = \mathbf{q}$.*

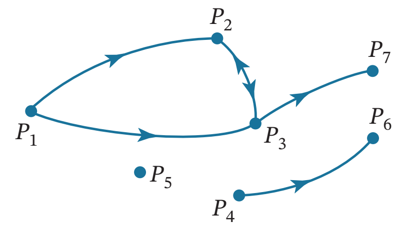
Example 7. Find the steady-state vector from Example 2.

Example 8. Find the steady-state vector from Example 1.

Example 9. Find the steady-state vector from Example 5.

10.5 Graph Theory

Remark 1. A directed graph is a finite set of elements, $\{P_1, P_2, \ldots, P_n\}$, together with a finite collection of ordered pairs (P_i, P_j) of distinct elements of this set, with no ordered pair being repeated. The elements of the set are called vertices, and the ordered pairs are called directed edges, of the directed graph. We use the notation $P_i \to P_j$ (which is read "P_i is connected to P_j") to indicate that the directed edge (P_i, P_j) belongs to the directed graph. Geometrically, we can visualize a directed graph (see the figure) by representing the vertices as points in the plane and representing the directed edge $P_i \to P_j$ by drawing a line or arc from vertex P_i to vertex P_j, with an arrow pointing from P_i to P_j. If both $P_i \to P_j$ and $P_j \to P_i$ hold (denoted $P_i \leftrightarrow P_j$), we draw a single line between P_i and P_j with two oppositely pointing arrows (as with P_2 and P_3 in the figure).

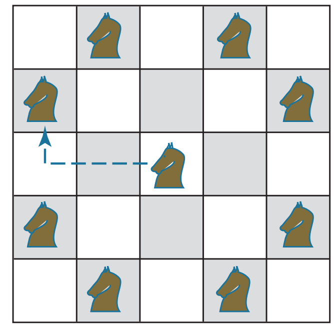


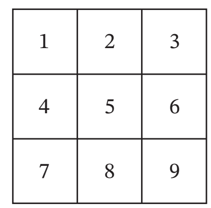
With a directed graph having n vertices, we may associate an $n \times n$ matrix $M = [m_{ij}]$, called the vertex matrix of the directed graph. Its elements are defined by

$$m_{ij} = \begin{cases} 1, & \text{if } P_i \to P_j, \\ 0, & \text{otherwise.} \end{cases}$$

Example 1. A certain family consists of a mother, father, daughter, and two sons. The family members have influence, or power, over each other in the following ways: the mother can influence the daughter and the oldest son; the father can influence the two sons; the daughter can influence the father; the oldest son can influence the youngest son; and the youngest son can influence the mother. We may model this family influence pattern with a directed graph whose vertices are the five family members. If family member A influences family member B, we write $A \to B$. Determine the resulting directed graph and vertex matrix of this directed graph.

Example 2. In chess the knight moves in an "L"-shaped pattern about the chessboard. For the board in the top figure it may move horizontally two squares and then vertically one square, or it may move vertically two squares and then horizontally one square. Thus, from the center square in the figure, the knight may move to any of the eight marked shaded squares. Suppose that the knight is restricted to the nine numbered squares in the bottom figure. If by $i \to j$ we mean that the knight may move from square i to square j, determine the resulting directed graph and vertex matrix that illustrates all possible moves that the knight may make among these nine squares.

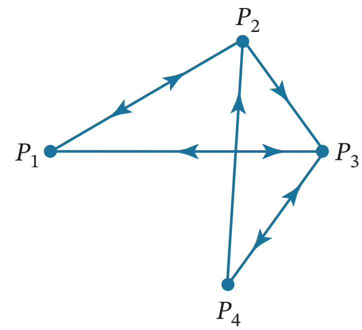




Remark 2. We call $P_i \to P_j$ in a directed graph a 1-step connection and $P_i \to P_k \to P_j$ a 2-step connection. Similarly, we call $P_i \to P_{k_1} \to P_{k_2} \to \cdots \to P_{k_{r-1}} \to P_j$ a r-step connection.

Theorem 10.5.1. *Let M be the vertex matrix of a directed graph and let $m_{ij}^{(r)}$ be the (i, j)-th element of M^r. Then $m_{ij}^{(r)}$ is equal to the number of r-step connections from P_i to P_j.*

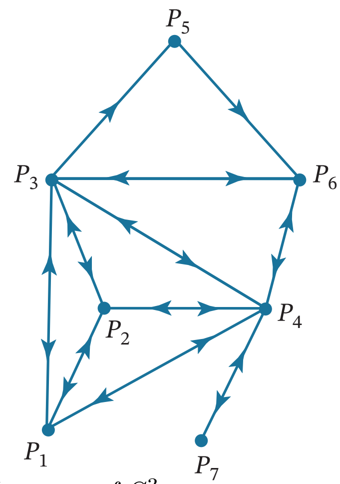
Example 3. The figure is the route map of a small airline that services the four cities P_1, P_2, P_3, P_4. Find the number of 1, 2, or 3-step connections from P_4 to P_3.



Definition 10.5.1. A subset of a directed graph is called a clique if it satisfies the following three conditions:

(i) The subset contains at least three vertices.
(ii) For each pair of vertices P_i and P_j in the subset, both $P_i \to P_j$ and $P_j \to P_i$ are true.
(iii) The subset is as large as possible; that is, it is not possible to add another vertex to the subset and still satisfy condition (ii).

Example 4. What are the cliques for the graph illustrated in the figure?



Remark 3. The matrix $S = [s_{ij}]$ related to a given directed graph is defined as follows:

$$s_{ij} = \begin{cases} 1, & \text{if } P_i \leftrightarrow P_j, \\ 0, & \text{otherwise.} \end{cases}$$

Theorem 10.5.2 (Identifying Cliques)**.** *Let $s_{ij}^{(3)}$ be the (i, j)-th element of S^3. Then a vertex P_i belongs to some clique if and only if $s_{ii}^{(3)} \neq 0$.*

Proof. If $s_{ii}^{(3)} \neq 0$, then there is at least one 3-step connection from P_i to itself in the modified directed graph determined by S. Suppose it is $P_i \to P_j \to P_k \to P_i$. In the modified directed graph, all of the directed relations are two-way, so we also have the connections $P_i \leftrightarrow P_j \leftrightarrow P_k \leftrightarrow P_i$. But this means that $\{P_i, P_j, P_k\}$ is either a clique or a subset of a clique. In either case, P_i must belong to some clique. The converse statement, "if P_i belongs to a clique, then $s_{ii}^{(3)} \neq 0$," follows in a similar manner. $\square$

Example 5. Suppose that a directed graph has as its vertex matrix

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

What are the cliques of the directed graph?

Example 6. Suppose that a directed graph has as its vertex matrix

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

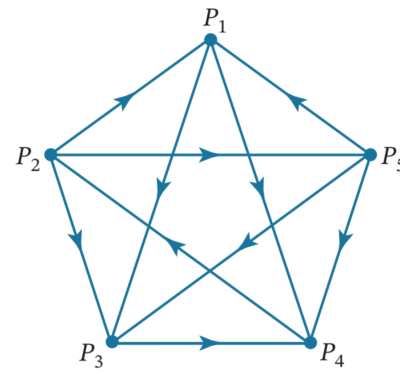
What are the cliques of the directed graph?

Definition 10.5.2. A dominance-directed graph is a directed graph such that for any distinct pair of vertices P_i and P_j, either $P_i \to P_j$ or $P_j \to P_i$, but not both.

Theorem 10.5.3 (Connections in Dominance-Directed Graphs)**.** *In any dominance-directed graph, there is at least one vertex from which there is a* 1*-step or* 2*-step connection to any other vertex.*

Proof. Consider a vertex (there may be several) with the largest total number of 1-step and 2-step connections to other vertices in the graph. By renumbering the vertices, we may assume that P_1 is such a vertex. Suppose there is some vertex P_i such that there is no 1-step or 2-step connection from P_1 to P_i. Then, in particular, $P_1 \to P_i$ is not true, so that by definition of a dominance-directed graph, it must be that $P_i \to P_1$. Next, let P_k be any vertex such that $P_1 \to P_k$ is true. Then we cannot have $P_k \to P_i$, as then $P_1 \to P_k \to P_i$ would be a 2-step from P_1 to P_i. Thus, it must be that $P_i \to P_k$. That is, P_i has 1-step connections to all the vertices to which P_1 has 1-step connections. The vertex P_i must then also have 2-step connections to all the vertices to which P_1 has 2-step connections. But because, in addition, we have that $P_i \to P_1$, this means that P_i has more 1-step connections and 2-step connections to other vertices than does P_1. However, this contradicts the way in which P_1 was chosen. Hence, there can be no vertex P_i to which P_1 has no 1-step or 2-step connection. □

Example 7. Suppose that five baseball teams play each other exactly once, and the results are as indicated in the dominance-directed graph of the figure. Use Theorem 10.5.3 to show that P_2 must have a 1-step or 2-step connection to any other vertex.



Definition 10.5.3. The <u>power</u> of a vertex of a dominance-directed graph is the total number of 1-step and 2-step connections from it to other vertices. Alternatively, the power of a vertex P_i is the sum of the entries of the ith row of the matrix $A = M + M^2$, where M is the vertex matrix of the directed graph.

Example 8. Rank the five baseball teams in Example 7 according to their powers.

10.6 Games of Strategy

Remark 1. In a two-person zero-sum matrix game the term *zero-sum* means that in each play of the game, the positive gain of one player is equal to the negative gain (loss) of the other player. The term *matrix game* is used to describe a two-person game in which each player has only a finite number of moves, so that all possible outcomes of each play, and the corresponding gains of the players, can be displayed in tabular or matrix form.

In a general game of this type, let player R have m possible moves and let player C have n possible moves. In a play of the game, each player makes one of his or her possible moves, and then a payoff is made from player C to player R, depending on the moves. For $i = 1, 2, \dots, m$, and $j = 1, 2, \dots, n$, let us set

$$a_{ij} = \text{payoff that player } C \text{ makes to player } R \text{ if player } R \text{ makes move } i \text{ and player } C \text{ makes move } j.$$

If an entry a_{ij} is negative, we mean that player C receives a payoff of $|a_{ij}|$ from player R. We arrange these mn possible payoffs in the form of an $m \times n$ matrix

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

which we call the payoff matrix of the game.

Each player is to make his or her moves on a probabilistic basis. In the general case we make the following definitions:

$$p_i = \text{probability that player } R \text{ makes move } i \quad (i = 1, 2, \dots, m)$$
$$q_j = \text{probability that player } C \text{ makes move } j \quad (j = 1, 2, \dots, n).$$

With the probabilities p_i and q_j we form two vectors:

$$\mathbf{p} = \begin{bmatrix} p_1 & p_2 \cdots & p_m \end{bmatrix} \quad \text{and} \quad \mathbf{q} = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix}.$$

We call the row vector $\mathbf{p}$ the strategy of player R and the column vector $\mathbf{q}$ the strategy of player C.

If we multiply each possible payoff by its corresponding probability and sum over all possible payoffs, we obtain the expression

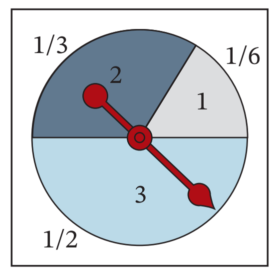
$$a_{11}p_1q_1 + a_{12}p_1q_2 + \cdots + a_{1n}p_1q_n + a_{21}p_2q_1 + \cdots + a_{mn}p_mq_n,$$

which is a weighted average of the payoffs to player R called the expected payoff to player R. We denote this expected payoff by $E(\mathbf{p}, \mathbf{q})$ to emphasize the fact that it depends on the strategies of the two players. From the definition of the payoff matrix A and the strategies $\mathbf{p}$ and $\mathbf{q}$, it can be verified that we may express the expected payoff in matrix notation as

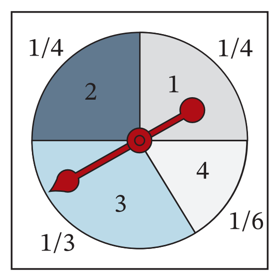
$$E(\mathbf{p}, \mathbf{q}) = \begin{bmatrix} p_1 & p_2 \cdots & p_m \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_n \end{bmatrix} = \mathbf{p}A\mathbf{q}.$$

Because $E(\mathbf{p}, \mathbf{q})$ is the expected payoff to player R, it follows that $-E(\mathbf{p}, \mathbf{q})$ is the expected payoff to player C.

Example 1. Consider the following carnival-type game where each player has a stationary wheel with a movable pointer on it as in the figure. We will call player R's wheel the *row-wheel* and player C's wheel the *column-wheel*. The row-wheel is divided into three sectors numbered 1, 2, and 3, and the column-wheel is divided into four sectors numbered 1, 2, 3, and 4. The fractions of the area occupied by the various sectors are indicated in the figure. To play the game, each player spins the pointer of his or her wheel and lets it come to rest at random. The number of the sector in which each pointer comes to rest is called the move of that player. Depending on the move each player makes, player C then makes a payment of money to player R according to Table 1.



Row-wheel of player R



Column-wheel of player C

Table 1 Payment to Player R

		Player C's Move			
		1	2	3	4
Player R's Move	1	\$3	\$5	−\$2	−\$1
	2	−\$2	\$4	−\$3	−\$4
	3	\$6	−\$5	\$0	\$3

Find the expected payoff to player R.

Theorem 10.6.1 (Fundamental Theorem of Zero-Sum Games)**.** *There exist strategies* $\mathbf{p}^*$ *and* $\mathbf{q}^*$ *such that*

$$E(\mathbf{p}^*, \mathbf{q}) \geq E(\mathbf{p}^*, \mathbf{q}^*) \geq E(\mathbf{p}, \mathbf{q}^*)$$

for all strategies $\mathbf{p}$ *and* $\mathbf{q}$.

Definition 10.6.1. If $\mathbf{p}^*$ and $\mathbf{q}^*$ are strategies such that

$$E(\mathbf{p}^*, \mathbf{q}) \geq E(\mathbf{p}^*, \mathbf{q}^*) \geq E(\mathbf{p}, \mathbf{q}^*)$$

for all strategies $\mathbf{p}$ and $\mathbf{q}$, then

(i) $\mathbf{p}^*$ is called an optimal strategy for player R.
(ii) $\mathbf{q}^*$ is called an optimal strategy for player C.
(iii) $v = E(\mathbf{p}^*, \mathbf{q}^*)$ is called the value of the game.

Definition 10.6.2. An entry a_{rs} in a payoff matrix A is called a saddle point if

(i) a_{rs} is the smallest entry in its row, and
(ii) a_{rs} is the largest entry in its column.

A game whose payoff matrix has a saddle point is called strictly determined.

Remark 2. If a matrix has a saddle point a_{rs}, it turns out that the following strategies are optimal strategies for the two players:

$$\mathbf{p}^* = \begin{bmatrix} 0 & 0 & \cdots & \underset{\nearrow\atop \boldsymbol{r}\textbf{th entry}}{1} & \cdots & 0 \end{bmatrix}, \qquad \mathbf{q}^* = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \\ \\ \\ \leftarrow \boldsymbol{s}\textbf{th entry} \\ \\ \\ \end{matrix}$$

That is, an optimal strategy for player R is to always make the rth move, and an optimal strategy for player C is to always make the sth move. Such strategies for which only one move is possible are called pure strategies. Strategies for which more than one move is possible are called mixed strategies.

Example 2. Two competing television networks, R and C, are scheduling one-hour programs in the same time period. Network R can schedule one of three possible programs, and network C can schedule one of four possible programs. Neither network knows which program the other will schedule. Both networks ask the same outside polling agency to give them an estimate of how all possible pairings of the programs will divide the viewing audience. The agency gives them each Table 2, whose (i, j)-th entry is the percentage of the viewing audience that will watch network R if network R's program i is paired against network C's program j. What program should each network schedule in order to maximize its viewing audience?

Table 2 Audience Percentage for Network R

		Network C's Program			
		1	2	3	4
Network R's Program	1	60	20	30	55
	2	50	75	45	60
	3	70	45	35	30

Theorem 10.6.2 (Optimal Strategies for a 2×2 Matrix Game)**.** *For a 2×2 game that is not strictly determined, optimal strategies for players R and C are*

$$\mathbf{p}^* = \begin{bmatrix} \dfrac{a_{22} - a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}} & \dfrac{a_{11} - a_{12}}{a_{11} + a_{22} - a_{12} - a_{21}} \end{bmatrix}$$

and

$$\mathbf{q}^* = \begin{bmatrix} \dfrac{a_{22} - a_{12}}{a_{11} + a_{22} - a_{12} - a_{21}} \\ \dfrac{a_{11} - a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}} \end{bmatrix}.$$

The value of the game is

$$v = \frac{a_{11}a_{22} - a_{12}a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}}.$$

Example 3. The federal government desires to inoculate its citizens against a certain flu virus. The virus has two strains, and the proportions in which the two strains occur in the virus population is not known. Two vaccines have been developed and each citizen is given only one of them. Vaccine 1 is 85% effective against strain 1 and 70% effective against strain 2. Vaccine 2 is 60% effective against strain 1 and 90% effective against strain 2. What inoculation policy should the government adopt?

10.7 Forest Management

Remark 1. The optimal sustainable yield of a forest is the largest yield that can be attained continually without depleting the forest. The column vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

is called the nonharvest vector, where x_i are the number of trees within the ith class that remain after each harvest and p_i is the economic value of a tree in the ith class. We set

$$x_1 + x_2 + \cdots + x_n = s$$

where s is predetermined by the amount of land available and the amount of space each tree requires. We define the following growth parameters g_i for $i = 1, 2, \ldots, n-1$:

$$\begin{aligned} g_i = {} & \text{the fraction of trees in the } i\text{th class that grow into} \\ & \text{the } (i+1)\text{-st class during a growth period.} \end{aligned}$$

Assuming that a tree can move at most one height class upward in one growth period, we form the following $n \times n$ growth matrix:

$$G = \begin{bmatrix} 1-g_1 & 0 & 0 & & \cdots & 0 \\ g_1 & 1-g_2 & 0 & & \cdots & 0 \\ 0 & g_2 & 1-g_3 & & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1-g_{n-1} & 0 \\ 0 & 0 & 0 & \cdots & g_{n-1} & 1 \end{bmatrix}.$$

The column vector

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

is called the harvest vector, where y_i are the number of trees removed from the ith class. If we define the following $n \times n$ replacement matrix

$$R = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

then the equation

$$G\mathbf{x} - \mathbf{y} + R\mathbf{y} = \mathbf{x}$$

represents the sustainable harvesting condition.

Theorem 10.7.1 (Optimal Sustainable Yield). *The optimal sustainable yield is achieved by harvesting all the trees from one particular height class and none of the trees from any other height class.*

Theorem 10.7.2 (Finding the Optimal Sustainable Yield). *The optimal sustainable yield is the largest value of*

$$\frac{p_k s}{\dfrac{1}{g_1} + \dfrac{1}{g_2} + \cdots + \dfrac{1}{g_{k-1}}}$$

for $k = 2, 3, \ldots, n$. The corresponding value of k is the number of the class that is completely harvested.

Example 1. For a Scots pine forest in Scotland with a growth period of six years, the following growth matrix was found (see M. B. Usher, "A Matrix Approach to the Management of Renewable Resources, with Special Reference to Selection Forests," *Journal of Applied Ecology,* vol. 3, 1966, pp. 355-367):

$$G = \begin{bmatrix} .72 & 0 & 0 & 0 & 0 & 0 \\ .28 & .69 & 0 & 0 & 0 & 0 \\ 0 & .31 & .75 & 0 & 0 & 0 \\ 0 & 0 & .25 & .77 & 0 & 0 \\ 0 & 0 & 0 & .23 & .63 & 0 \\ 0 & 0 & 0 & 0 & .37 & 1.00 \end{bmatrix}.$$

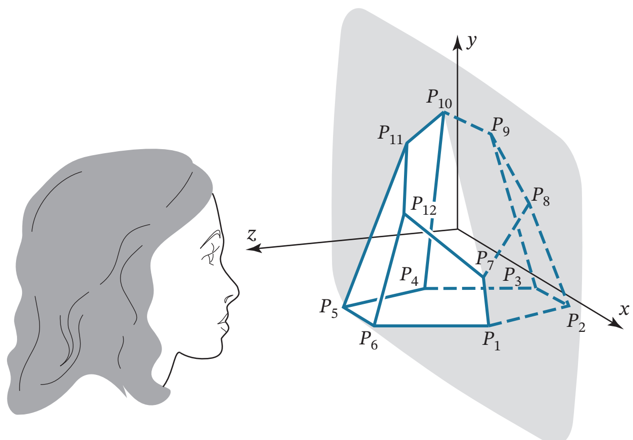
Suppose that the prices of trees in the five tallest height classes are

$$p_2 = \$50, \qquad p_3 = \$100, \qquad p_4 = \$150, \qquad p_5 = \$200, \qquad p_6 = \$250.$$

Which class should be completely harvested to obtain the optimal sustainable yield, and what is that yield?

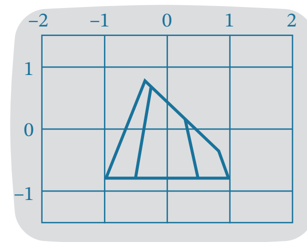
10.8 Computer Graphics

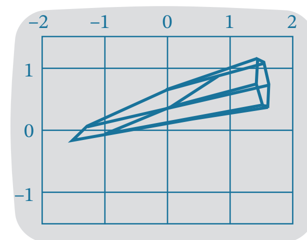
Remark 1. Suppose that we want to visualize a three-dimensional object by displaying various views of it on a video screen. The object we have in mind to display is to be determined by a finite number of straight line segments. As an example, consider the truncated right pyramid with hexagonal base illustrated in the figure.



We first introduce an xyz-coordinate system in which to embed the object. As in the figure, we orient the coordinate system so that its origin is at the center of the video screen and the xy-plane coincides with the plane of the screen. Consequently, an observer will see only the projection of the view of the three-dimensional object onto the two-dimensional xy-plane.

Example 1. The top view represents line segments of the truncated right pyramid with hexagonal base as they would appear on a video screen.





(a) The bottom view is the top view subject to the following five transformations:

1. Scale by a factor of $\frac{1}{2}$ in the x-direction, 2 in the y-direction, and $\frac{1}{3}$ in the z-direction.
2. Translate $\frac{1}{2}$ unit in the x-direction.
3. Rotate 20° about the x-axis.
4. Rotate -45° about the y-axis.
5. Rotate 90° about the z-axis.

Construct the five matrices M_1, M_2, M_3, M_4, and M_5 associated with these five transformations.

(b) If P is the coordinate matrix of the original view and P' is the coordinate matrix of the transformed view, express P' in terms of M_1, M_2, M_3, M_4, M_5, and P.

10.9 Equilibrium Temperature Distributions

Theorem 10.9.1 (The Mean-Value Property)**.** *Let a plate be in thermal equilibrium and let P be a point inside the plate. Then if C is any circle with center at P that is completely contained in the plate, the temperature at P is the average value of the temperature on the circle.*

Remark 1. A plate can be overlaid with a succession of finer and finer square nets or meshes. The points of intersection of the net lines are called *mesh points*. We classify them as boundary mesh points if they fall on the boundary of the plate or as interior mesh points if they lie in the interior of the plate.

Theorem 10.9.2 (Discrete Mean-Value Property)**.** *At each interior mesh point, the temperature is approximately the average of the temperatures at the four neighboring mesh points.*

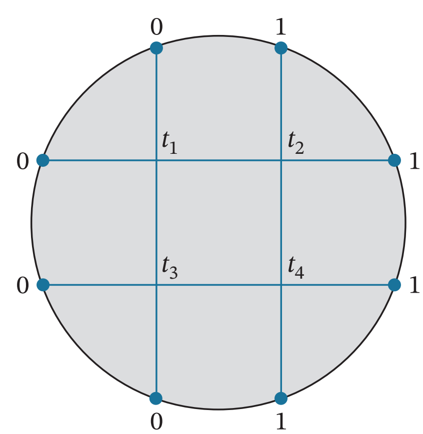
Remark 2. The technique of generating successive approximations to the solution of the equation

$$\mathbf{t} = M\mathbf{t} + \mathbf{b}$$

where $\mathbf{t}$ and $\mathbf{b}$ are column vectors whose numbers of entries are equal to the number of interior mesh points and M is a matrix whose number of rows and columns is equal to the number of interior mesh points, is called Jacobi iteration.

Example 1. A plate in the form of a circular disk has boundary temperatures of $0°$ on the left of its circumference and $1\,\text{deg}$ on the right half of its circumference. A net with four interior mesh points is overlaid on the disk (see the figure).

(a) Using the discrete mean-value property, write the 4×4 linear system $\mathbf{t} = M\mathbf{t} + \mathbf{b}$ that determines the approximate temperatures at the four interior mesh points.



(b) Solve the linear system in part (a).

(c) Use the Jacobi iteration scheme with $\mathbf{t}^{(0)} = \mathbf{0}$ to generate the iterates $\mathbf{t}^{(1)}$, $\mathbf{t}^{(2)}$, $\mathbf{t}^{(3)}$, $\mathbf{t}^{(4)}$, and $\mathbf{t}^{(5)}$ for the linear system in part (a). What is the "error vector" $\mathbf{t}^{(5)} - \mathbf{t}$, where $\mathbf{t}$ is the solution found in part (b)?

(d) By certain advanced methods, it can be determined that the exact temperatures to four decimal places at the four mesh points are $t_1 = t_3 = .2871$ and $t_2 = t_4 = .7129$. What are the percentage errors in the values found in part (b)?

Remark 3. By a discrete random walk along a net we mean a directed path along the net lines that joins a succession of mesh points such that the direction of departure from each mesh point is chosen at random.

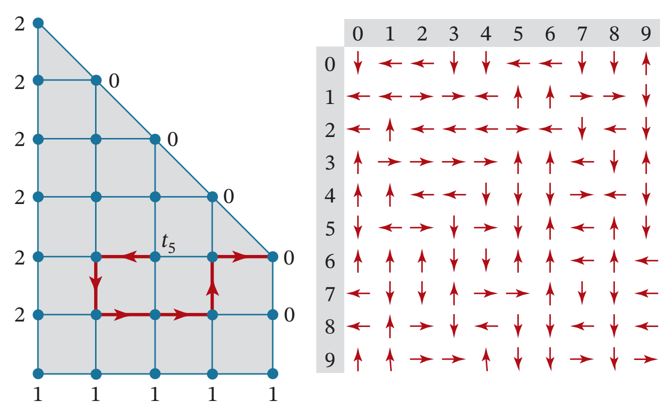
Theorem 10.9.3 (Random Walk Property)**.** *Let $W_1, W_2, \ldots, W_n$ be a succession of random walks, all of which begin at a specified interior mesh point. Let $t_1^*, t_2^*, \ldots, t_n^*$ be the temperatures at the boundary mesh points first encountered along each of these random walks. Then the average value $(t_1^* + t_2^* + \cdots + t_n^*)/n$ of these boundary temperatures approaches the temperature at the specified interior mesh point as the number of random walks n increases without bound.*

Example 2. The random walk illustrated in Figure (a) can be described by six arrows

$$\leftarrow\downarrow\rightarrow\rightarrow\uparrow\rightarrow$$

that specify the directions of departure from the successive mesh points along the path. Figure (b) is an array of 100 computer-generated, random oriented arrows arranged in a 10×10 array. Use these arrows to determine random walks to approximate the temperature t_5. Proceed as follows:

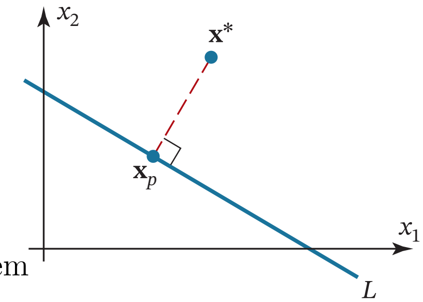
1. Take the last two digits of your telephone number. Use the last digit to specify a row and the other to specify a column.
2. Go to the arrow in the array with that row and column number.
3. Using this arrow as a starting point, move through the array of arrows as you would read a book (left to right and top to bottom). Beginning at the point labeled t_5 in Figure (a) and using this sequence of arrows to specify a sequence of directions, move from mesh point to mesh point until you reach a boundary mesh point. This completes your first random walk. Record the temperature at the boundary mesh point. (If you reach the end of the arrow array, continue with the arrow in the upper left corner.)
4. Return to the interior mesh point labeled t_5 and begin where you left off in the arrow array; generate your next random walk. Repeat this process until you have completed 10 random walks and have recorded 10 boundary temperatures.
5. Calculate the average of the 10 boundary temperatures recorded. (The exact value is $t_5 = .7491$.)



(a) *(b)*

10.10 Computed Tomography

Theorem 10.10.1 (Orthogonal Projection Formula)**.** *Let L be a line in R^2 with equation $\mathbf{a}^T\mathbf{x} = b$, and let $\mathbf{x}^*$ be any point in R^2 (see the figure). Then the orthogonal projection, $\mathbf{x}_p$, of $\mathbf{x}^*$ onto L is given by*

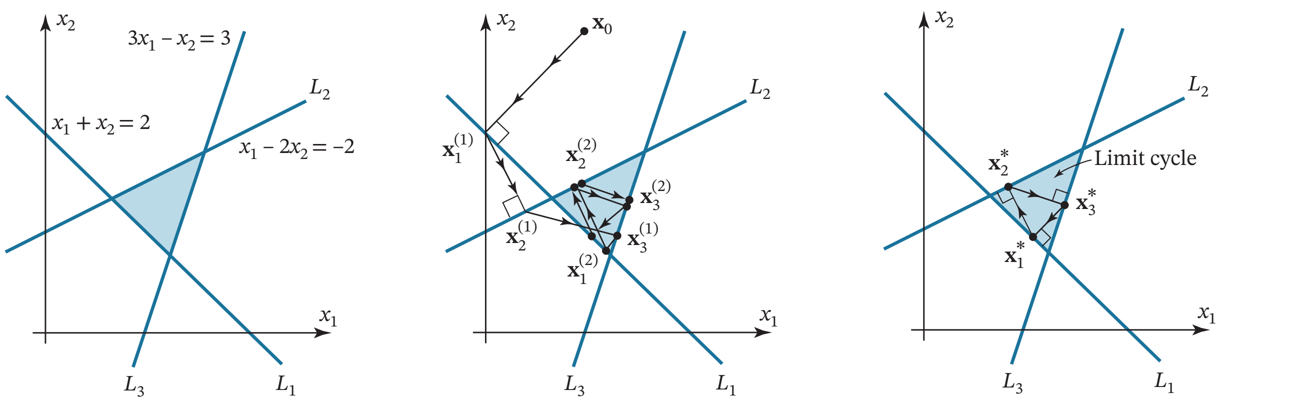


$$\mathbf{x}_p = \mathbf{x}^* + \frac{(b - \mathbf{a}^T\mathbf{x}^*)}{\mathbf{a}^T\mathbf{a}}\mathbf{a}.$$

Example 1. Find an approximate solution of the linear system

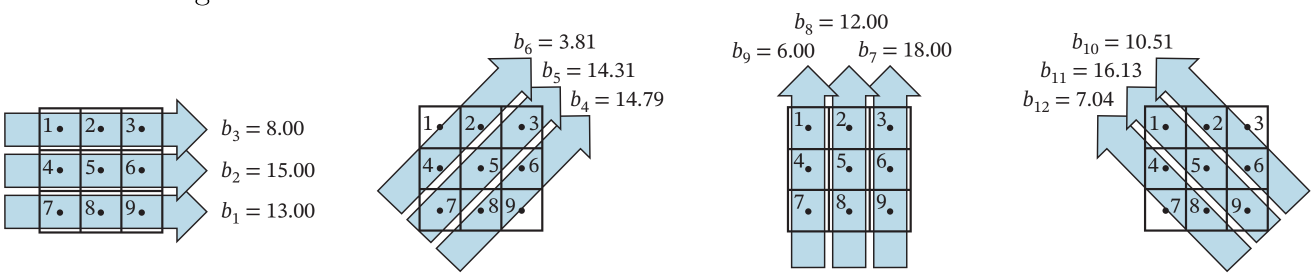
$$\begin{aligned} L_1 : \quad x_1 + x_2 &= 2 \\ L_2 : \quad x_1 - 2x_2 &= -2 \\ L_3 : \quad 3x_1 - x_2 &= 3 \end{aligned}$$

shown in the figure.



(a) **(b)** **(c)**

Example 2. Find the unknown pixel densities of the 9 pixels arranged in the 3×3 array illustrated in the figure. These 9 pixels are scanned using the parallel mode with 12 beams whose measured beam densities are indicated in the figure.



10.11 Fractals

Remark 1. We call a set in R^2 bounded if it can be enclosed by a suitably large circle and closed if it contains all of its boundary points. Two sets in R^2 will be called congruent if they can be made to coincide exactly by translating and rotating them appropriately within R^2.

If $T : R^2 \to R^2$ is the linear operator that scales by a factor of s (see Table 7 of Section 4.9), and if Q is a set in R^2, then the set $T(Q)$ (the set of images of points in Q under T is called a dilation of the set Q if $s > 1$ and a contraction of Q if $0 < s < 1$. In either case we say that $T(Q)$ is the set Q scaled by the factor s.

Definition 10.11.1. A closed and bounded subset of the Euclidean plane R^2 is said to be self-similar if it can be expressed in the form

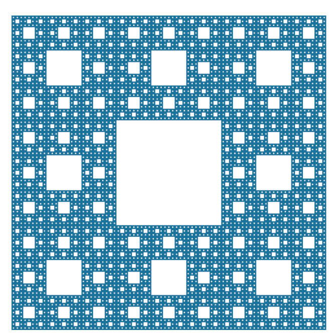
$$S = S_1 \cup S_2 \cup S_3 \cup \cdots \cup S_k$$

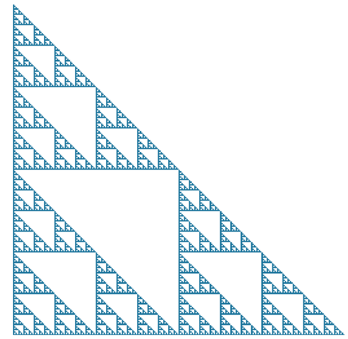
where $S_1, S_2, S_3, \ldots, S_k$ are nonoverlapping sets, each of which is congruent to S scaled by the same factor s $(0 < s < 1)$.

Example 1. A line segment in R^2 can be expressed as the union of two nonoverlapping congruent line segments. Determine the values of k and s for this self-similar set.

Example 2. A square can be expressed as the union of four nonoverlapping congruent squares. Determine the values of k and s for this self-similar set.

Example 3. The set suggested by the figure, the Sierpinski "carpet," was first described by the Polish mathematician Waclaw Sierpinski (1882–1969). Determine the values of k and s for this self-similar set.





Example 4. The figure illustrates another set described by Sierpinski. Determine the values of k and s for this self-similar set.

Remark 2. The definition of the dimension of a subspace given in Section 4.5 is a special case of a more general concept called topological dimension, which is applicable to sets in R^n that are not necessarily subspaces. We denote the topological dimension of a set S by $d_T(S)$.

Example 5. What are the topological dimensions of the sets given in Examples 1-4?

Definition 10.11.2. The Hausdorff dimension of a self-similar set S is denoted by $d_H(S)$ and is defined by

$$d_H(S) = \frac{\ln k}{\ln(1/s)}.$$

Example 6. What are the Hausdorff dimensions of the sets given in Examples 1-4?

Definition 10.11.3. A fractal is a subset of a Euclidean space whose Hausdorff dimension and topological dimension are not equal.

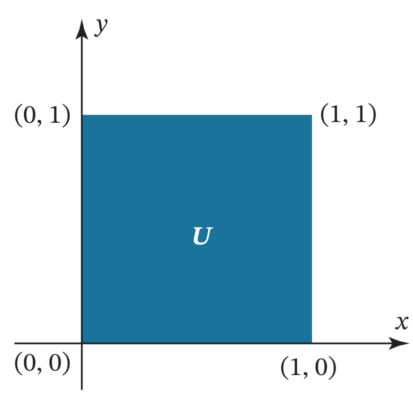
Definition 10.11.4. A similitude with scale factor s is a mapping of R^2 into R^2 of the form

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = s\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix}$$

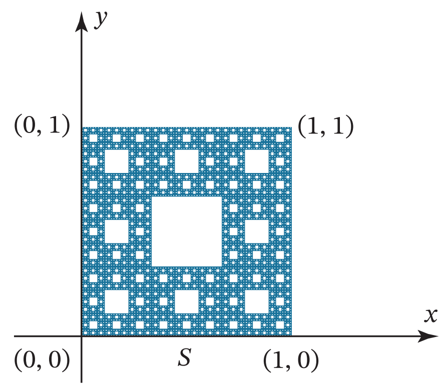
where s, θ, e, and f are scalars.

Example 7. Consider the line segment S connecting the points $(0,0)$ and $(1,0)$ in the xy-plane. Find similitudes whose union is S.

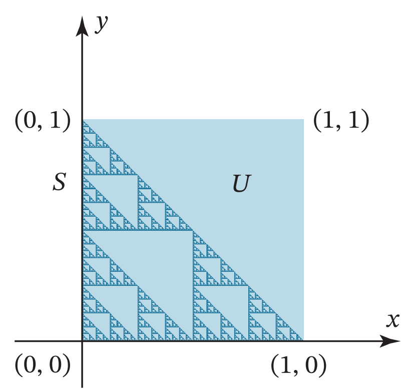
Example 8. Consider the unit square U in the xy-plane, as shown in the figure. Find similitudes whose union is U.



Example 9. Consider the Sierpinski carpet S over the unit square U of the xy-plane, as shown in the figure. Find similitudes whose union is S.



Example 10. Consider the Sierpinski triangle S fitted inside the unit square U of the xy-plane, as shown in the figure. Find similitudes whose union is S.



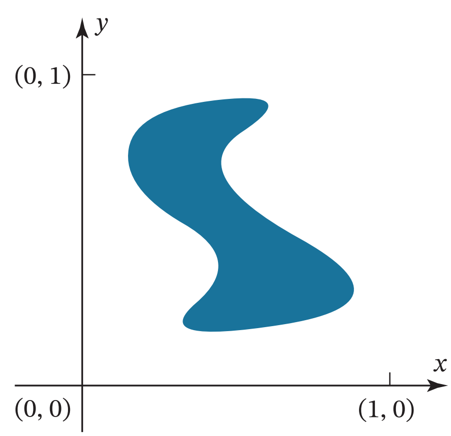
Theorem 10.11.1. *If $T_1, T_2, \ldots, T_k$ are contracting similitudes with the same scale factor, then there is a unique nonempty closed and bounded set in S in the Euclidean plane such that*

$$S = T_1(S) \cup T_2(S) \cup T_3(S) \cup \cdots \cup T_k(S).$$

Furthermore, if the sets $T_1(S), T_2(S), T_3(S), \ldots, T_k(S)$ are nonoverlapping, then S is self-similar.

Example 11. Use similitudes to construct the Sierpinski carpet, starting with the unit square in the xy-plane.

Example 12. Use similitudes to construct the Sierpinski triangle, starting with the arbitrary closed and bounded set S_0 in the figure.



S_0

Example 13. Consider the following two similitudes:

$$T_1\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$T_2\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \frac{1}{2}\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} .3 \\ .3 \end{bmatrix}$$

Describe the actions of these two similitudes on the unit square U for various values of θ.

Definition 10.11.5. An affine transformation is a mapping of R^2 into R^2 of the form

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e \\ f \end{bmatrix}$$

where a, b, c, d, e, and f are scalars.

10.12 Chaos

Remark 1. Arnold's cat map is the transformation $\Gamma : R^2 \to R^2$ defined by the formula

$$\Gamma : (x, y) \to (x + y, x + 2y) \bmod 1$$

or, in matrix notation,

$$\Gamma\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \bmod 1.$$

Under Arnold's cat map each pixel point of the unit square

$$S = \{(x, y) \mid 0 \leq x < 1, 0 \leq y < 1\}$$

is transformed into another pixel point of S. To see why this is so, observe that the image of the pixel point $(m/p, n/p)$ under Γ is given in matrix form by

$$\Gamma\left(\begin{bmatrix} \frac{m}{p} \\ \frac{n}{p} \end{bmatrix}\right) = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} \frac{m}{p} \\ \frac{n}{p} \end{bmatrix} \bmod 1 = \begin{bmatrix} \frac{m+n}{p} \\ \frac{m+2n}{p} \end{bmatrix} \bmod 1.$$

Example 1. Determine the successive iterates of the point $\left(\frac{27}{76}, \frac{58}{76}\right)$ under Arnold's cat map.

Remark 2. We say that a set D of points in S is dense in S if every circle centered at any point of S encloses points of D, no matter how small the radius of the circle is taken. It can be shown that the rational points are dense in S and the iterates of most (but not all) of the irrational points are dense in S.

Definition 10.12.1. A mapping T of S onto itself is said to be <u>chaotic</u> if:

(i) S contains a dense set of periodic points of the mapping T.
(ii) There is a point in S whose iterates under T are dense in S.

10.13 Cryptography

Remark 1. The study of encoding and decoding secret messages is called cryptography. In the language of cryptography, codes are called ciphers, uncoded messages are called plaintext, and coded messages are called ciphertext. The process of converting from plaintext to ciphertext is called enciphering, and the reverse process of converting from ciphertext to plaintext is called deciphering.

The simplest ciphers, called substitution ciphers, are those that replace each letter of the alphabet by a different letter. A system of cryptography in which the plaintext is divided into sets of n letters, each of which is replaced by a set of n cipher letters, is called a polygraphic system. Hill ciphers are a class of polygraphic systems based on matrix transformations.

Example 1. Use the matrix

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

to obtain the Hill cipher for the plaintext message

I AM HIDING.

Definition 10.13.1. If m is a positive integer and a and b are any integers, then we say that a is equivalent to b modulo m, written

$$a = b \pmod{m}$$

if $a - b$ is an integer multiple of m.

Example 2. Determine values for m that make these equivalences true:

$$\begin{aligned} 7 &= 2 && \pmod{m} \\ 19 &= 3 && \pmod{m} \\ -1 &= 25 && \pmod{m} \\ 12 &= 0 && \pmod{m} \end{aligned}$$

Remark 2. For any modulus m it can be proved that every integer a is equivalent, modulo m, to exactly one of the integers

$$0, 1, 2, \ldots, m-1.$$

We call this integer the residue of a modulo m, and we write

$$Z_m = \{0, 1, 2, \ldots, m-1\}$$

to denote the set of residues modulo m.

Theorem 10.13.1. *For any integer a and modulus m, let*

$$R = \textit{remainder of } \frac{|a|}{m}.$$

Then the residue r of a modulo m is given by

$$r = \begin{cases} R & \textit{if } a \geq 0 \\ m - R & \textit{if } a < 0 \textit{ and } R \neq 0 \\ 0 & \textit{if } a < 0 \textit{ and } R = 0. \end{cases}$$

Example 3. Find the residue modulo 26 of (a) 87, (b) -38, and (c) -26.

Definition 10.13.2. If a is a number in Z_m, then a number a^{-1} in Z_m is called a reciprocal or multiplicative inverse of a modulo m if
$aa^{-1} = a^{-1}a = 1 \pmod{m}$.

Example 4. Find the reciprocal of the number 3 modulo 26, if it exists.

Example 5. Find the reciprocal of the number 4 modulo 26, if it exists.

Theorem 10.13.2. *A square matrix A with entries in Z_m is invertible modulo m if and only if the residue of* $\det(A)$ *modulo m has a reciprocal modulo m.*

Theorem 10.13.3. *A square matrix A with entries in Z_m is invertible modulo m if and only if m and the residue of* $\det(A)$ *modulo m have no common prime factors.*

Theorem 10.13.4. *A square matrix A with entries in Z_{26} is invertible modulo* 26 *if and only if the residue of* $\det(A)$ *modulo* 26 *is not divisible by* 2 *or* 13.

Example 6. Find the inverse of

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

modulo 26.

Example 7. Decode the following Hill 2-cipher, which was enciphered by the matrix in Example 6:

$$GTNKGKDUSK.$$

Theorem 10.13.5 (Determining the Deciphering Matrix)**.** *Let* $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ *be linearly independent plaintext vectors, and let* $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n$ *be the corresponding ciphertext vectors in a Hill* n*-cipher. If*

$$P = \begin{bmatrix} \mathbf{p}_1^T \\ \mathbf{p}_2^T \\ \vdots \\ \mathbf{p}_n^T \end{bmatrix}$$

is the $n \times n$ *matrix with row vectors* $\mathbf{p}_1^T, \mathbf{p}_2^T, \ldots, \mathbf{p}_n^T$ *and if*

$$C = \begin{bmatrix} \mathbf{c}_1^T \\ \mathbf{c}_2^T \\ \vdots \\ \mathbf{c}_n^T \end{bmatrix}$$

is the $n \times n$ *matrix with row vectors* $\mathbf{c}_1^T, \mathbf{c}_2^T, \ldots, \mathbf{c}_n^T$*, then the sequence of elementary row operations that reduces* C *to* I *transforms* P *to* $(A^{-1})^T$*.*

Example 8. The following Hill 2-cipher is intercepted:

$$IOSBTGXESPXHOPDE.$$

Decipher the message, given that it starts with the word *DEAR*.

10.14 Genetics

Remark 1. In this section we will assume that inherited traits are governed by a set of two genes, which we designate by A and a. Under autosomal inheritance each individual in the population of either gender possesses two of these genes, the possible pairings being designated AA, Aa, and aa. This pair of genes is called the individual's genotype, and it determines how the trait controlled by the genes is manifested in the individual.

Table 1

Genotype of Offspring	Genotypes of Parents					
	AA–AA	AA–Aa	AA–aa	Aa–Aa	Aa–aa	aa–aa
AA	1	$\frac{1}{2}$	0	$\frac{1}{4}$	0	0
Aa	0	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$	0
aa	0	0	0	$\frac{1}{4}$	$\frac{1}{2}$	1

Example 1. Suppose that a farmer has a large population of plants consisting of some distribution of all three possible genotypes AA, Aa, and aa. The farmer desires to undertake a breeding program in which each plant in the population is always fertilized with a plant of genotype AA and is then replaced by one of its offspring.

For $n = 0, 1, 2, \ldots$, let us set

$$\begin{aligned} a_n &= \text{fraction of plants of genotype } AA \text{ in } n\text{th generation} \\ b_n &= \text{fraction of plants of genotype } Aa \text{ in } n\text{th generation} \\ c_n &= \text{fraction of plants of genotype } aa \text{ in } n\text{th generation.} \end{aligned}$$

Derive an expression for the distribution of the three possible genotypes in the population after any number of generations.

Example 2. Modify Example 1 so that instead of each plant being fertilized with one of genotype AA, each plant is fertilized with a plant of its own genotype. Derive an expression for the distribution of the three possible genotypes in the population after any number of generations.

10.15 Age-Specific Population Growth

Remark 1. Suppose the maximum age attained by any female in a population is L years and divide the population into n age classes. We define the age distribution vector $\mathbf{x}^{(k)}$ at time t_k by

$$\mathbf{x}^{(k)} = \begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \\ \vdots \\ x_n^{(k)} \end{bmatrix}$$

where $x_i^{(k)}$ is the number of females in the ith age class at time t_k. Then

$$\mathbf{x}^{(k)} = L\mathbf{x}^{(k-1)}, \qquad k = 1, 2, \ldots$$

where L is the Leslie matrix

$$L = \begin{bmatrix} a_1 & a_2 & a_3 & \cdots & a_{n-1} & a_n \\ b_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & b_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & b_{n-1} & 0 \end{bmatrix}.$$

Example 1. Suppose that the oldest age attained by the females in a certain animal population is 15 years and we divide the population into three age classes with equal durations of five years. Let the Leslie matrix for this population be

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

If there are initially 1000 females in each of the three age classes, then find the number of females in each age class after 15 years.

Theorem 10.15.1 (Existence of a Positive Eigenvalue)**.** *A Leslie matrix L has a unique positive eigenvalue λ_1. This eigenvalue has multiplicity* 1 *and an eigenvector $\mathbf{x}_1$ all of whose entries are positive.*

Theorem 10.15.2 (Eigenvalues of a Leslie Matrix)**.** *If λ_1 is the unique positive eigenvalue of a Leslie matrix L, and λ_k is any other real or complex eigenvalue of L, then $|\lambda_k| \leq \lambda_1$.*

Example 2. Find the eigenvalues of

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

Theorem 10.15.3 (Dominant Eigenvalue)**.** *If two successive entries a_i and a_{i+1} in the first row of a Leslie matrix L are nonzero, then the positive eigenvalue of L is dominant.*

Example 3. Find the limiting proportion of the age distribution of the population in Example 1.

Example 4. In this example we use birth and parameters from the year 1965 for Canadian females. Because few women over 50 years of age bear children, we restrict ourselves to the portion of the female population between 0 and 50 years of age. The birth and death parameters are as follows:

Age Interval	a_i	b_i
$[0, 5)$	0.00000	0.99651
$[5, 10)$	0.00024	0.99820
$[10, 15)$	0.05861	0.99802
$[15, 20)$	0.28608	0.99729
$[20, 25)$	0.44791	0.99694
$[25, 30)$	0.36399	0.99621
$[30, 35)$	0.22259	0.99460
$[35, 40)$	0.10457	0.99184
$[40, 45)$	0.02826	0.98700
$[45, 50)$	0.00240	—

Using numerical techniques, we can approximate the positive eigenvalue and corresponding eigenvector by

$$\lambda_1 = 1.07622 \quad \text{and} \quad \mathbf{x}_1 = \begin{bmatrix} 1.00000 \\ 0.92594 \\ 0.85881 \\ 0.79641 \\ 0.73800 \\ 0.68364 \\ 0.63281 \\ 0.58482 \\ 0.53897 \\ 0.49429 \end{bmatrix}.$$

Interpret these results in the context of this example.

10.16 Harvesting of Animal Populations

Definition 10.16.1. A harvesting policy in which an animal population is periodically harvested is said to be sustainable if the yield of each harvest is the same and the age distribution of the population remaining after each harvest is the same.

Remark 1. To describe this harvesting model mathematically, let

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

be the age distribution vector of the population at the beginning of the growth period. Then

$$(I - H)L\mathbf{x} = \mathbf{x}$$

where L is the Leslie matrix describing the growth of the population, H is the harvesting matrix

$$H = \begin{bmatrix} h_1 & 0 & 0 & \cdots & 0 \\ 0 & h_2 & 0 & \cdots & 0 \\ 0 & 0 & h_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & h_n \end{bmatrix},$$

and h_i, for $i = 1, 2, \ldots, n$, is the fraction of females from the ith class that is harvested.

Example 1. For a certain species of domestic sheep in New Zealand with a growth period of 1 year, the following Leslie matrix was found (see G. Caughley, "Parameters for Seasonally Breeding Populations," *Ecology*, *48*, 1967, pp. 834-839).

$$L = \begin{bmatrix} .000 & .045 & .391 & .472 & .484 & .546 & .543 & .502 & .468 & .459 & .433 & .421 \\ .845 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .975 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & .965 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & .950 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .926 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .895 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .850 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .786 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .691 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .561 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .370 & 0 \end{bmatrix}.$$

The sheep have a lifespan of 12 years, so they are divided into 12 age classes of duration 1 year each. By the use of numerical techniques, the unique positive eigenvalue of L can be found to be

$$\lambda_1 = 1.176.$$

Determine the uniform harvesting policy for this population.

Example 2. In some populations only the youngest females are of any economic value, so the harvester seeks to harvest only the females from the youngest age class. Apply this type of sustainable harvesting policy to the sheep population in Example 1.

Theorem 10.16.1 (Optimal Sustainable Yield)**.** *An optimal sustainable harvesting policy is one in which either one or two age classes are harvested. If two age classes are harvested, then the older age class is completely harvested.*

10.17 A Least Squares Model for Human Hearing

Theorem 10.17.1. (Minimizing the Mean Square Error on $[0, 2\pi]$). *If $f(t)$ is continuous on $[0, 2\pi]$, then the trigonometric function $g(t)$ of the form*

$$g(t) = \frac{1}{2}a_0 + a_1 \cos t + \cdots + a_n \cos nt + b_1 \sin t + \cdots + b_n \sin nt$$

that minimizes the mean square error

$$\int_0^{2\pi} [f(t) - g(t)]^2 \, dt$$

has coefficients

$$\begin{aligned} a_k &= \frac{1}{\pi} \int_0^{2\pi} f(t) \cos kt \, dt, \qquad k = 0, 1, 2, \ldots, n \\ b_k &= \frac{1}{\pi} \int_0^{2\pi} f(t) \sin kt \, dt, \qquad k = 1, 2, \ldots, n. \end{aligned}$$

Theorem 10.17.2. (Minimizing the Mean Square Error on $[0, T]$). *If $f(t)$ is continuous on $[0, T]$, then the trigonometric function $g(t)$ of the form*

$$g(t) = \frac{1}{2}a_0 + a_1 \cos \frac{2\pi}{T}t + \cdots + a_n \cos \frac{2n\pi}{T}t + b_1 \sin \frac{2\pi}{T}t + \cdots + b_n \sin \frac{2n\pi}{T}t$$

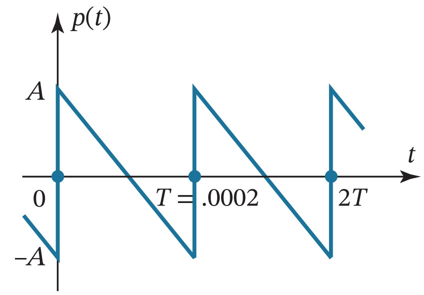
that minimizes the mean square error

$$\int_0^{2\pi} [f(t) - g(t)]^2 \, dt$$

has coefficients

$$\begin{aligned} a_k &= \frac{2}{T} \int_0^{T} f(t) \cos \frac{2k\pi t}{T} \, dt, \qquad k = 0, 1, 2, \ldots, n \\ b_k &= \frac{2}{T} \int_0^{T} f(t) \sin \frac{2k\pi t}{T} \, dt, \qquad k = 1, 2, \ldots, n. \end{aligned}$$

Example 1. Let a sound wave $p(t)$ have a saw-tooth pattern with a basic frequency of 5000 cps (see the figure). Assume units are chosen so that the normal atmospheric pressure is at the zero level and the maximum amplitude of the wave is A. The basic period of the wave is $T = 1/5000 = .0002$ second. From $t = 0$ to $t = T$, the function $p(t)$ has the equation



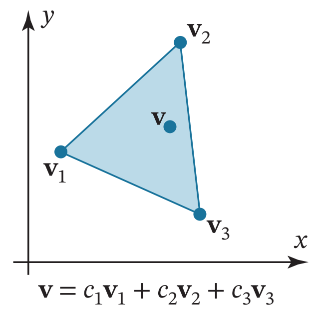
$$p(t) = \frac{2A}{T}\left(\frac{T}{2} - t\right).$$

Investigate how the sound wave $p(t)$ is perceived by the human ear.

10.18 Warps and Morphs

Remark 1. Let the three vertices of a triangle be given by the three noncollinear points $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ (see the figure). We will call this triangle the begin-triangle. If $\mathbf{v}$ is any point in the begin-triangle, then there are unique constants c_1 and c_2 such that

$$\mathbf{v} - \mathbf{v}_3 = c_1(\mathbf{v}_1 - \mathbf{v}_3) + c_2(\mathbf{v}_2 - \mathbf{v}_3).$$



If we set $c_3 = 1 - c_1 - c_2$, then we can rewrite this equation as

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3$$

where

$$c_1 + c_2 + c_3 = 1.$$

We say that $\mathbf{v}$ is a convex combination of the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ if these equations are satisfied and, in addition, the coefficients c_1, c_2, and c_3 are nonnegative. It can be shown that $\mathbf{v}$ lies in the triangle determined by $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ if and only if it is a convex combination of those three vectors.

Example 1. Determine whether the vector $\mathbf{v}$ is a convex combination of the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$.

(a) $\mathbf{v} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \quad \mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 3 \\ 5 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}.$

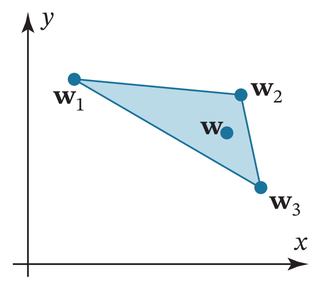
(b) $\mathbf{v} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \quad \mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 3 \\ 5 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}.$

(c) $\mathbf{v} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_1 = \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -2 \\ -2 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 3 \\ 0 \end{bmatrix}.$

(d) $\mathbf{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_1 = \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -2 \\ -2 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 3 \\ 0 \end{bmatrix}.$

Remark 2. Given three noncollinear points $\mathbf{w}_1$, $\mathbf{w}_2$, and $\mathbf{w}_3$ of an end-triangle (see the figure), there is a unique affine transformation that maps $\mathbf{v}_1$ to $\mathbf{w}_1$, $\mathbf{v}_2$ to $\mathbf{w}_2$, and $\mathbf{v}_3$ to $\mathbf{w}_3$. That is, there is a unique 2×2 invertible matrix M and a unique vector $\mathbf{b}$ such that

$$\mathbf{w}_i = M\mathbf{v}_i + \mathbf{b} \quad \text{for } i = 1, 2, 3.$$



$\mathbf{w} = c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + c_3\mathbf{w}_3$

Moreover, it can be shown that the image $\mathbf{w}$ of the vector $\mathbf{v}$ under this affine transformation is

$$\mathbf{w} = c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + c_3\mathbf{w}_3.$$

To each point in the begin-triangle we assign a gray level, say 0 for white and 100 for black, with any other gray level lying between 0 and 100. In particular, let a scalar-valued function ρ_0, called the picture-density of the begin-triangle, be defined so that $\rho_0(\mathbf{v})$ is the gray level at the point $\mathbf{v}$ in the begin-triangle. We can now define a picture in the end-triangle, called a warp of the original picture, with a picture-density ρ_1 by defining the gray level at the point $\mathbf{w}$ within the end-triangle to be the gray level of the point $\mathbf{v}$ in the begin-triangle that maps onto $\mathbf{w}$. In equation form, the picture-density ρ_1 is determined by

$$\rho_1(\mathbf{w}) = \rho_0(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3).$$

Suppose we are given a picture contained within some rectangular region of the plane. We choose n points $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ within the rectangle, which we call vertex points, so that they fall on key elements or features of the picture we wish to warp. Once the vertex points are chosen, we complete a triangulation of the region.

A time-varying warp is the set of warps generated when the vertex points of a beginning picture are moved continually in time from their original positions to specified final positions. A time-varying morph can be described as a blending of two time-varying warps of two different pictures using two triangulations that match corresponding features in the two pictures.

Example 2. Find the 2×2 matrix M and two-dimensional vector $\mathbf{b}$ that define the affine transformation that maps the three vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ to the three vectors $\mathbf{w}_1$, $\mathbf{w}_2$, and $\mathbf{w}_3$. Do this by setting up a system of six linear equations for the four entries of the matrix M and the two entries of the vector $\mathbf{b}$.

(a) $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2 \\ 1 \end{bmatrix},$

$\mathbf{w}_1 = \begin{bmatrix} 4 \\ 3 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} 9 \\ 5 \end{bmatrix}, \quad \mathbf{w}_3 = \begin{bmatrix} 5 \\ 3 \end{bmatrix}.$

(b) $\mathbf{v}_1 = \begin{bmatrix} -2 \\ 2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2 \\ 1 \end{bmatrix},$

$\mathbf{w}_1 = \begin{bmatrix} -8 \\ 1 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \mathbf{w}_3 = \begin{bmatrix} 5 \\ 4 \end{bmatrix}.$

(c) $\mathbf{v}_1 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 3 \\ 5 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$

$\mathbf{w}_1 = \begin{bmatrix} 0 \\ -2 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} 5 \\ 2 \end{bmatrix}, \quad \mathbf{w}_3 = \begin{bmatrix} 3 \\ -3 \end{bmatrix}.$

(d) $\mathbf{v}_1 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -4 \\ -2 \end{bmatrix},$

$\mathbf{w}_1 = \begin{bmatrix} \frac{5}{2} \\ -1 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} \frac{7}{2} \\ 3 \end{bmatrix}, \quad \mathbf{w}_3 = \begin{bmatrix} -\frac{7}{2} \\ -9 \end{bmatrix}.$

10.19 Internet Search Engines

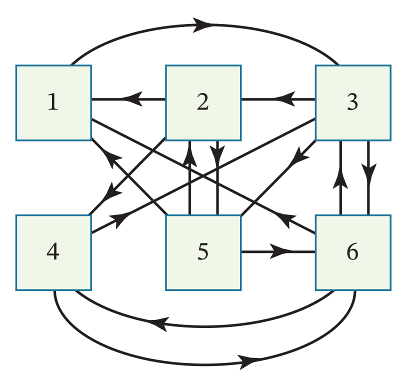
Remark 1. A network of links between web pages can provide a means of measuring their relative importance. A diagram called a webgraph shows the links among the web pages. A directed path from the ith page to the jth page means that the ith page has an outgoing link to the jth page (i.e., it references that page). The adjacency matrix of a webgraph with n pages is the $n \times n$ matrix A whose ijth entry a_{ij} is 1 if the jth page has an outgoing link to the ith page and 0 otherwise.

Definition 10.19.1. If a webgraph with n pages is "surfed" by clicking a mouse, then the state vector $\mathbf{x}^{(k)}$ is the $n \times 1$ column vector whose ith entry is the probability that the surfer is on the ith page after k random mouse clicks.

Definition 10.19.2. The probability transition matrix $B = [b_{ij}]$ associated with an adjacency matrix $A = [a_{ij}]$ is the matrix obtained by dividing each entry of A by the sum of the entries in the same column; that is,

$$b_{ij} = \frac{a_{ij}}{\sum_{k=1}^{n} a_{kj}}.$$

Example 1. Suppose we know with certainty that Alice is initially on Page 2 of the webgraph in the figure. Determine her successive state vectors.

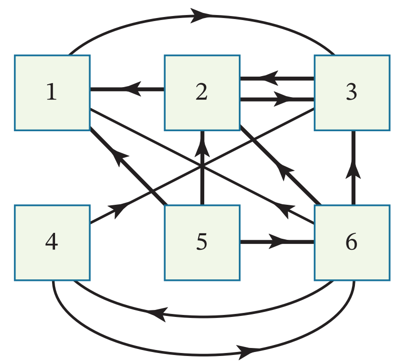


Remark 2. Suppose there is a probability δ, called the damping factor, that one will go to the next page in a network by choosing a link on the current page and a probability of $1-\delta$ that the next page will be chosen at random. Then there is a new probability transition matrix $M = [m_{ij}]$ in which

$$m_{ij} = \delta b_{ij} + \frac{1-\delta}{n}$$

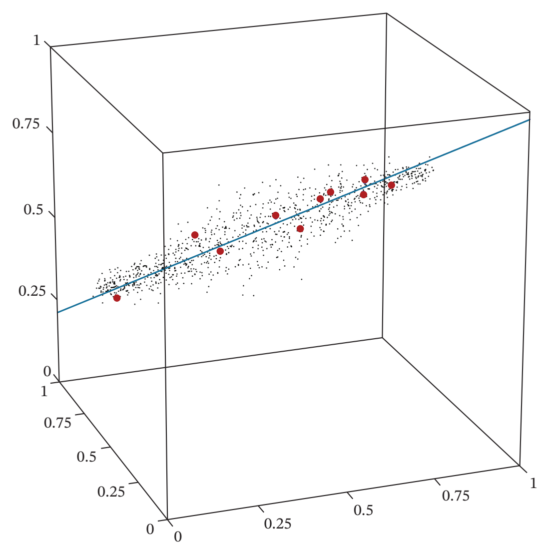
with b_{ij} as given in Definition 10.19.2.

Example 2. Consider the webgraph in the figure. Determine the successive state vectors with and without a damping factor of $\delta = 0.85$.



10.20 Facial Recognition

Example 1. Suppose we have grayscale images of 1000 faces and we reduce the resolution of each image to only 3 pixels. The facial images can then be represented by vectors (or points) $\mathbf{p}_i$, $i = 1, 2, \ldots, 1000$ in the unit box in R^3 (see the figure).
Let us pick the following 10 face vectors as our training set (shown as red dots in the figure).



$$\mathbf{p}_1 = \begin{bmatrix} 0.7272 \\ 0.2826 \\ 0.7404 \end{bmatrix}, \mathbf{p}_2 = \begin{bmatrix} 0.6763 \\ 0.3489 \\ 0.7039 \end{bmatrix}, \mathbf{p}_3 = \begin{bmatrix} 0.5431 \\ 0.4740 \\ 0.5756 \end{bmatrix},$$

$$\mathbf{p}_4 = \begin{bmatrix} 0.1425 \\ 0.8826 \\ 0.2801 \end{bmatrix}, \mathbf{p}_5 = \begin{bmatrix} 0.3034 \\ 0.6640 \\ 0.5243 \end{bmatrix}, \mathbf{p}_6 = \begin{bmatrix} 0.5806 \\ 0.4154 \\ 0.6729 \end{bmatrix}, \mathbf{p}_7 = \begin{bmatrix} 0.4948 \\ 0.5350 \\ 0.5946 \end{bmatrix},$$

$$\mathbf{p}_8 = \begin{bmatrix} 0.6670 \\ 0.3386 \\ 0.7375 \end{bmatrix}, \mathbf{p}_9 = \begin{bmatrix} 0.3495 \\ 0.5876 \\ 0.4980 \end{bmatrix}, \mathbf{p}_{10} = \begin{bmatrix} 0.5816 \\ 0.3480 \\ 0.7132 \end{bmatrix}.$$

(a) Find the mean $\mathbf{m}$ of the training set, and use it to find the caricature vectors $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_{10}$.

(b) Find the 3×10 caricature matrix Q, the 3×3 correlation matrix $C = QQ^T$, and the 10×10 matrix $A = Q^T Q$.

(c) Find the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ and corresponding normalized eigenvectors $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ of the correlation matrix C.

(d) Find the 10 eigenvalues of the matrix A and verify that the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ of the matrix C in part (c) are the nonzero eigenvalues of A and that corresponding eigenvectors are $Q\mathbf{u}_1, Q\mathbf{u}_2, Q\mathbf{u}_3$ with norms $\sqrt{\lambda_1}, \sqrt{\lambda_2}, \sqrt{\lambda_3}$. [*Note:* In this example it is easier to work with the 3×3 matrix C rather than the 10×10 matrix A. However, that is a result of the small size of our training set. In most facial-recognition applications A will be very much smaller than C.]

(e) Compute the three eigenfaces $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ determined by the training set.

(f) Compute the 3 features of each of the 10 faces in the training set.

(g) Let f_{ij} denote the jth feature of the ith face vector in the training set, so that the ith caricature vector is $\mathbf{q}_i = f_{i1}\mathbf{e}_1 + f_{i2}\mathbf{e}_2 + f_{i3}\mathbf{e}_3$. Suppose that $\mathbf{q}_i$ is approximated by $f_{i1}\mathbf{e}_1$ (that is, each caricature vector is approximated by the feature corresponding to the largest eigenvalue of C). Find the Euclidean distance between each caricature vector and its approximation. [*Note:* In the figure, the blue line passes through the average coordinates of the 1000 pictured points and lies in the direction of the eigenface $\mathbf{e}_1$. Thus, the distances computed in this part are the distances of the red training-set points to the line.]

Index

Bibliography

[1] Anton, H., Rorres, C., & Kaul, A. *Elementary Linear Algebra: Applications Version* Hoboken, NJ, USA: Wiley, 2019. Print.