Notes for Linear Algebra with Application

CHAPTER 1 Matrices and System of Equations

Index:

Kirchhoff's Laws; *Ohm's law*; adjacency matrix; mathematical induction;

1.1. Systems of Linear Equations

A ***linear*** **equation** in n unknowns is an equation of the form

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

where $a_1, a_2, \cdots, a_n$ and b are real numbers and $x_1, x_2, \cdots, x_n$ are variables. A *linear* system of m equations in n unknowns is then a system of the form

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

where the a_{ij}'s and the b_i's are all real numbers. We will refer to above systems of the form as $m \times n$ linear systems. For example, a 2×2 system,

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

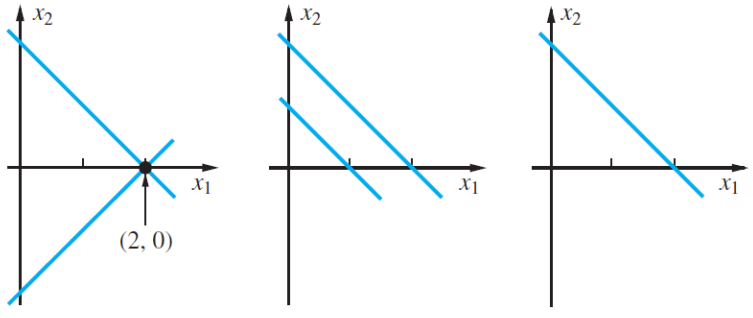
By a **solution** of an $m \times n$ system, we mean an ordered n-tuple of numbers $(x_1, x_2, \cdots, x_n)$ that satisfies all the equations of the system.

If a linear system has no solution, we say that the system is ***inconsistent***. If the system has at least one solution, we say that it is ***consistent***.

The set of all solutions of a linear system is called the *solution set* of the system. If a system is inconsistent, its solution set is empty. A consistent system will have a nonempty solution set. To solve a consistent system, we must find its solution set.

2×2 Systems

Each equation can be represented graphically as a line in the plane. The ordered pair (x_1, x_2) will be a solution of the system if and only if it lies on both lines



(i) (ii) (iii)

The situation is the same for $m \times n$ systems. An $m \times n$ system may or may not be consistent. If it is consistent, it must have either exactly one solution or infinitely many solutions. These are the **only** possibilities. We will see why this is so in Section 1.2 when we study the row echelon form. Of more immediate concern is the problem of finding all solutions of a given system. To tackle this problem, we introduce the notion of *equivalent systems*.

Equivalent Systems

> Two systems of equations involving the same variables are said to be **equivalent** if they have the same solution set.

Clearly, if we interchange the order in which two equations of a system are written, this will have no effect on the solution set.
If one equation of a system is multiplied through by a nonzero real number, this will have no effect on the solution set, and the new system will be equivalent to the original system.
If a multiple of one equation is added to another equation, the new system will be equivalent to the original system.
To summarize, there are three operations that can be used on a system to **obtain** an equivalent system:

I. The order in which any two equations are written may be **interchanged**
II. Both sides of an equation may be **multiplied** by the same nonzero real number
III. A **multiple** of one equation may be **added** to (or **subtracted** from) another

Given a system of equations, we may use these operations to obtain an equivalent system that is easier to solve.

$n \times n$ Systems

We will show that if an $n \times n$ system has exactly one solution, then operations **I** and **III** can be used to obtain an equivalent "strictly triangular system."

> A system is said to be in **strict triangular form** if, in the kth equation, the coefficients of the first $k-1$ variables are all zero and the coefficient of x_k is nonzero $(k = 1, \cdots, n)$.

The system

$$\begin{aligned} 3x_1 + 2x_2 + x_3 &= 1 \\ x_2 - x_3 &= 2 \\ 2x_3 &= 4 \end{aligned}$$

is in strict triangular form. Because of the strict triangular form, the system is easy to solve.
Any $n \times n$ strictly triangular system can be solved in the same manner as the last example. First, the nth equation is solved for the value of x_n. This value is used in the $(n-1)$st equation to solve for x_{n-1}. The values x_n and x_{n-1} are used in the $(n-2)$nd equation to solve for x_{n-2}, and so on. We will refer to this method of solving a strictly triangular system as ***back substitution***.
In general, given a system of n linear equations in n unknowns, we will use operations **I** and **III** to try to obtain an equivalent system that is strictly triangular.
Let us look back at the system of equations in the last example. We can associate with that system a 3×3 array of numbers whose entries are the coefficients of the x_i's: (Page 7)

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

We will refer to this array as the *coefficient matrix* of the system. The term ***matrix*** means simply a rectangular array of numbers. A matrix having m rows and n columns is said to be $m \times n$. A matrix is said to be *square* if it has the same number of rows and columns, that is, if $m = n$.
If we attach to the coefficient matrix an additional column whose entries are the numbers on the right-hand side of the system, we obtain the new matrix

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

We will refer to this new matrix as the ***augmented*** *matrix*. In general, when an $m \times n$ matrix B is attached to an $m \times n$ matrix A in this way, the augmented matrix is denoted by $(A|B)$. Thus, if

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1r} \\ b_{21} & b_{22} & \cdots & b_{2r} \\ \vdots & & & \\ b_{m1} & b_{m2} & \cdots & b_{mr} \end{bmatrix}$$

then

$$(A|B) = \left[\begin{array}{ccc|ccc} a_{11} & \cdots & a_{1n} & b_{11} & \cdots & b_{1r} \\ \vdots & & & \vdots & & \\ a_{m1} & \cdots & a_{mn} & b_{m1} & \cdots & b_{mr} \end{array}\right]$$

With each system of equations we may associate an augmented matrix of the form

$$\left[\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & & & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{array}\right]$$

The system can be solved by performing operations on the augmented matrix. The x_i's are placeholders that can be omitted until the end of the computation. Corresponding to the three operations used to obtain equivalent systems, the following row operations may be applied to the augmented matrix:

Elementary Row Operations

I. Interchange two rows.
II. Multiply a row by a nonzero real number.
III. Replace a row by its **sum** with a **multiple** of another row.

Returning to the example, we find that the first row is used to eliminate the elements in the first column of the remaining rows. We refer to the first row as the *pivotal row*. The first nonzero entry in the pivotal row is called the *pivot*. (Page 8)

$$\left.\begin{array}{r}(\text{pivot } a_{11} = 1) \\ \text{entries to be eliminated} \\ a_{21} = 3 \text{ and } a_{31} = 2\end{array}\right\} \rightarrow \left[\begin{array}{ccc|c} \mathbf{1} & \mathbf{2} & \mathbf{1} & \mathbf{3} \\ 3 & -1 & -3 & -1 \\ 2 & 3 & 1 & 4 \end{array}\right] \leftarrow \text{pivotal row}$$

At this step we choose the second row as our new pivotal row and apply row operation **III** to eliminate the last element in the second column (Page 8).

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ \mathbf{0} & \mathbf{-7} & \mathbf{-6} & \mathbf{-10} \\ 0 & -1 & -1 & -2 \end{array}\right] \leftarrow \text{pivotal row}$$

Since it is not possible to eliminate any entries by using 0 as a pivot element, we will use row operation **I** to interchange the first two rows of the augmented matrix (Page 8).

$$\left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & -7 & -6 & -10 \\ 0 & 0 & -\frac{1}{7} & -\frac{4}{7} \end{array}\right]$$

In general, if an $n \times n$ linear system can be reduced to strictly triangular form, then it will have a unique solution that can be obtained by performing **back substitution** on the triangular system. We can think of the reduction process as an algorithm involving $n - 1$ steps (Page 9).

If the elimination process can be carried out as described, we will arrive at an equivalent strictly triangular system after $n - 1$ steps. However, the procedure will break down if, at any step, all possible choices for a pivot element are equal to 0. When this happens, the alternative is to reduce the system to certain special echelon, or staircase-shaped, forms. These echelon forms will be studied in the next section. They will also be used for $m \times n$ systems, where $m \neq n$

Step 1

$$\begin{pmatrix} x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \\ x & x & x & x & x \end{pmatrix} \rightarrow \begin{pmatrix} x & x & x & x & x \\ 0 & x & x & x & x \\ 0 & x & x & x & x \\ 0 & x & x & x & x \end{pmatrix}$$

Step 2

$$\begin{pmatrix} x & x & x & x & x \\ 0 & x & x & x & x \\ 0 & x & x & x & x \\ 0 & x & x & x & x \end{pmatrix} \rightarrow \begin{pmatrix} x & x & x & x & x \\ 0 & x & x & x & x \\ 0 & 0 & x & x & x \\ 0 & 0 & x & x & x \end{pmatrix}$$

Step 3

$$\begin{pmatrix} x & x & x & x & x \\ 0 & x & x & x & x \\ 0 & 0 & x & x & x \\ 0 & 0 & x & x & x \end{pmatrix} \rightarrow \begin{pmatrix} x & x & x & x & x \\ 0 & x & x & x & x \\ 0 & 0 & x & x & x \\ 0 & 0 & 0 & x & x \end{pmatrix}$$

1.2. Row Echelon Form

In Section 1.1 we learned a method for reducing an $n \times n$ linear system to strict triangular form. However, this method will **fail** if, at any stage of the reduction process, all the possible choices for a pivot element in a given column are 0.

EXAMPLE 1: Consider the system represented by the augmented matrix (Page 11)

$$\left(\begin{array}{rrrrr|r} \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ -1 & -1 & 0 & 0 & 1 & -1 \\ -2 & -2 & 0 & 0 & 3 & 1 \\ 0 & 0 & 1 & 1 & 3 & -1 \\ 1 & 1 & 2 & 2 & 4 & 1 \end{array}\right) \leftarrow \text{pivotal row}$$

$$\left(\begin{array}{rrrrr|r} 1 & 1 & 1 & 1 & 1 & 1 \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{1} & \mathbf{2} & \mathbf{0} \\ 0 & 0 & 2 & 2 & 5 & 3 \\ 0 & 0 & 1 & 1 & 3 & -1 \\ 0 & 0 & 1 & 1 & 3 & 0 \end{array}\right) \leftarrow \text{pivotal row}$$

$$\left(\begin{array}{rrrrr|r} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 2 & 0 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{3} \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right)$$

$$\left(\begin{array}{rrrrr|r} \textcircled{1} & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & \textcircled{1} & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & \textcircled{1} & 3 \\ 0 & 0 & 0 & 0 & 0 & -4 \\ 0 & 0 & 0 & 0 & 0 & -3 \end{array}\right)$$

The coefficient matrix that we end up with is not in strict triangular form; it is in staircase, or echelon, form. The horizontal and vertical line segments in the array for the coefficient matrix indicate the structure of the staircase form. Note that the vertical drop is 1 for each step, but the horizontal span for a step can be more than 1.

Since there are **no** 5-tuples that could satisfy the **last two rows**, the system is **inconsistent**.

Suppose now that we **change** the right-hand side of the system in the last example so as to obtain a

consistent system. For example, if we start with

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

then the reduction process will yield the echelon-form augmented matrix

$$\left[\begin{array}{rrrrr|r} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Thus the solution set will be the set of all 5-tuples **satisfying** the **first three equations**

$$x_1 + x_2 + x_3 + x_4 + x_5 = 1$$
$$x_3 + x_4 + 2x_5 = 0$$
$$x_5 = 3$$

The variables corresponding to the **first nonzero elements** in each row of the reduced matrix will be referred to as ***lead variables***. Thus $x_1, x_3,$ and x_5 are the lead variables. The **remaining** variables corresponding to the columns skipped in the reduction process will be referred to as ***free*** *variables*. Hence, x_2 and x_4 are the free variables. If we transfer the **free variables** over to the right-hand side in above equations, we obtain the system

$$x_1 + x_3 + x_5 = 1 - x_2 - x_4$$
$$x_3 + 2x_5 = -x_4$$
$$x_5 = 3$$

System above is strictly triangular in the unknowns $x_1, x_3,$ and x_5.

> A matrix is said to be **in row echelon form** if
> 1. The first nonzero entry in each nonzero row is 1
> 2. If row k does not consist entirely of zeros, the number of leading zero entries in row $k + 1$ is greater than the number of leading zero entries in row k
> 3. If there are rows whose entries are all zero, they are below the rows having nonzero entries

The following matrices are not in row echelon form:

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

> The process of using row operations **I, II**, and **III** to transform a linear system into one whose augmented matrix is in **row echelon form** is called **Gaussian elimination**.

If the row echelon form of the augmented matrix contains a row of the form

$$[0 \quad 0 \quad \cdots \quad 0 | \quad 1]$$

the system is **inconsistent**. **Otherwise**, the system will be consistent. If the system is consistent and the nonzero rows of the row echelon form of the matrix form a strictly triangular system, the system will have a **unique solution**.

Overdetermined Systems

A linear system is said to be *overdetermined* if there are **more equations** than unknowns. Overdetermined systems are *usually* (but **not** always) **inconsistent** (Page 14).

Underdetermined Systems

A system of m linear equations in n unknowns is said to be *underdetermined* if there are **fewer equations** than unknowns ($m < n$). Although it is possible for underdetermined systems to be inconsistent, they are usually **consistent** with **infinitely many solutions**. It is not possible for an underdetermined system to have a **unique** solution. The reason for this is that any row echelon form of the coefficient matrix will involve $r \leq m$ nonzero rows. Thus there will be r lead variables and $n - r$ free variables, where $n - r \geq n - m > 0$. If the system is consistent, we can assign the free variables arbitrary values and solve for the lead variables. Therefore, **a consistent underdetermined system will have infinitely many solutions**.

EXAMPLE 5: Solve the following underdetermined systems: (Page 15)

$$x_1 + x_2 + x_3 + x_4 + x_5 = 2$$

(b) $x_1 + x_2 + x_3 + 2x_4 + 2x_5 = 3$

$$x_1 + x_2 + x_3 + 2x_4 + 3x_5 = 2$$

Note: the transition procedure is shown by the arrow "→" (not equality "=")

$$\left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{-1} \end{array}\right) \rightarrow \left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 0 & 3 \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{2} \\ 0 & 0 & 0 & 0 & 1 & -1 \end{array}\right)$$

$$\rightarrow \left(\begin{array}{ccccc|c} 1 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{array}\right)$$

$$x_1 = 1 - x_2 - x_3$$
$$x_4 = 2$$
$$x_5 = -1$$

Thus, for any real numbers α and β, the 5-tuple

$$(1 - \alpha - \beta, \alpha, \beta, 2, -1)$$

is a solution of the system.

In the case where the row echelon form of a consistent system has free variables, **the standard procedure** is to **continue** the elimination process until **all the entries** above each leading 1 have been **eliminated**. The resulting reduced matrix is said to be in *reduced row echelon form*.

Reduced Row Echelon Form

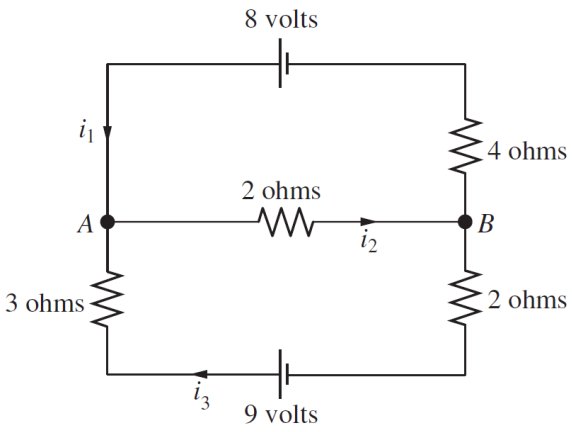
> A matrix is said to be in **reduced row echelon form** if
> 1. The matrix is in row echelon form.
> 2. The first nonzero entry in each row is the only nonzero entry in its column.

The following matrices are in reduced row echelon form:

$$\left(\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right), \quad \left(\begin{array}{cccc} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 1 \end{array}\right), \quad \left(\begin{array}{cccc} 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right), \quad \left(\begin{array}{cccc} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

The **process** of using elementary row operations to transform a matrix into **reduced row echelon form** is called ***Gauss–Jordan reduction***.

APPLICATION 2 Electrical Networks (Page 19)



Kirchhoff's Laws

1. At every node the sum of the incoming currents equals the sum of the outgoing currents
2. Around every closed loop, the algebraic sum of the voltage gains must equal the algebraic sum of the voltage drops

The voltage drops E for each resistor are given by *Ohm's law*:

$$E = iR$$

Homogeneous Systems

A system of linear equations is said to be *homogeneous* if the **constants** on the **right-hand side** are all zero. Homogeneous systems are **always consistent**. It is a trivial matter to find a solution; just set all the variables equal to zero. Thus, **if** an $m \times n$ homogeneous system has **a unique solution**, it must be the trivial solution $(0, 0, \cdots, 0)$.

Theorem 1.2.1 (Page 20)

An $m \times n$ homogeneous system of linear equations has a nontrivial solution if $n > m$.

APPLICATION 4 Economic Models for Exchange of Goods (Page 21)

1.3. Matrix Arithmetic

In this section, we introduce the standard notations used for matrices and vectors and define arithmetic operations (addition, subtraction, and multiplication) with matrices. We will also introduce two additional operations: *scalar multiplication* and *transposition*.

The entries of a matrix are called ***scalars***. They are usually either **real** or **complex numbers**.

Matrix Notation

We will use capital letters A, B, C, and so on. In general, a_{ij} will denote the entry of the matrix A that is in the ith row and the jth column. We will refer to this entry as the (i, j) entry of A. Thus, if A is an $m \times n$ matrix, then

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

We will sometimes shorten this to $A = (a_{ij})$. Similarly, a matrix B may be referred to as (b_{ij}), a matrix C as (c_{ij}), and so on.

Vectors

Matrices that have **only one row** or **one column** are of **special interest**, since they are used to represent **solutions** of linear systems. We will refer to an n-tuple of real numbers as a ***vector***. If an n-tuple is represented in terms of a $\mathbf{1} \times \boldsymbol{n}$ **matrix**, then we will refer to it as a ***row*** *vector*. Alternatively, if the n-tuple is represented by an $\boldsymbol{n} \times \mathbf{1}$ **matrix**, then we will refer to it as a ***column*** *vector*. For example, the solution of the linear system

$$\begin{aligned} x_1 + x_2 &= 3 \\ x_1 - x_2 &= 1 \end{aligned}$$

can be represented by the row vector $(2, 1)$ **or** the column vector $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

In working with matrix equations, it is generally more convenient to represent the solutions in terms of column vectors ($n \times 1$ matrix). The **set** of **all** $n \times 1$ **matrices** of **real** numbers is called ***Euclidean*** *n-space* and is usually denoted by R^n. Since we will be working almost exclusively with column vectors in the future, we will generally omit the word "column" and refer to the **elements** of R^n as **simply** *vectors*, rather than as column vectors. The standard notation for a column vector is a boldface lowercase letter, as in

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

For row vectors, there is no universal standard notation. In this book, we will represent both row and column vectors with boldface lowercase letters and to distinguish a row vector from a column vector we will place a horizontal arrow above the letter. For example,

$$\vec{\mathbf{x}} = (x_1, x_2, x_3, x_4) \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

are row and column vectors, respectively, with four entries each.

Given an $m \times n$ matrix A, it is often necessary to refer to a particular row or column. The standard notation for the jth column vector of A is $\mathbf{a}_j$. There is no universally accepted standard notation for the ith row vector of a matrix A. In this book, since we use horizontal arrows to indicate row vectors, we denote the ith row vector of A by $\vec{\mathbf{a}}_i$.

If A is an $m \times n$ matrix, then the row vectors of A are given by

$$\vec{\mathbf{a}}_i = (a_{i1}, a_{i2}, \cdots, a_{in}) \quad i = 1, \cdots, m$$

and the column vectors are given by

$$\mathbf{a}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix} \quad j = 1, \cdots, n$$

The matrix A can be represented in terms of either its column vectors or its row vectors:

$$A = (\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n) \quad \text{or} \quad A = \begin{bmatrix} \vec{\mathbf{a}}_1 \\ \vec{\mathbf{a}}_2 \\ \vdots \\ \vec{\mathbf{a}}_m \end{bmatrix}$$

Similarly, if B is an $n \times r$ matrix, then

$$B = (\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_r) = \begin{bmatrix} \vec{\mathbf{b}}_1 \\ \vec{\mathbf{b}}_2 \\ \vdots \\ \vec{\mathbf{b}}_n \end{bmatrix}$$

Equality

For two matrices to be equal, they must have the same **dimensions** and their corresponding entries must agree.

> Two $m \times n$ matrices A and B are said to be **equal** if $a_{ij} = b_{ij}$ for each i and j.

Scalar Multiplication

If A is a matrix and α is a scalar, then αA is the matrix formed by **multiplying** each of the **entries** of A by α.

> If A is an $m \times n$ matrix and α is a scalar, then αA is the $m \times n$ matrix whose (i, j) entry is αa_{ij}.

Matrix Addition

Two matrices with the same **dimensions** can be added by adding their corresponding entries.

If $A = (a_{ij})$ and $B = (b_{ij})$ are both $m \times n$ matrices, then the **sum** $A + B$ is the $m \times n$ matrix whose (i, j) entry is $a_{ij} + b_{ij}$ for each ordered pair (i, j).

If O represents the matrix, with the same dimensions as A, whose entries are all 0, then

$$A + O = O + A = A$$

We will refer to O as the *zero matrix*. It acts as an **additive identity** on the set of all $m \times n$ matrices. Furthermore, each $m \times n$ matrix A has an **additive inverse**. Indeed,

$$A + (-1)A = O = (-1)A + A$$

It is customary to denote the additive inverse by $-A$. Thus,

$$-A = (-1)A$$

Matrix Multiplication and Linear Systems

We have yet to define the most important operation: the multiplication of two matrices. Much of the **motivation** behind the definition comes from the applications to linear systems of equations. If we have a system of **one** linear equation in **one** unknown, it can be written in the form

$$ax = b$$

We **generally** think of a, x, and b as being scalars; **however**, they could **also** be treated as 1×1 matrices. Our goal now is to generalize equation above so that we can represent an $m \times n$ linear system by a single matrix equation of the form

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

where A is a $m \times n$ matrix, $\mathbf{x}$ is an unknown vector in R^n, and $\mathbf{b}$ is in R^m.

Case 1. One Equation in Several Unknowns

Let us begin by examining the case of one equation in several variables. Consider, for example, the equation

$$3x_1 + 2x_2 + 5x_3 = 4$$

If we set

$$A = [3 \quad 2 \quad 5] \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

and define the product $A\mathbf{x}$ by

$$A\mathbf{x} = [3 \quad 2 \quad 5] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 3x_1 + 2x_2 + 5x_3$$

then the equation $3x_1 + 2x_2 + 5x_3 = 4$ can be written as the matrix equation

$$A\mathbf{x} = 4$$

For **a** linear equation with n unknowns of the form

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

if we let

$$A = [a_1 \quad a_2 \quad \cdots \quad a_n] \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

and define the product $A\mathbf{x}$ by

$$A\mathbf{x} = a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

then the system can be written in the form $A\mathbf{x} = \mathbf{b}$

Note that the **result** of multiplying **a row vector** on the left by **a column vector** on the right is a scalar. Consequently, this type of multiplication is often referred to as a ***scalar product***.

Case 2. M Equations in N Unknowns

Consider now an $m \times n$ linear system

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

It is **desirable** to write the system as a matrix equation

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

where $A = (a_{ij})$ is known, $\mathbf{x}$ is a $n \times 1$ matrix of unknowns, and $\mathbf{b}$ is an $m \times 1$ matrix representing the right-hand side of the system. Thus, if we set

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

and define the product $A\mathbf{x}$ by

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

Given an $m \times n$ matrix A and a vector $\mathbf{x}$ in R^n, it is possible to compute a product $A\mathbf{x}$ by above formula (Page 32)

$$A\mathbf{x} = \begin{bmatrix} \vec{\mathbf{a}}_1\mathbf{x} \\ \vec{\mathbf{a}}_2\mathbf{x} \\ \vdots \\ \vec{\mathbf{a}}_m\mathbf{x} \end{bmatrix}$$

An **alternative way** to represent the linear system in this case as a matrix equation is to express the product $A\mathbf{x}$ as a **sum** of **column vectors**:

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

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

Thus, we have

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

> If $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n$ are vectors in R^m and $c_1, c_2, \cdots, c_n$ are scalars, then a sum of the form
> $$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_n\mathbf{a}_n$$
> is said to be a **linear combination** of the vectors $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n$.

The **product** $A\mathbf{x}$ is a linear combination of the **column vectors of** $\boldsymbol{A}$. Some books even use this linear combination representation as the definition of matrix vector multiplication

> If A is an $m \times n$ matrix and $\mathbf{x}$ is a vector in R^n, then
> $$A\mathbf{x} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$$

The matrix equation above about $A\mathbf{x}$ provides **a nice way** of characterizing whether a linear system of equations is consistent. Indeed, the following theorem is a direct consequence of that matrix equation

Theorem 1.3.1

Consistency Theorem for Linear Systems

*A linear system $A\mathbf{x} = \mathbf{b}$ is **consistent** if and only if $\mathbf{b}$ can be written as **a linear combination** of the column vectors of A.*

Matrix Multiplication

More generally, it is possible to multiply a matrix A times a matrix B if the number of columns of A equals the number of rows of B. **The first column** of the product is determined by the **first column of $\boldsymbol{B}$**; that is, the first column of AB is $A\mathbf{b}_1$, the second column of AB is $A\mathbf{b}_2$, and so on. Thus the product AB is the matrix whose columns are $A\mathbf{b}_1, A\mathbf{b}_2, \cdots, A\mathbf{b}_n$

$$AB = (A\mathbf{b}_1, A\mathbf{b}_2, \cdots, A\mathbf{b}_n)$$

The (i,j) entry of AB is the ith entry of the column vector $A\mathbf{b}_j$. It is determined by multiplying the ith row vector of A times the jth column vector of B

If $A = (a_{ij})$ is an $m \times n$ matrix and $B = (b_{ij})$ is an $n \times r$ matrix, then the product $AB = C = (c_{ij}) = (A\mathbf{b}_1, A\mathbf{b}_2, \cdots, A\mathbf{b}_r)$ is the $m \times r$ matrix whose entries are defined by

$$c_{ij} = \vec{\mathbf{a}}_i \mathbf{b}_j = \sum_{k=1}^{n} a_{ik} b_{kj}$$

Multiplication of matrices is not commutative (Page 36).

Notational Rules

Just as in ordinary algebra, if an expression involves both multiplication and addition and there are no parentheses to indicate the order of the operations, multiplications are carried out **before** additions. This is true for both scalar and matrix multiplications.

The Transpose of a Matrix

Given an $m \times n$ matrix A, it is often useful to form a new $n \times m$ matrix whose columns are the rows of A

The **transpose** of an $m \times n$ matrix A is the $n \times m$ matrix B defined by

$$b_{ji} = a_{ij}$$

for $j = 1, \cdots, n$ and $i = 1, \cdots, m$. The transpose of A is denoted by A^T.

It follows from above that the jth row of A^T has the same entries, respectively, as the jth column of A, and the ith column of A^T has the same entries, respectively, as the ith row of A.

An $n \times n$ matrix A is said to be **symmetric** if $A^T = A$.

The following are some examples of symmetric matrices:

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

EXERCISES

16. A matrix A is said to be *skew symmetric* if $A^T = -A$. Its diagonal entries must all be 0.

1.4. Matrix Algebra

Warning: In general, $AB \neq BA$. Matrix multiplication is *not* commutative.

Algebraic Rules

Theorem 1.4.1

Each of the following statements is valid for any scalars α and β and for any matrices A, B, and C for which the indicated operations are defined

1. $A + B = B + A$
2. $(A + B) + C = A + (B + C)$
3. $(AB)C = A(BC)$
4. $A(B + C) = AB + AC$
5. $(A + B)C = AC + BC$
6. $(\alpha\beta)A = \alpha(\beta A)$
7. $\alpha(AB) = (\alpha A)B = A(\alpha B)$
8. $(\alpha + \beta)A = \alpha A + \beta A$
9. $\alpha(A + B) = \alpha A + \alpha B$

Proof of rule 4 and 3 (Page 47)

Notation

Since $(AB)C = A(BC)$, we may simply omit the parentheses and write ABC. The same is true for a product

of four or more matrices. In the case where an $n \times n$ matrix is multiplied by **itself** a number of times, it is **convenient** to use exponential notation. Thus, if k is a positive integer, then

$$A^k = \underbrace{AA \cdots A}_{k \text{ times}}$$

If

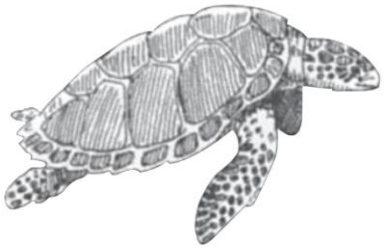
$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$
$$A^2 = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$
$$A^3 = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix}$$

and in general

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

APPLICATION 2

Ecology: Demographics of the Loggerhead Sea Turtle (Page 50)



The Identity Matrix

Just as the number 1 acts as an identity for the multiplication of real numbers, there is a special matrix I that acts as **an identity** for matrix multiplication; that is,

$$IA = AI = A$$

for any $n \times n$ matrix A. It is easy to verify that, if we define I to be an $n \times n$ matrix with 1's on the **main diagonal** and 0's **elsewhere**, then I satisfies equation above for any $n \times n$ matrix A. More formally, we have the following definition:

> The $n \times n$ **identity matrix** is the matrix $I = (\delta_{ij})$, where
> $$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

The column vectors of the $n \times n$ identity matrix I are the **standard vectors** used to define a coordinate system in Euclidean n-space. The **standard notation** for the jth column vector of I is $\mathbf{e}_j$, rather than the usual $\mathbf{i}_j$. Thus, the $n \times n$ identity matrix can be written

$$I = (\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n)$$

Matrix Inversion

A real number a is said to have a multiplicative inverse if there exists a number b such that $ab = 1$. Any nonzero number a has a multiplicative inverse $b = \frac{1}{a}$. We generalize the **concept** of **multiplicative inverses** to matrices with the following definition

> An $n \times n$ matrix A is said to be **nonsingular** or **invertible** if there exists a matrix B such that $AB = BA = I$. The matrix B is said to be a **multiplicative inverse** of A.

If B and C are both multiplicative inverses of A, then

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

Thus, a matrix can have **at most one** multiplicative inverse. We will refer to the multiplicative inverse of a nonsingular matrix A as simply the ***inverse*** of A and denote it by A^{-1}.

> An $n \times n$ matrix is said to be **singular** if it **does not have** a multiplicative inverse.

Note

Only square matrices have **multiplicative inverses**. One should not use the terms ***singular*** and ***nonsingular*** when referring to **nonsquare** matrices.

Often we will be working with products of nonsingular matrices. It turns out that any product of **nonsingular** matrices is **nonsingular**.

Theorem 1.4.2

*If A and B are **nonsingular** $n \times n$ matrices, then AB is **also nonsingular** and $(AB)^{-1} = B^{-1}A^{-1}$.*

Proof:

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

If $A_1, \cdots, A_k$ are **all nonsingular** $n \times n$ matrices, then the **product** $A_1A_2 \cdots A_k$ is **nonsingular** and

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

Algebraic Rules for Transposes

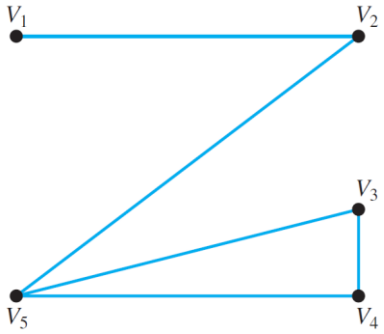
Algebraic Rules for Transposes
1. $(A^T)^T = A$
2. $(\alpha A)^T = \alpha A^T$
3. $(A + B)^T = A^T + B^T$
4. $(AB)^T = B^T A^T$

Symmetric Matrices and Networks

Recall that a matrix A is symmetric if $A^T = A$. One type of application that leads to symmetric matrices is problems involving networks. These problems are often solved using the techniques of an area of mathematics called *graph theory*.

APPLICATION 3

Networks and Graphs (Page 56)



An alternative is to use a matrix representation for the network. If the graph contains a total of n vertices, we can define an $n \times n$ matrix A by

$$a_{ij} = \begin{cases} 1 & \text{if } \{V_i, V_j\} \text{ is an edge of the graph} \\ 0 & \text{if there is no edge joining } V_i \text{ and } V_j \end{cases}$$

The matrix A is called the *adjacency matrix* of the graph. The adjacency matrix for the graph above is given by

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

Note that the matrix A is symmetric.

Theorem 1.4.3 (Page 57)

If A is an $n \times n$ adjacency matrix of a graph and $a_{ij}^{(k)}$ represents the (i, j) entry of A^k, then $a_{ij}^{(k)}$ is equal to the number of walks of length k from V_i to V_j.

EXERCISES

15. Let A be a nonsingular matrix. Show that A^{-1} is also nonsingular and $(A^{-1})^{-1} = A$.

Solution:

Consider A a nonsingular matrix. Then we have that there exist A^{-1} such that $AA^{-1} = A^{-1}A = I$.

Since $A^{-1}A = AA^{-1} = I$ it follows from the uniqueness of the inverse that A^{-1} is nonsingular with $(A^{-1})^{-1} = A$.

16. Prove that if A is nonsingular then A^T is nonsingular and

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

Hint: $(AB)^T = B^T A^T$.

Solution:

Consider A a nonsingular matrix. Then we have that there exist A^{-1} such that $AA^{-1} = A^{-1}A = I$.

Since $(AB)^T = B^T A^T$, we have that

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

We also have that $I = I^T$.

Hence we have that $A^T (A^{-1})^T = (A^{-1})^T A^T = I$ and by the uniqueness of the inverse we have that $(A^T)^{-1} = (A^{-1})^T$ and therefore A^T is nonsingular.

1.5. Elementary Matrices

In this section, we view the **process** of **solving a linear system** in terms of matrix multiplications **rather than** row operations. Given a linear system $A\mathbf{x} = \mathbf{b}$, we can multiply both sides by a sequence of special matrices to **obtain** an equivalent system in row echelon form. The special matrices we will use are called ***elementary matrices***.

Equivalent Systems

Given an $m \times n$ linear system $A\mathbf{x} = \mathbf{b}$, we can obtain an equivalent system by multiplying both sides of the equation by a **nonsingular** $m \times m$ matrix M:

$$A\mathbf{x} = \mathbf{b}$$
$$MA\mathbf{x} = M\mathbf{b}$$

However, if $\hat{\mathbf{x}}$ is a solution of second equation above, then

$$M^{-1}(MA\hat{\mathbf{x}}) = M^{-1}(M\mathbf{b})$$
$$A\hat{\mathbf{x}} = \mathbf{b}$$

and it follows that the **two systems** are equivalent.

We will show next that any of the three elementary row operations can be accomplished by **multiplying** A **on the left** by a **nonsingular matrix**.

Elementary Matrices

If we start with the identity matrix I and then perform exactly one elementary row operation, the resulting matrix is called an *elementary* matrix.

There are **three types** of elementary matrices corresponding to the **three types** of elementary row operations.

Type I: An elementary matrix of type I is a matrix obtained by interchanging **two rows** of I.

EXAMPLE 1 The matrix

$$E_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is an elementary matrix of type I since it was obtained by interchanging the first two rows of I. If A is a 3×3 matrix, then

$$E_1 A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$AE_1 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{12} & a_{11} & a_{13} \\ a_{22} & a_{21} & a_{23} \\ a_{32} & a_{31} & a_{33} \end{bmatrix}$$

Multiplying A on the left by E_1 interchanges the first and second rows of A. Right multiplication of A by E_1

is equivalent to the elementary column operation of interchanging the first and second columns.

Type II: An elementary matrix of type II is a matrix obtained by **multiplying** a **row** of I by a nonzero constant.

EXAMPLE 2

$$E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

is an elementary matrix of type II. If A is a 3×3 matrix, then

$$E_2A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 3a_{31} & 3a_{32} & 3a_{33} \end{bmatrix}$$

$$AE_2 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & 3a_{13} \\ a_{21} & a_{22} & 3a_{23} \\ a_{31} & a_{32} & 3a_{33} \end{bmatrix}$$

Multiplication on the left by E_2 performs the elementary row operation of multiplying the third row by 3, while multiplication on the right by E_2 performs the elementary column operation of multiplying the third column by 3.

Type III: An elementary matrix of type III is a matrix obtained from I by **adding** a **multiple** of one **row** to another **row**.

EXAMPLE 3

$$E_3 = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is an elementary matrix of type III. If A is a 3×3 matrix, then

$$E_3A = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11} + 3a_{31} & a_{12} + 3a_{32} & a_{13} + 3a_{33} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$AE_3 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & 3a_{11} + a_{13} \\ a_{21} & a_{22} & 3a_{21} + a_{23} \\ a_{31} & a_{32} & 3a_{31} + a_{33} \end{bmatrix}$$

Multiplication on the left by E_3 adds 3 times the third row to the first row. Multiplication on the right adds 3 times the first column to the third column.

In general, suppose that E is an $n \times n$ elementary matrix. We can think of E as being obtained from I by either a row operation or a column operation. If A is an $n \times r$ matrix, *premultiplying A by E has the effect of performing that same row operation on A*. If B is an $m \times n$ matrix, *postmultiplying B by E is equivalent to performing that same column operation on B*.

Theorem 1.5.1 (Page 63)

*If E is an **elementary matrix**, then E is **nonsingular** and E^{-1} is an elementary matrix of the same type.*

Definition

A matrix B is **row equivalent** to a matrix A if there **exists** a finite sequence $E_1, E_2, \cdots, E_k$ of **elementary matrices** such that

$$B = E_kE_{k-1}\cdots E_1A$$

In other words, B is row equivalent to A if B can be obtained from A by a finite number of row operations. In particular, if two augmented matrices $(A|\mathbf{b})$ and $(B|\mathbf{c})$ are row equivalent, then $A\mathbf{x} = \mathbf{b}$ and $B\mathbf{x} = \mathbf{c}$ are equivalent systems.

The following properties of row equivalent matrices are easily established

1. If A is row equivalent to B, then B is row equivalent to A
2. If A is row equivalent to B, and B is row equivalent to C, then A is row equivalent to C

Theorem 1.5.2

Equivalent Conditions for Nonsingularity

*Let A be an $n \times n$ matrix. The following are **equivalent**:*

(a) *A is nonsingular*
(b) $A\mathbf{x} = \mathbf{0}$ *has only the trivial solution* $\mathbf{0}$
(c) *A is row equivalent to I*

Corollary 1.5.3 (Page 65)

The system $A\mathbf{x} = \mathbf{b}$ *of* n *linear equations in* n *unknowns has a unique solution* ($\hat{\mathbf{x}} = A^{-1}\mathbf{b}$) *if and only if A is nonsingular*.

If A is nonsingular then A is row equivalent to I and hence there exist elementary matrices $E_1, \cdots, E_k$ such that

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

Multiplying both sides of this equation **on the right** by A^{-1}, we obtain

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

Thus the **same series** of elementary row operations that transforms a nonsingular matrix A into I **will** transform I into A^{-1}. This gives us a method for computing A^{-1}. If we **augment** A by I and perform the elementary row operations that transform A into I on the augmented matrix, **then** I will be transformed into A^{-1}. That is, the reduced row echelon form of the augmented matrix $(A|I)$ **will** be $(I|A^{-1})$ (Page 65).

Diagonal and Triangular Matrices

An $n \times n$ matrix A is said to be *upper triangular* if $a_{ij} = 0$ for $i > j$ and *lower triangular* if $a_{ij} = 0$ for $i < j$. Also, A is said to be *triangular* if it is either upper triangular or lower triangular. For example, the 3×3 matrices

$$\begin{bmatrix} 3 & 2 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 5 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 & 0 \\ 6 & 0 & 0 \\ 1 & 4 & 3 \end{bmatrix}$$

are both triangular.

A triangular matrix may have 0's on the diagonal. However, for a linear system $A\mathbf{x} = \mathbf{b}$ to be in strict triangular form, the coefficient matrix A must be upper triangular with nonzero diagonal entries.

An $n \times n$ matrix A is *diagonal* if $a_{ij} = 0$ whenever $i \neq j$. The matrices

$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

are all diagonal. A diagonal matrix is both upper triangular and lower triangular.

Triangular Factorization

If an $n \times n$ matrix A can be reduced to strict upper triangular form using only row operation III, then it is possible to represent the reduction process in terms of a matrix factorization (Page 67).

EXERCISES

21. In general, matrix multiplication is not commutative (i.e., $AB \neq BA$). However, in certain special cases the commutative property does hold.

(a) if D_1 and D_2 are $n \times n$ diagonal matrices, then $D_1 D_2 = D_2 D_1$.

(b) if A is an $n \times n$ matrix and

$$B = a_0 I + a_1 A + a_2 A^2 + \cdots + a_k A^k$$

where $a_0, a_1, \cdots, a_k$ are scalars, then $AB = BA$.

28. Given a vector $\mathbf{x} \in \mathrm{R}^{n+1}$, the $(n+1) \times (n+1)$ matrix V defined by

$$v_{ij} = \begin{cases} 1 & \text{if } j = 1 \\ x_i^{j-1} & \text{for } j = 2, \cdots, n+1 \end{cases}$$

is called the Vandermonde matrix.

(a) If

$$V\mathbf{c} = \mathbf{y}$$

and

$$p(x) = c_1 + c_2 x + \cdots + c_{n+1} x^n$$

then

$$p(x_i) = y_i, \quad i = 1,2, \cdots, n+1$$

(b) Suppose that $x_1, x_2, \cdots, x_{n+1}$ are all distinct. If $\mathbf{c}$ is a solution of $V\mathbf{x} = \mathbf{0}$ then the coefficients $c_1, c_2, \cdots, c_{n+1}$ must all be zero, and hence V must be nonsingular.

(Since all the roots $x_1, x_2, \cdots, x_{n+1}$ are all distinct, it follows that the polynomial $p(x_i)$ has at least $n+1$ distinct roots. However, this polynomial has the degree less or equal to n, which is strictly less than the number of its roots. This implies that $p(x_i)$ is zero polynomial, making all of its coefficients equal to zero.)

1.6. Partitioned Matrices

Often it is useful to think of a matrix as being composed of a number of submatrices. A matrix C can be partitioned into smaller matrices by drawing horizontal lines between the rows and vertical lines between the columns. The smaller matrices are often referred to as *blocks*.

$$AB = (A\mathbf{b}_1 \quad A\mathbf{b}_2 \quad \cdots \quad A\mathbf{b}_r)$$

Block Multiplication

Let A be an $m \times n$ matrix and B an $n \times r$ matrix. It is often useful to partition A and B and express the product in terms of the submatrices of A and B. Consider the following four cases

Case 1. If $B = [B_1 \quad B_2]$, where B_1 is an $n \times t$ matrix and B_2 is an $n \times (r-t)$ matrix, then

$$A[B_1 \quad B_2] = [AB_1 \quad AB_2]$$

Case 2. If $A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$, where A_1 is a $k \times n$ matrix and A_2 is an $(m-k) \times n$ matrix, then

$$\begin{bmatrix} A_1 \\ A_2 \end{bmatrix} B = \begin{bmatrix} A_1 B \\ A_2 B \end{bmatrix}$$

Case 3. Let $A = [A_1 \quad A_2]$ and $B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$, where A_1 is an $m \times s$ matrix, A_2 is an $m \times (n-s)$ matrix, B_1 is an $s \times r$ matrix, and B_2 is an $(n-s) \times r$ matrix. If $C = AB$, then

$$[A_1 \quad A_2] \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = A_1 B_1 + A_2 B_2$$

Case 4. Let A and B both be partitioned as follows:

$$A = \left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right) \begin{matrix} k \\ m-k \end{matrix}, \qquad B = \left(\begin{array}{c|c} B_{11} & B_{12} \\ \hline B_{21} & B_{22} \end{array}\right) \begin{matrix} s \\ n-s \end{matrix}$$
$$\begin{matrix} s & n-s \end{matrix} \qquad \begin{matrix} t & r-t \end{matrix}$$

Therefore,

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}$$

In general, if the blocks have the proper dimensions, the block multiplication can be carried out in the same manner as ordinary matrix multiplication (Page 74).

Outer Product Expansions

Given two vectors $\mathbf{x}$ and $\mathbf{y}$ in R^n, it is **possible** to perform a matrix multiplication of the vectors **if** we **transpose one** of the vectors first. The matrix product $\mathbf{x}^T\mathbf{y}$ is the product of a row vector (a $1 \times n$ matrix) and a column vector (an $n \times 1$ matrix). The result will be a 1×1 matrix, or simply a scalar:

$$\mathbf{x}^T\mathbf{y} = [x_1 \quad x_2 \quad \cdots \quad x_n] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = x_1y_1 + x_2y_2 + \cdots + x_ny_n$$

This type of product is referred to as a ***scalar product*** or an ***inner product***.

It is **also useful** to multiply a column vector times a row vector. The matrix product $\mathbf{x}\mathbf{y}^T$ is the product of an $n \times 1$ matrix times a $1 \times n$ matrix. The result is a full $n \times n$ matrix

$$\mathbf{x}\mathbf{y}^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} [y_1 \quad y_2 \quad \cdots \quad y_n] = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & \cdots & x_2y_n \\ \vdots & & & \\ x_ny_1 & x_ny_2 & \cdots & x_ny_n \end{bmatrix}$$

The product $\mathbf{x}\mathbf{y}^T$ is referred to as the ***outer product*** of $\mathbf{x}$ and $\mathbf{y}$.

We are now ready to generalize the idea of an outer product from vectors to matrices. Suppose that we start with an $m \times n$ matrix X and a $k \times n$ matrix Y (Page 77)

$$XY^T = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_n] \begin{bmatrix} \mathbf{y}_1^T \\ \mathbf{y}_2^T \\ \vdots \\ \mathbf{y}_n^T \end{bmatrix} = \mathbf{x}_1\mathbf{y}_1^T + \mathbf{x}_2\mathbf{y}_2^T + \cdots + \mathbf{x}_n\mathbf{y}_n^T$$

CHAPTER 2 Determinants

With each square matrix, it is possible to associate a **real** number called the **determinant** of the matrix. The value of this number will **tell** us **whether the matrix is singular**.

2.1. The Determinant of a Matrix

With each $n \times n$ matrix A it is possible to associate a scalar, $\det(A)$, whose value will tell us whether the matrix is nonsingular. Before proceeding to the general definition, let us consider the following cases

Case 1. 1×1 Matrices If $A = (a)$ is a 1×1 matrix, then A will have a **multiplicative inverse** if and only if $a \neq 0$. Thus, if we define

$$\det(A) = a$$

then A will be nonsingular if and only if $\det(A) \neq 0$.

Case 2. 2×2 Matrices Let

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

By Theorem 1.5.2, A will be nonsingular **if and only if** it is row equivalent to I. Then, **if** $a_{11} \neq 0$, we can test whether A is row equivalent to I by performing the following operations:

1. Multiply the second row of A by a_{11}

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{11}a_{21} & a_{11}a_{22} \end{bmatrix}$$

2. Subtract a_{21} times the first row from the new second row

$$\begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{11}a_{22} - a_{21}a_{12} \end{bmatrix}$$

Since $a_{11} \neq 0$, the resulting matrix will be row equivalent to I if and only if

$$a_{11}a_{22} - a_{21}a_{12} \neq 0$$

If $a_{11} \neq 0$, we can switch the two rows of A. The resulting matrix

$$\begin{bmatrix} a_{21} & a_{22} \\ 0 & a_{12} \end{bmatrix}$$

will be row equivalent to I if and only if $a_{21}a_{12} \neq 0$. This requirement is equivalent to condition $a_{11}a_{22} - a_{21}a_{12} \neq 0$. Thus, if A is any 2×2 matrix and we define

$$\det(A) = a_{11}a_{22} - a_{12}a_{21}$$

then A is nonsingular **if and only if** $\det(A) \neq 0$.

Notation

We can refer to the determinant of a specific matrix by enclosing the array between vertical lines. For example, if

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

then

$$\begin{vmatrix} 3 & 4 \\ 2 & 1 \end{vmatrix}$$

represents the determinant of A.

Case 3. 3×3 Matrices We can test whether a 3×3 matrix is nonsingular by performing row operations to see if the matrix is row equivalent to the identity matrix I. To carry out the elimination in the first column of an arbitrary 3×3 matrix A, let us first assume that $a_{11} \neq 0$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \rightarrow \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & \dfrac{a_{11}a_{22} - a_{21}a_{12}}{a_{11}} & \dfrac{a_{11}a_{23} - a_{21}a_{13}}{a_{11}} \\ 0 & \dfrac{a_{11}a_{32} - a_{31}a_{12}}{a_{11}} & \dfrac{a_{11}a_{33} - a_{31}a_{13}}{a_{11}} \end{bmatrix}$$

The matrix on the right will be row equivalent to I if and only if

$$a_{11} \begin{vmatrix} \dfrac{a_{11}a_{22} - a_{21}a_{12}}{a_{11}} & \dfrac{a_{11}a_{23} - a_{21}a_{13}}{a_{11}} \\ \dfrac{a_{11}a_{32} - a_{31}a_{12}}{a_{11}} & \dfrac{a_{11}a_{33} - a_{31}a_{13}}{a_{11}} \end{vmatrix} \neq 0$$

Thus, if we define

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

then, for the case $a_{11} \neq 0$, the matrix will be nonsingular if and only if $\det(A) \neq 0$.

What if $a_{11} = 0$? Consider the following possibilities:

1. $a_{11} = 0, a_{21} \neq 0$
2. $a_{11} = a_{21} = 0, a_{31} \neq 0$
3. $a_{11} = a_{21} = a_{31} = 0$

In case 1, it is not difficult to show that A is row equivalent to I if and only if

$$-a_{12}a_{21}a_{33} + a_{12}a_{31}a_{23} + a_{13}a_{21}a_{32} - a_{13}a_{31}a_{22} \neq 0$$

But this condition is the same as previous condition ($a_{11} \neq 0$) with $a_{11} = 0$.

In case 2, it follows that

$$A = \begin{bmatrix} 0 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

is row equivalent to I if and only if

$$a_{31}(a_{12}a_{23} - a_{22}a_{13}) \neq 0$$

Again, this is a special case of previous condition ($a_{11} \neq 0$) with $a_{11} = a_{21} = 0$.

Clearly, in case 3 the matrix A cannot be row equivalent to I and hence must be singular. The result will be

$$\det(A) = 0$$

We would now like to define the determinant of an $n \times n$ matrix.

For a 2×2 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

Define two 1×1 matrices

$$M_{11} = (a_{22}) \quad \text{and} \quad M_{12} = (a_{21})$$

$$\det(A) = a_{11}a_{22} - a_{12}a_{21} = a_{11}\det(M_{11}) - a_{12}\det(M_{12})$$

For a 3×3 matrix A,

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

$$\det(A) = a_{11}\det(M_{11}) - a_{12}\det(M_{12}) + a_{13}\det(M_{13})$$

where

$$M_{11} = \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}, \quad M_{12} = \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix}, \quad M_{13} = \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$$

For the case $n > 3$, we introduce the following definition

> Let $A = (a_{ij})$ be an $n \times n$ matrix and let M_{ij} denote the $(n-1) \times (n-1)$ matrix **obtained** from A by deleting the **row** and **column** containing a_{ij}. The determinant of M_{ij} is called the **minor** of a_{ij}. We define the **cofactor** A_{ij} of a_{ij} by
>
> $$A_{ij} = (-1)^{i+j}\det(M_{ij})$$

In view of this definition, for a 2×2 matrix A, we may rewrite $\det(A)$ in the form

$$\det(A) = a_{11}A_{11} + a_{12}A_{12} \quad (n = 2)$$

which is called the *cofactor expansion* of $\det(A)$ along the first row of A. Actually, there is no reason that we must expand along a row of the matrix (Page 90).

For a 3×3 matrix A, we have

$$\det(A) = a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13}$$

The **determinant** of an $n \times n$ matrix A, denoted $\det(A)$, is a **scalar** associated with the matrix A that is defined inductively as

$$\det(A) = \begin{cases} a_{11} & \text{if } n = 1 \\ a_{11}A_{11} + a_{12}A_{12} + \cdots + a_{1n}A_{1n} & \text{if } n > 1 \end{cases}$$

where

$$A_{1j} = (-1)^{1+j} \det(M_{1j}) \quad j = 1, \cdots, n$$

are the cofactors associated with the entries in the first row of A.

Theorem 2.1.1

If A is an $n \times n$ matrix with $n \geq 2$, then $\det(A)$ can be expressed as a cofactor expansion using **any row** or **column** of A

$$\begin{aligned} \det(A) &= a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in} \\ &= a_{1j}A_{1j} + a_{2j}A_{2j} + \cdots + a_{nj}A_{nj} \end{aligned}$$

for $i = 1, \cdots, n$ and $j = 1, \cdots, n$.

For $n \leq 3$, we have seen that an $n \times n$ matrix A is nonsingular if and only if $\det(A) \neq 0$. In the next section we will show that this result holds for all values of n.

Theorem 2.1.2

If A is an $n \times n$ matrix, then $\det(A^T) = \det(A)$.

Proof:

The proof is by induction on n. Clearly, the result holds if $n = 1$, since a 1×1 matrix is necessarily symmetric. Assume that the result holds for all $k \times k$ matrices and that A is a $(k+1) \times (k+1)$ matrix. Expanding $\det(A)$ along the first row of A, we get

$$\det(A) = a_{11} \det(M_{11}) - a_{12} \det(M_{12}) + - \cdots \pm a_{1,k+1} \det(M_{1,k+1})$$

Since the M_{ij}'s are all $k \times k$ matrices, it follows from the induction hypothesis that

$$\det(A) = a_{11} \det(M_{11}^T) - a_{12} \det(M_{12}^T) + - \cdots \pm a_{1,k+1} \det(M_{1,k+1}^T)$$

The right-hand side of the above equation is just the expansion by minors of $\det(A^T)$ using the first column of A^T. Therefore,

$$\det(A^T) = \det(A).$$

Theorem 2.1.3

If A is an $n \times n$ triangular matrix, then the determinant of A equals the product of the diagonal elements of A.

Theorem 2.1.4

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

i. If A has a row or column consisting entirely of zeros, then $\det(A) = 0$.
ii. If A has two identical rows or two identical columns, then $\det(A) = 0$.

EXERCISES

10. Use mathematical induction to prove that if A is an $(n+1) \times (n+1)$ matrix with two identical rows, then $\det(A) = 0$.

Solution:

Basis: $n = 1$

Let $A_{2\times 2} = [a_{ij}]$. If we assume that both rows of the matrix A are equal, we have the following:

$$\begin{aligned} \det(A) &= \begin{vmatrix} a_{11} & a_{12} \\ a_{11} & a_{12} \end{vmatrix} \\ &= 0 \end{aligned}$$

Induction step: $n \Longrightarrow n+1$

Let's assume that the determinant of every $n \times n$ matrix that has two identical rows is equal to zero and let $A_{(n+1)\times(n+1)} = [a_{ij}]$. Let's assume that $\vec{\boldsymbol{a}}_k = \vec{\boldsymbol{a}}_l$ for $k < l$. If we use cofactor expansion along the m-th row, where $k \neq m \neq l$, we have the following:

$$\begin{aligned} \det(A) &= a_{m1}A_{m1} + a_{m2}A_{m2} + \cdots + a_{m,n+1}A_{m,n+1} \\ &= a_{m1}0 + a_{m2}0 + \cdots + a_{m,n+1}0 \end{aligned}$$

$$= 0$$

Here, $A_{mi} = 0$, for $i = 1, \cdots, n+1$, because these are $n \times n$ matrices that have at least two identical rows (remainders of rows $\vec{a}_k$ and $\vec{a}_l$). Thus, it follows that $\det(A) = 0$ for every $A_{(n+1)\times(n+1)}$ matrix that has two identical rows. (same as column)

11. Let A and B be 2×2 matrices.
    (a) $\det(A + B) \neq \det(A) + \det(B)$
    (b) $\det(AB) = \det(A)\det(B)$
    (c) $\det(AB) = \det(BA)$

2.2. Properties of Determinants

Lemma 2.2.1

Let A be an $n \times n$ matrix. If A_{jk} denotes the cofactor of a_{jk} for $k = 1, \cdots, n$, then

$$a_{i1}A_{j1} + a_{i2}A_{j2} + \cdots + a_{in}A_{jn} = \begin{cases} \det(A) & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Let us now consider the effects of each of the **three row operations** on the value of the **determinant**.

Row Operation I

In general, if A is an $n \times n$ matrix and E_{ij} is the $n \times n$ elementary matrix formed by **interchanging** the ith and jth rows of I, then

$$\det(E_{ij}A) = -\det(A)$$

In particular,

$$\det(E_{ij}) = \det(E_{ij}I) = -\det(I) = -1$$

Thus, for any elementary matrix E of type I,

$$\det(EA) = -\det(A) = \det(E)\det(A)$$

Row Operation II

Let E denote the elementary matrix of type II formed from I by **multiplying** the ith **row** by the nonzero scalar α. If $\det(EA)$ is expanded by cofactors along the ith row, then

$$\begin{aligned} \det(EA) &= \alpha a_{i1}A_{i1} + \alpha a_{i2}A_{i2} + \cdots + \alpha a_{in}A_{in} \\ &= \alpha(a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in}) \\ &= \alpha \det(A) \end{aligned}$$

In particular,

$$\det(E) = \det(EI) = \alpha \det(I) = \alpha$$

and hence,

$$\det(EA) = \alpha \det(A) = \det(E)\det(A)$$

Row Operation III

Let E be the elementary matrix of type III formed from I by **adding** c **times** the ith **row** to the jth **row**. Since E is triangular and its diagonal elements are all 1, it follows that $\det(E) = 1$. We will show that

$$\det(EA) = \det(A) = \det(E)\det(A)$$

If $\det(EA)$ is expanded by cofactors along the jth row, it follows from Lemma 2.2.1 that

$$\begin{aligned} \det(EA) &= (a_{j1} + ca_{i1})A_{j1} + (a_{j2} + ca_{i2})A_{j2} + \cdots + (a_{jn} + ca_{in})A_{jn} \\ &= (a_{j1}A_{j1} + \cdots + a_{jn}A_{jn}) + c(a_{i1}A_{j1} + \cdots + a_{in}A_{jn}) \\ &= \det(A) \end{aligned}$$

SUMMARY

In summation, if E is an elementary matrix, then

$$\det(EA) = \det(E)\det(A)$$

where

$$\det(E) = \begin{cases} -1 & \text{if } E \text{ is of type I} \\ \alpha \neq 0 & \text{if } E \text{ is of type II} \\ 1 & \text{if } E \text{ is of type III} \end{cases}$$

Similar results hold for **column** operations. Indeed, if E is an elementary matrix, then E^T is also an elementary matrix and

$$\det(AE) = \det((AE)^T) = \det(E^T A^T)$$
$$= \det(E^T)\det(A^T) = \det(E)\det(A)$$

Thus, the effects that **row or column operations** have on the value of the determinant can be summarized as follows:

I. **Interchanging** two rows (or columns) of a matrix changes the sign of the determinant

II. **Multiplying** a **single** row or column of a matrix by **a scalar** has the effect of multiplying the value of the determinant by that scalar

III. **Adding** a **multiple** of one row (or column) to another does not change the value of the determinant (Note here we can use the equality "=")

Note

As a consequence of **III**, if one row (or column) of a matrix is a multiple of another, the determinant of the matrix must equal zero.

Main Results

Theorem 2.2.2

An $n \times n$ matrix A is **singular** if and only if

$$\det(A) = 0$$

From the proof of Theorem 2.2.2, we can obtain a method for computing $\det(A)$. We reduce A to row echelon form

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

If the last row of U consists entirely of zeros, A is singular and $\det(A) = 0$. Otherwise, A is nonsingular and

$$\det(A) = [\det(E_k)\det(E_{k-1}) \cdots \det(E_1)]^{-1}$$

Actually, if A is nonsingular, it is simpler to reduce A to triangular form. This can be done using only row operations I and III. Thus,

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

and hence,

$$\det(A) = \pm \det(T) = \pm t_{11} t_{22} \cdots t_{nn}$$

where the t_{ii}'s are the diagonal entries of T. The sign will be positive if row operation I has been used an even number of times and negative otherwise.

We now have two methods for evaluating the determinant of an $n \times n$ matrix A. If $n > 3$ and A has nonzero entries, elimination is the most efficient method

	Cofactors		Elimination	
n	Additions	Multiplications	Additions	Multiplications and Divisions
2	1	2	1	3
3	5	9	5	10
4	23	40	14	23
5	119	205	30	44
10	3,628,799	6,235,300	285	339

Theorem 2.2.3

If A and B are $n \times n$ matrix, then

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

SECTION 2.2 **EXERCISES**

5. Let A be an $n \times n$ matrix and α a scalar. Show that

$$\det(\alpha A) = \alpha^n \det(A)$$

6. Let A be a nonsingular matrix. Show that

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

12. Consider the 3×3 Vandermonde matrix

$$V = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{bmatrix}$$

(a) That $\det(V) = (x_2 - x_1)(x_3 - x_1)(x_3 - x_2)$. *Hint*: Make use of row operation III.

Remark:

For an $n \times n$ Vandermonde matrix V,

$$\det(V) = \det\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \vdots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{bmatrix} = \prod_{n \geq i > j \geq 1} (x_i - x_j) \quad (n > 2)$$

16. A matrix A is said to be *skew symmetric* if $A^T = -A$. If A is an $n \times n$ skew-symmetric matrix and n is odd, show that A must be singular.

18. Let A be a $k \times k$ matrix and let B be an $(n-k) \times (n-k)$ matrix. Let

$$E = \begin{bmatrix} I_k & O \\ O & B \end{bmatrix}, \quad F = \begin{bmatrix} A & O \\ O & I_{n-k} \end{bmatrix},$$

$$C = \begin{bmatrix} A & O \\ O & B \end{bmatrix}$$

where I_k and I_{n-k} are the $k \times k$ and $(n-k) \times (n-k)$ identity matrices.

(a) $\det(E) = \det(B)$

(b) $\det(F) = \det(A)$

(c) $\det(C) = \det(A)\det(B)$

2.3. Additional Topics and Applications

The Adjoint of a Matrix

Let A be $n \times n$ matrix. We define a new matrix called the *adjoint* of A by

$$\operatorname{adj} A = \begin{bmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & & & \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{bmatrix}$$

Thus, to form the adjoint, we must replace each term by its cofactor and then transpose the resulting matrix. By Lemma 2.2.1, and it follows that

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

If A is nonsingular, $\det(A)$ is a nonzero scalar, and we may write

$$A\left(\frac{1}{\det(A)} \operatorname{adj} A\right) = I$$

Thus,

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj} A \quad \text{when } \det(A) \neq 0$$

Cramer's Rule

Theorem 2.3.1 Cramer's Rule

Let A be a nonsingular $n \times n$ matrix, and let $\mathbf{b} \in \mathrm{R}^n$. Let A_i be the matrix obtained by replacing the ith column of A by $\mathbf{b}$. If $\mathbf{x}$ is the unique solution of $A\mathbf{x} = \mathbf{b}$, then

$$x_i = \frac{\det(A_i)}{\det(A)} \quad \text{for } i = 1,2,\cdots,n$$

since (Page 103)

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

APPLICATION 1 Coded Messages

A common way of sending a coded message is to assign an integer value to each letter of the alphabet and to send the message as a string of integers. For example, the message

SEND MONEY

might be coded as

$$5, 8, 10, 21, 7, 2, 10, 8, 3$$

Here the S is represented by a 5, the E by an 8, and so on. Unfortunately, this type of code is generally easy to break. In a longer message we might be able to guess which letter is represented by a number on the basis of the relative frequency of occurrence of that number. For example, if 8 is the most frequently occurring number in the coded message, then it is likely that it represents the letter E, the letter that occurs most frequently in the English language.

We can disguise the message further by using matrix multiplications. If A is a matrix whose entries are all integers and whose determinant is ± 1, then, since $A^{-1} = \pm \text{adj } A$, the entries of A^{-1} will be integers. We can use such a matrix to transform the message. The transformed message will be more difficult to decipher. (Page 104)

The Cross Product (Page 105)

Given two vectors $\mathbf{x}$ and $\mathbf{y}$ in R^3, one can define a third vector, the *cross product*, denoted $\mathbf{x} \times \mathbf{y}$, by

$$\mathbf{x} \times \mathbf{y} = \begin{Bmatrix} x_2y_3 - y_2x_3 \\ y_1x_3 - x_1y_3 \\ x_1y_2 - y_1x_2 \end{Bmatrix}$$

$$\mathbf{x}^T(\mathbf{x} \times \mathbf{y}) = \mathbf{y}^T(\mathbf{x} \times \mathbf{y}) = 0$$

In calculus books, it is standard to use row vectors

$$\mathbf{x} = (x_1, x_2, x_3) \quad \text{and} \quad \mathbf{y} = (y_1, y_2, y_3)$$

and to define the cross product to be the row vector

$$\mathbf{x} \times \mathbf{y} = (x_2y_3 - y_2x_3)\mathbf{i} - (x_1y_3 - y_1x_3)\mathbf{j} + (x_1y_2 - y_1x_2)\mathbf{k}$$

$$= \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}$$

APPLICATION 2 Newtonian Mechanics (Page 106)

$$\mathbf{x}^T\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\| \cos\theta$$

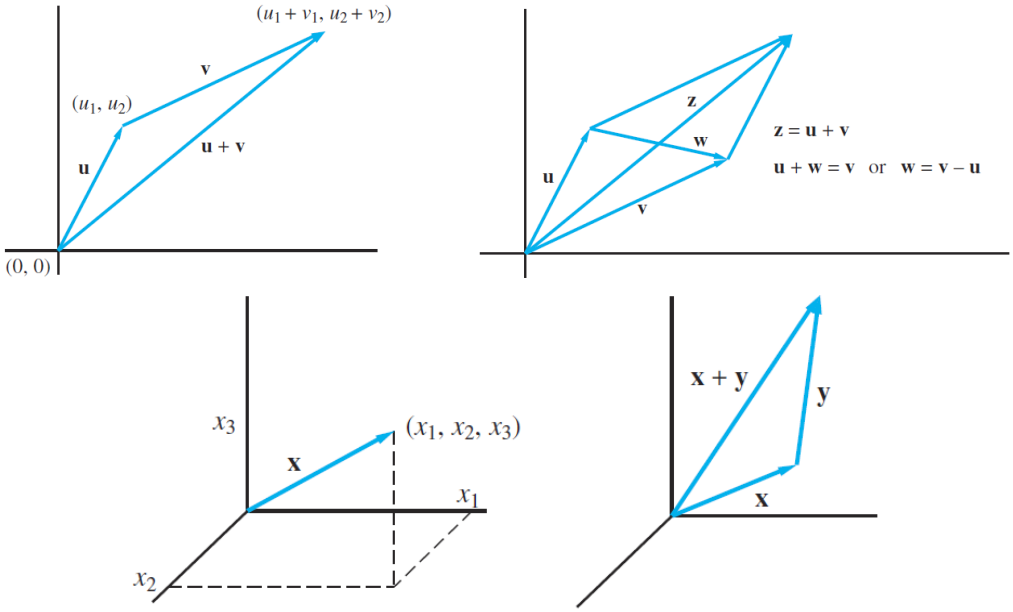
$$\|\mathbf{x} \times \mathbf{y}\| = \|\mathbf{x}\|\|\mathbf{y}\| \sin\theta$$

CHAPTER 3 Vector Spaces

3.1 Definition and Examples

Euclidean Vector Spaces

Perhaps the most elementary vector spaces are the Euclidean vector spaces R^n, $n = 1,2,\cdots$.



In general, scalar multiplication and addition in R^n are, respectively, defined by

$$\alpha\mathbf{x} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{bmatrix} \quad \text{and} \quad \mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$$

for any $\mathbf{x}, \mathbf{y} \in \mathrm{R}^n$ and any scalar α.

The Vector Space $\mathrm{R}^{m\times n}$

We can also view R^n as the set of all $n \times 1$ matrices with real entries. The addition and scalar multiplication of vectors in R^n is just the usual addition and scalar multiplication of matrices. More generally, let $\mathrm{R}^{m\times n}$ denote the set of all $m \times n$ matrices with real entries. If $A = (a_{ij})$ and $B = (b_{ij})$, then the sum $A + B$ is defined to be the $m \times n$ matrix $C = (c_{ij})$, where $c_{ij} = a_{ij} + b_{ij}$. Given a scalar α, we can define αA to be the $m \times n$ matrix whose (i, j) entry is αa_{ij}. Thus, by defining operations on the set $\mathrm{R}^{m\times n}$, we have created a mathematical system. The operations of addition and scalar multiplication of $\mathrm{R}^{m\times n}$ obey certain algebraic rules. These rules form the axioms that are used to define the concept of a vector space.

Vector Space Axioms

Let V be a **set** on which the operations of addition and scalar multiplication are defined. By this we mean that, with **each** pair of elements $\mathbf{x}$ and $\mathbf{y}$ in V, we can associate a **unique** element $\mathbf{x}+\mathbf{y}$ that is also in V, and with **each** element $\mathbf{x}$ in V and **each** scalar α, we can associate a unique element $\alpha\mathbf{x}$ in V. The set V together with the operations of addition and scalar multiplication is said to form a **vector space** if the following axioms are **satisfied**:

A.1 $\mathbf{x}+\mathbf{y}=\mathbf{y}+\mathbf{x}$ for any $\mathbf{x}$ and $\mathbf{y}$ in V.
A.2 $(\mathbf{x}+\mathbf{y})+\mathbf{z}=\mathbf{x}+(\mathbf{y}+\mathbf{z})$ for any $\mathbf{x},\mathbf{y}$, and $\mathbf{z}$ in V.
A.3 There exists an element $\mathbf{0}$ in V such that $\mathbf{x}+\mathbf{0}=\mathbf{x}$ for each $\mathbf{x}\in V$.
A.4 For each $\mathbf{x}\in V$, there exists an element $-\mathbf{x}$ in V such that $\mathbf{x}+(-\mathbf{x})=\mathbf{0}$.
A.5 $\alpha(\mathbf{x}+\mathbf{y})=\alpha\mathbf{x}+\alpha\mathbf{y}$ for each scalar α and any $\mathbf{x}$ and $\mathbf{y}$ in V.
A.6 $(\alpha+\beta)\mathbf{x}=\alpha\mathbf{x}+\beta\mathbf{y}$ for any scalars α and β and any $\mathbf{x}\in V$.
A.7 $(\alpha\beta)\mathbf{x}=\alpha(\beta\mathbf{x})$ for any scalars α and β and any $\mathbf{x}\in V$.
A.8 $1\mathbf{x}=\mathbf{x}$ for all $\mathbf{x}\in V$.

We will refer to the set V as the **universal set** for the vector space. Its elements are called **vectors** and are usually denoted by boldface letters such as $\mathbf{u},\mathbf{v},\mathbf{w},\mathbf{x},\mathbf{y}$, and $\mathbf{z}$. The term *scalar* will generally refer to a **real number**, although in **some cases** it will be used to refer to **complex numbers**. Scalars will generally be represented by lowercase italic letters such as a,b, and c or lowercase Greek letters such as α,β, and γ. Often the term *real vector space* is used to indicate that the set of scalars is the set of real numbers. The boldface symbol $\mathbf{0}$ was used in Axiom 3 in order to **distinguish** the zero vector from the scalar 0.

An important component of the definition is the closure properties of the two operations. These properties can be summarized as follows:

C.1 If $\mathbf{x}\in V$ and α is a scalar, then $\alpha\mathbf{x}\in V$.
C.2 If $\mathbf{x},\mathbf{y}\in V$, then $\mathbf{x}+\mathbf{y}\in V$.

R^n and $\mathrm{R}^{m\times n}$, with the usual addition and scalar multiplication of matrices, are both vector spaces.

The Vector Space $C[a,b]$

Let $C[a,b]$ denote the set of all real-valued functions that are defined and continuous on the closed interval $[a,b]$. In this case, our universal set is a set of functions. Thus, our vectors are the functions in $C[a,b]$. The sum $f+g$ of two functions in $C[a,b]$ is defined by

$$(f+g)(x)=f(x)+g(x)$$

for all x in $[a,b]$. The new function $f+g$ is an element of $C[a,b]$ since the sum of two continuous functions is continuous. If f is a function in $C[a,b]$ and α is a real number, define αf by

$$(\alpha f)(x)=\alpha f(x)$$

for all x in $[a,b]$. Clearly, αf is in $C[a,b]$ since a constant times a continuous function is always continuous. Thus we have defined the operations of addition and scalar multiplication on $C[a,b]$.

The Vector Space P_n

Let P_n denote the set of all polynomials of degree less than n. Define $p+q$ and αp, respectively, by

$$(p+q)(x)=p(x)+q(x)$$

and

$$(\alpha p)(x)=\alpha p(x)$$

for all real numbers x. In this case, the zero vector is the zero polynomial, (Page 117)

$$z(x)=0x^{n-1}+0x^{n-2}+\cdots+0x+0$$

Additional Properties of Vector Spaces

Theorem 3.1.1

If V is a vector space and $\mathbf{x}$ is any element of V, then

i. $0\mathbf{x}=\mathbf{0}$
ii. $\mathbf{x}+\mathbf{y}=\mathbf{0}$ implies that $\mathbf{y}=-\mathbf{x}$ (i. e., the additive inverse of $\mathbf{x}$ is unique)
iii. $(-1)\mathbf{x}=-\mathbf{x}$

EXERCISES

16. We can define a one-to-one correspondence between the elements of P_n and R^n by

$$p(x) = a_1 + a_2x + \cdots + a_nx^{n-1}$$
$$\leftrightarrow (a_1, \cdots, a_n)^T = \mathbf{a}$$

Show that if $p \leftrightarrow \mathbf{a}$ and $q \leftrightarrow \mathbf{b}$, then

(a) $\alpha p \leftrightarrow \alpha\mathbf{a}$ for any scalar α.

(b) $p + q \leftrightarrow \mathbf{a} + \mathbf{b}$.

[In general, two vector spaces are said to be *isomorphic* if their elements can be put into a one-to-one correspondence that is preserved under scalar multiplication and addition as in (a) and (b).]

3.2 Subspaces

Given a vector space V, it is often possible to form another vector space by taking a **subset** S of V and **using** the operations of V. **Since** V is a vector space, the operations of addition and scalar multiplication always produce another vector in V. For a new system using a subset S of V **as** its universal set to be a vector space, the set S **must** be closed under the operations of addition and scalar multiplication.

> If S is a nonempty **subset** of a vector space V, and S **satisfies** the conditions
> i. $\alpha\mathbf{x} \in S$ whenever $\mathbf{x} \in S$ for **any** scalar α
> ii. $\mathbf{x} + \mathbf{y} \in S$ whenever $\mathbf{x} \in S$ and $\mathbf{y} \in S$
> then S is said to be a **subspace** of V.

Condition (i) says that S is closed under scalar multiplication. Condition (ii) says that S is closed under addition. Thus, if we **use the operations from** V and **the elements of** S, to do arithmetic, then we will always end up with elements of S. A subspace of V, then, is a subset S that is closed under the operations of V.

Let S be a subspace of a vector space V. Using the operations of addition and scalar multiplication as defined on V, we can form a new mathematical system with S as the universal set. It is easily **seen** that all eight axioms will **remain valid** for this new system. Every subspace of a vector space is **a vector space** in its own right.

Remarks

1. In a vector space V, it can be readily verified that $\{\mathbf{0}\}$ and V are subspaces of V. All other subspaces are referred to as *proper subspaces*. We refer to $\{\mathbf{0}\}$ as the *zero subspace*.

2. To show that a subset S of a vector space forms a subspace, we must show that S is nonempty and that the closure properties (i) and (ii) in the definition are satisfied. Since every subspace **must contain** the zero vector, we **can verify** that S is nonempty by showing that $\mathbf{0} \in S$.

EXAMPLE 2: (Page 120)

Let $S = \{(x_1, x_2, x_3)^T \mid x_1 = x_2\}$. The set S is nonempty since $\mathbf{0} = (0, 0, 0)^T \in S$. To show that S is a subspace of $\mathbb{R}^3$, we need to verify that the two closure properties hold:

(i) If $\mathbf{x} = (a, a, b)^T$ is any vector in S, then

$$\alpha\mathbf{x} = (\alpha a, \alpha a, \alpha b)^T \in S$$

(ii) If $(a, a, b)^T$ and $(c, c, d)^T$ are arbitrary elements of S, then

$$(a, a, b)^T + (c, c, d)^T = (a + c, a + c, b + d)^T \in S$$

Since S is nonempty and satisfies the two closure conditions, it follows that S is a subspace of $\mathbb{R}^3$. ■

The Null Space of a Matrix

Let A be an $m \times n$ matrix. Let $N(A)$ denote the set of all solutions to the homogeneous system $A\mathbf{x} = \mathbf{0}$. Thus,

$$N(A) = \{\mathbf{x} \in \mathrm{R}^n \mid A\mathbf{x} = \mathbf{0}\}$$

We claim that $N(A)$ is a subspace of R^n. Clearly, $\mathbf{0} \in N(A)$, so $N(A)$ is nonempty, and $\alpha\mathbf{x} \in N(A)$, $\mathbf{x} + \mathbf{y} \in N(A)$. The **set of all solutions of the homogeneous system** $A\mathbf{x} = \mathbf{0}$ forms a subspace of R^n. The subspace $N(A)$ is called the *null space* of A (Page 122).

EXAMPLE 9: Determine $N(A)$ if (Page 122)

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

Using Gauss–Jordan reduction to solve $A\mathbf{x} = \mathbf{0}$, we obtain

$$\left[\begin{array}{cccc|c} 1 & 1 & 1 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 \end{array}\right] \rightarrow \left[\begin{array}{cccc|c} 1 & 1 & 1 & 0 & 0 \\ 0 & -1 & -2 & 1 & 0 \end{array}\right]$$

$$\rightarrow \left[\begin{array}{cccc|c} 1 & 0 & -1 & 1 & 0 \\ 0 & -1 & -2 & 1 & 0 \end{array}\right] \rightarrow \left[\begin{array}{cccc|c} 1 & 0 & -1 & 1 & 0 \\ 0 & 1 & 2 & -1 & 0 \end{array}\right]$$

The reduced row echelon form involves two free variables, x_3 and x_4.

$$x_1 = x_3 - x_4$$
$$x_2 = -2x_3 + x_4$$

Thus, if we set $x_3 = \alpha$ and $x_4 = \beta$, then

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

is a solution of $A\mathbf{x} = \mathbf{0}$. The vector space $N(A)$ consists of all vectors of the form

$$\alpha \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} -1 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

where α and β are scalars.

The Span of a Set of Vectors

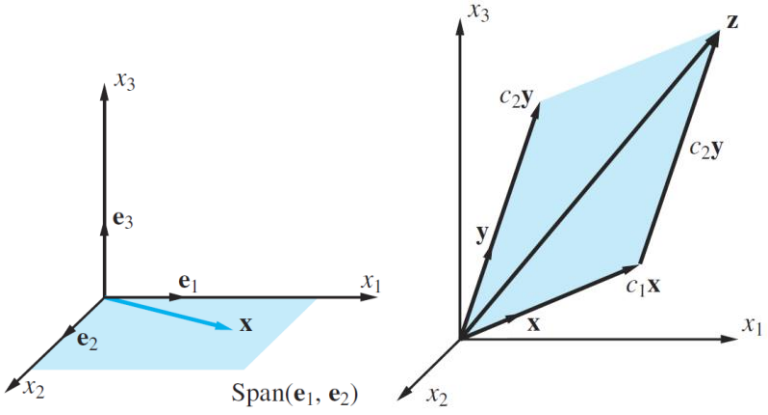
> Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ be vectors in a vector space V. A sum of the form $\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n$, where $\alpha_1, \alpha_2, \cdots, \alpha_n$ are scalars, is called a **linear combination** of $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$. The set of all linear combination of $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ is called the **span** of $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$. The span of $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ will be denoted by $\text{Span}(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n)$.

In Example 9, we saw that the null space of A was the span of the vectors $(1, -2, 1, 0)^T$ and $(-1, 1, 0, 1)^T$.

Theorem 3.2.1

If $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ are elements of a vector space V, then $\text{Span}(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n)$ is a subspace of V.

A vector $\mathbf{x}$ in R^3 is in $\text{Span}(\mathbf{e}_1, \mathbf{e}_2)$ if and only if it lies in the x_1x_2-plane in 3-space. Thus, we can think of the x_1x_2-plane as the geometrical representation of the subspace $\text{Span}(\mathbf{e}_1, \mathbf{e}_2)$ as shown below. Similarly, given two vectors $\mathbf{x}$ and $\mathbf{y}$, if $(0,0,0)$, (x_1, x_2, x_3), and (y_1, y_2, y_3) are not collinear, these points determine a plane. If $\mathbf{z} = c_1\mathbf{x} + c_2\mathbf{y}$, then $\mathbf{z}$ is a sum of vectors parallel to $\mathbf{x}$ and $\mathbf{y}$ and hence must lie on the plane determined by the two vectors also shown below. In general, if two vectors $\mathbf{x}$ and $\mathbf{y}$ can be used to determine a plane in 3-space, that plane is the geometrical representation of $\text{Span}(\mathbf{x}, \mathbf{y})$.



Spanning Set for a Vector Space

Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ be vectors in a vector space V. We will refer to $\text{Span}(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ as the subspace of V *spanned* by $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$. It may happen that $\text{Span}(\mathbf{v}_1, \cdots, \mathbf{v}_n) = V$, in which case we say that the vectors $\mathbf{v}_1, \cdots, \mathbf{v}_n$ *span* V, or that $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is a *spanning set* for V. Thus, we have the following definition:

> The set $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is a **spanning set** for V if and only if every vector in V can be written as a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$.

EXAMPLE 11: Which of the following are spanning sets for R^3? (Page 125)

(a) $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, (1, 2, 3)^T\}$

(b) $\{(1, 1, 1)^T, (1, 1, 0)^T, (1, 0, 0)^T\}$

(c) $\{(1, 0, 1)^T, (0, 1, 0)^T\}$

(d) $\{(1, 2, 4)^T, (2, 1, 3)^T, (4, -1, 1)^T\}$

Linear Systems Revisited

Let S be the solution set to a consistent $m \times n$ linear system $A\mathbf{x} = \mathbf{b}$. In the case that $\mathbf{b} = \mathbf{0}$ we have that $S = N(A)$ and consequently the solution set forms a subspace of R^n. If $\mathbf{b} \neq \mathbf{0}$, then S does not form a subspace of R^n; however, if one can find a particular solution $\mathbf{x}_0$, then it is possible to represent any solution vector in terms of $\mathbf{x}_0$ and a vector $\mathbf{z}$ from the null space of A.

Let $A\mathbf{x} = \mathbf{b}$ be a consistent linear system and let $\mathbf{x}_0$ be a particular solution to the system. If there is another solution $\mathbf{x}_1$ to the system, then the difference vector $\mathbf{z} = \mathbf{x}_1 - \mathbf{x}_0$ must be in $N(A)$ since

$$A\mathbf{z} = A\mathbf{x}_1 - A\mathbf{x}_0 = \mathbf{b} - \mathbf{b} = \mathbf{0}$$

Thus if there is a second solution, it must be of the form $\mathbf{x}_1 = \mathbf{x}_0 + \mathbf{z}$ where $\mathbf{z} \in N(A)$.

In general, if $\mathbf{x}_0$ is a particular solution to $A\mathbf{x} = \mathbf{b}$ and $\mathbf{z}$ is any vector in $N(A)$, then setting $\mathbf{y} = \mathbf{x}_0 + \mathbf{z}$, we have

$$A\mathbf{y} = A\mathbf{x}_0 + A\mathbf{z} = \mathbf{b} + \mathbf{0} = \mathbf{b}$$

So $\mathbf{y} = \mathbf{x}_0 + \mathbf{z}$ must also be a solution to the system $A\mathbf{x} = \mathbf{b}$.

Theorem 3.2.2

If the linear system $A\mathbf{x} = \mathbf{b}$ is consistent and $\mathbf{x}_0$ is a particular solution, then a vector $\mathbf{y}$ will also be a solution if and only if $\mathbf{y} = \mathbf{x}_0 + \mathbf{z}$ where $\mathbf{z} \in N(A)$ (Page 127).

EXERCISES

7. Show that $C^n[a, b]$ is a subspace of $C[a, b]$.

A function f is n times differentiable on $[a, b]$ then that means $f \in C^n[a, b]$.

3.3 Linear Independence

To begin with, we restrict ourselves to vector spaces that can be generated from a finite set of elements. Each vector in the vector space can be built up from the elements in this generating set using only the operations of addition and scalar multiplication. The generating set is usually referred to as a spanning set. In particular, it is desirable to find a *minimal* spanning set. By minimal we mean a spanning set with no unnecessary elements (i.e., all the elements in the set are needed in order to span the vector space). To see how to find a minimal spanning set, it is **necessary** to **consider** how the vectors in the collection *depend* on each other. Consequently, we introduce the concepts of *linear dependence* and *linear independence*. These simple concepts provide the **keys** to understanding the structure of vector spaces.

We can generalize this example (Page 130) by making the following observations:

I. **If** $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ span a vector space V and one of these vectors can be written as a linear combination of the other $n - 1$ vectors, then those $n - 1$ vectors span V

II. Given n vectors $\mathbf{v}_1, \cdots, \mathbf{v}_n$, it is possible to write one of the vectors as a linear combination of the other $n - 1$ vectors if and only if there exist scalars $c_1, \cdots, c_n$, not all zero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$$

The vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ in a vector space V are said to be **linearly independent** if

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$$

implies that all the scalars $c_1, \cdots, c_n$ must equal 0.

It follows from (I) and (II) that, **if** $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is a **minimal** spanning set, then $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ are **linearly independent**. Conversely, if $\mathbf{v}_1, \cdots, \mathbf{v}_n$ are linearly independent and span V, then $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is a minimal spanning set for V. A minimal spanning set is called a ***basis***.

EXAMPLE 1: The vectors $\begin{bmatrix}1\\1\end{bmatrix}$ and $\begin{bmatrix}1\\2\end{bmatrix}$ are linearly independent, since if

$$c_1\begin{bmatrix}1\\1\end{bmatrix} + c_2\begin{bmatrix}1\\2\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}$$

then

$$c_1 + c_2 = 0$$
$$c_1 + 2c_2 = 0$$

and the only solution to this system is $c_1 = 0$, $c_2 = 0$.

The vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ in a vector space V are said to be **linearly dependent** if there exist scalars $c_1, c_2, \cdots, c_n$, not all zero, such that

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$$

EXAMPLE 2: Let $\mathbf{x} = (1,2,3)^T$. The vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$, and $\mathbf{x}$ are linearly dependent, since

$$\mathbf{e}_1 + 2\mathbf{e}_2 + 3\mathbf{e}_3 - \mathbf{x} = \mathbf{0}$$

(In this case $c_1 = 1$, $c_2 = 2$, $c_3 = 3$, $c_4 = -1$.)

Given a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ in a vector space V, it is trivial to find scalars $c_1, c_2, \cdots, c_n$ such that

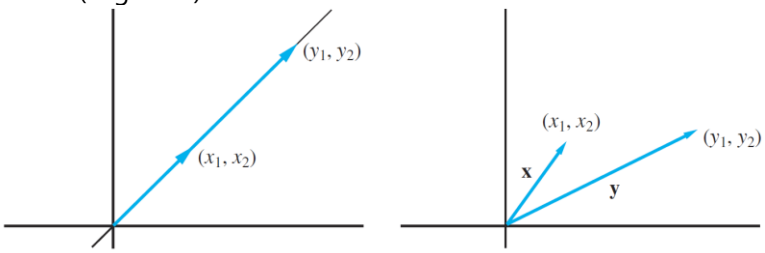
$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$$

Just take

$$c_1 = c_2 = \cdots = c_n = 0.$$

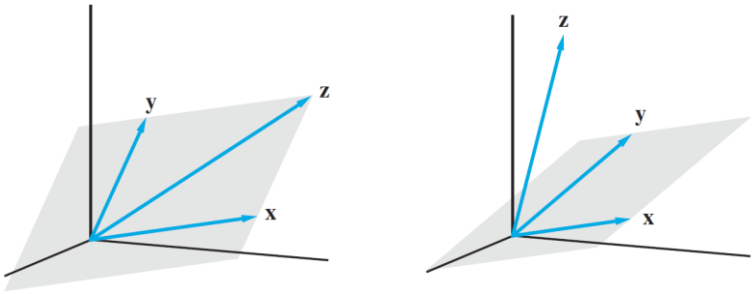
If there are nontrivial choices of scalars for which the linear combination $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$ equals the zero vector, then $\mathbf{v}_1, \cdots, \mathbf{v}_n$ are linearly dependent. If the *only* way the linear combination $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$ can equal the zero vector is for all the scalars $c_1, \cdots, c_n$ to be 0, then $\mathbf{v}_1, \cdots, \mathbf{v}_n$ are linearly independent

Geometric Interpretation (Page 133)



(a) **x** and **y** linearly dependent

(b) **x** and **y** linearly independent



Theorems and Examples (Page 134)

Theorem 3.3.1

Let $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ be n vectors in R^n and let $X = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$. The vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ will be linearly dependent if and only if X is singular.

We can use Theorem 3.3.1 to **test** whether n vectors are linearly independent in R^n. Simply form a matrix X whose columns are the vectors being tested. To determine whether X is singular, calculate the value of $\det(X)$. If $\det(X) = 0$, the vectors are linearly dependent. If $\det(X) \neq 0$, the vectors are linearly independent.

To determine whether k vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k$ in R^n are linearly independent we can rewrite the equation

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_k\mathbf{x}_k = \mathbf{0}$$

as a linear system $X\mathbf{c} = \mathbf{0}$, where $X = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k)$. If $k \neq n$, then the matrix X is not square, so we cannot **use** determinants to decide whether the vectors are linearly independent. The system is homogeneous, so it has the trivial solution $\mathbf{c} = \mathbf{0}$. It will have nontrivial solutions if and only if the row echelon forms of X involve free variables. If there are nontrivial solutions, then the vectors are linearly dependent. If there are no free variables, then $\mathbf{c} = \mathbf{0}$ is the only solution, and hence the vectors must be linearly independent.

Next, we consider a very important property of linearly independent vectors: Linear combinations of linearly independent vectors are unique. More precisely, we have the following theorem:

Theorem 3.3.2

Let $\mathbf{v}_1, \cdots, \mathbf{v}_n$ be vectors in a vector space V. A vector $\mathbf{v} \in \mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n)$ can be written uniquely as a linear combination of $\mathbf{v}_1, \cdots, \mathbf{v}_n$ if and only if $\mathbf{v}_1, \cdots, \mathbf{v}_n$ are linearly independent.

Vector Spaces of Functions

The Vector Space P_n (Page 137)

The Vector Space $C^{(n-1)}[a, b]$

Let $f_1, f_2, \cdots, f_n$ be functions in $C^{(n-1)}[a, b]$, and define the function $W[f_1, f_2, \cdots, f_n](x)$ on $[a, b]$ by

$$W[f_1, f_2, \cdots, f_n](x) = \begin{vmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f_1'(x) & f_2'(x) & \cdots & f_n'(x) \\ \vdots & & & \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \cdots & f_n^{(n-1)}(x) \end{vmatrix}$$

The function $W[f_1, f_2, \cdots, f_n]$ is called the **Wronskian** of $f_1, f_2, \cdots, f_n$.

Theorem 3.3.3

Let $f_1, f_2, \cdots, f_n$ be elements of $C^{(n-1)}[a, b]$. If there exists a point x_0 in $[a, b]$ such that $W[f_1, f_2, \cdots, f_n](x_0) \neq 0$, then $f_1, f_2, \cdots, f_n$ are linearly independent.

The converse of Theorem 3.3.3 is not valid (Page 139).

EXERCISES

13. Prove that any finite set of vectors that contains the zero vector must be linearly dependent.

3.4 Basis and Dimension

The vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ form a **basis** for a vector space V if and only if

i. $\mathbf{v}_1, \cdots, \mathbf{v}_n$ are linearly independent.

ii. $\mathbf{v}_1, \cdots, \mathbf{v}_n$ **span** V.

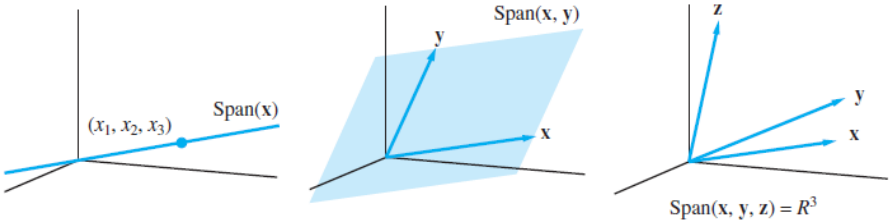
Theorem 3.4.1

If $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is a spanning set for a vector space V, then any collection of m vectors in V, where $m > n$, is linearly dependent.

Corollary 3.4.2

If both $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ are bases for a vector space V, then $n = m$.

Let V be a vector space. If V has a basis consisting of n vectors, we say that V has **dimension** n. The subspace $\{\mathbf{0}\}$ of V is said to have dimension 0. V is said to be **finite dimensional** if there is a finite set of vectors that spans V; otherwise, we say that V is **infinite dimensional**.



Theorem 3.4.3

If V is a vector space of dimension $n > 0$, then

I. any set of n linearly independent vectors spans V.

II. any n vectors that span V are linearly independent.

Theorem 3.4.4

If V is a vector space of dimension $n > 0$, then (**find** minimal spanning set)

i. no set of fewer than n vectors can span V.

ii. any subset of fewer than n linearly independent vectors can be extended to form a basis for V.

iii. any spanning set containing more than n vectors can be pared down to form a basis for V.

Standard Bases

We refer to the basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ as the **standard basis** because it is the **most natural one** to use for representing vectors in R^3. More generally, the **standard basis** for R^n is the set $\{\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n\}$.

The most natural way to represent matrices in $\mathrm{R}^{2\times 2}$ is in terms of the basis $\{E_{11}, E_{12}, E_{21}, E_{22}\}$, which is the **standard basis** for $\mathrm{R}^{2\times 2}$.

The standard way to represent a polynomial in P_n is $\{1, x, x^2, \cdots, x^{n-1}\}$.

Although these standard bases appear to be the **simplest** and most natural to use, they **are not** the most **appropriate** bases for many applied problems.

3.5 Change of Basis

Many applied problems can be simplified by changing from one coordinate system to another. Changing coordinate systems in a vector space is essentially the same as changing from one basis to another.

We will show that this can be accomplished by multiplying a given coordinate vector $\mathbf{x}$ by a nonsingular matrix S. The product $\mathbf{y} = S\mathbf{x}$ will be the coordinate vector for the new coordinate system.

Changing Coordinates in R^2

The standard basis for R^2 is $\{\mathbf{e}_1, \mathbf{e}_2\}$. Any vector $\mathbf{x}$ in R^2 can be expressed as a linear combination

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2$$

The scalars x_1 and x_2 can be thought of as the *coordinates* of $\mathbf{x}$ with respect to the standard basis. Actually, for any basis $[\mathbf{y}, \mathbf{z}]$ for R^2, it follows from Theorem 3.3.2 that a given vector $\mathbf{x}$ can be represented uniquely as a linear combination

$$\mathbf{x} = \alpha\mathbf{y} + \beta\mathbf{z}$$

The scalars α and β are the coordinates of $\mathbf{x}$ with respect to the basis $\{\mathbf{y}, \mathbf{z}\}$. Let us order the basis elements so that $\mathbf{y}$ is considered the first basis vector and $\mathbf{z}$ is considered the second, and denote the ordered basis by $[\mathbf{y}, \mathbf{z}]$. We can then refer to the vector $(\alpha, \beta)^T$ as the ***coordinate vector*** of $\mathbf{x}$ with respect to $[\mathbf{y}, \mathbf{z}]$. Note that, if we reverse the order of the basis vectors and take $[\mathbf{z}, \mathbf{y}]$, then we must also reorder the coordinate vector.

APPLICATION 1 Population Migration

Suppose that the total population of a large metropolitan area remains relatively fixed; however, each year 6 percent of the people living in the city move to the suburbs and 2 percent of the people living in the

suburbs move to the city. If, initially, 30 percent of the population lives in the city and 70 percent lives in the suburbs, what will these percentages be in 10 years? 30 years? 50 years? What are the long-term implications? The changes in population can be determined by matrix multiplications. If we set

$$A = \begin{bmatrix} 0.94 & 0.02 \\ 0.06 & 0.98 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_0 = \begin{bmatrix} 0.30 \\ 0.70 \end{bmatrix}$$

This application is an example of a type of mathematical model called a *Markov process*. The sequence of vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots$ is called a *Markov chain*. The matrix A has a special structure in that its entries are nonnegative and its columns all add up to 1. Such matrices are called *stochastic matrices*. (Page 148)

Changing Coordinates

Once we have decided to work with a new basis, we have the problem of finding the coordinates with respect to that basis. Suppose, for example, that instead of using the standard basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ for R^2, we wish to use a different basis, say

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Indeed, we may want to switch back and forth between the two coordinate systems. Let us consider the following two problems (Page 150)

I. Given a vector $\mathbf{x} = (x_1, x_2)^T$, find its coordinates with respect to $\mathbf{u}_1$ and $\mathbf{u}_2$

II. Given a vector $c_1\mathbf{u}_1 + c_2\mathbf{u}_2$, find its coordinates with respect to $\mathbf{e}_1$ and $\mathbf{e}_2$

We will solve **II** first: given any **coordinate vector** $\mathbf{c}$ with respect to $\{\mathbf{u}_1, \mathbf{u}_2\}$, to find the corresponding coordinate vector (Tricky!) $\mathbf{x}$ with respect to $\{\mathbf{e}_1, \mathbf{e}_2\}$, we simply multiply U times $\mathbf{c}$:

$$\mathbf{x} = U\mathbf{c}$$

$$U = (\mathbf{u}_1, \mathbf{u}_2) = \begin{bmatrix} 3 & 1 \\ 2 & 1 \end{bmatrix}$$

The matrix U is called the *transition matrix* from the ordered basis $\{\mathbf{u}_1, \mathbf{u}_2\}$ to the standard basis $\{\mathbf{e}_1, \mathbf{e}_2\}$.

To solve problem **I**, we must find the transition matrix from $\{\mathbf{e}_1, \mathbf{e}_2\}$ to $\{\mathbf{u}_1, \mathbf{u}_2\}$. The matrix U is nonsingular, since its column vectors, $\mathbf{u}_1$ and $\mathbf{u}_2$, are linearly independent. It follows that

$$\mathbf{c} = U^{-1}\mathbf{x}$$

Thus, given a vector

$$\mathbf{x} = (x_1, x_2)^T = x_1\mathbf{e}_1 + x_2\mathbf{e}_2$$

we need only multiply by U^{-1} to find its coordinate vector with respect to $\{\mathbf{u}_1, \mathbf{u}_2\}$.

U^{-1} is the transition matrix from $\{\mathbf{e}_1, \mathbf{e}_2\}$ to $\{\mathbf{u}_1, \mathbf{u}_2\}$.

Now let us consider the general problem of changing from one ordered basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ of R^2 to another ordered basis $\{\mathbf{u}_1, \mathbf{u}_2\}$. In this case, we assume that, for a given vector $\mathbf{x}$, its coordinates with respect to $\{\mathbf{v}_1, \mathbf{v}_2\}$ are known:

$$\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$$

Now we wish to represent $\mathbf{x}$ as a sum $d_1\mathbf{u}_1 + d_2\mathbf{u}_2$

$$\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = d_1\mathbf{u}_1 + d_2\mathbf{u}_2$$

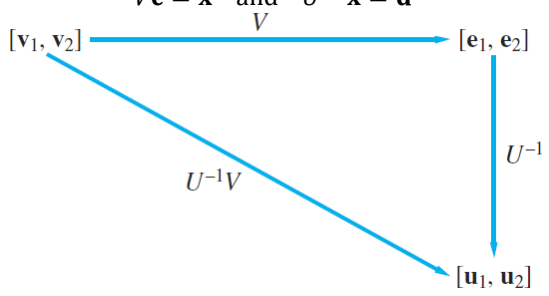
If we set $V = (\mathbf{v}_1, \mathbf{v}_2)$ and $U = (\mathbf{u}_1, \mathbf{u}_2)$, then

$$V\mathbf{c} = U\mathbf{d}$$

$$\mathbf{d} = U^{-1}V\mathbf{c}$$

The change of basis from $\{\mathbf{v}_1, \mathbf{v}_2\}$ to $\{\mathbf{u}_1, \mathbf{u}_2\}$ can also be viewed as a two-step process. First we change from $\{\mathbf{v}_1, \mathbf{v}_2\}$ to the standard basis, $\{\mathbf{e}_1, \mathbf{e}_2\}$, and then we change from the standard basis to $\{\mathbf{u}_1, \mathbf{u}_2\}$.

$$V\mathbf{c} = \mathbf{x} \quad \text{and} \quad U^{-1}\mathbf{x} = \mathbf{d}$$



Change of Basis for a General Vector Space

We begin by defining coordinate vectors for an n-dimensional vector space

> Let V be a vector space and let $E = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ be an ordered basis for V. If $\mathbf{v}$ is any element of V, then $\mathbf{v}$ can be written in the form
>
> $$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$$
>
> where $c_1, c_2, \cdots, c_n$ are scalars. Thus, we can associate with each vector $\mathbf{v}$ a unique vector $\mathbf{c} = (c_1, c_2, \cdots, c_n)^T$ in R^n. The vector $\mathbf{c}$ defined in this way is called the **coordinate vector** of $\mathbf{v}$ with respect to the ordered basis E and is denoted $[\mathbf{v}]_E$. The c_i's are called the **coordinates** of $\mathbf{v}$ relative to E.

In the case of R^n the transition matrices will be $n \times n$.

If V is any n-dimensional vector space, it is possible to change from one basis to another by means of an $n \times n$ transition matrix. We will show that such a transition matrix is necessarily nonsingular. To see how this is done, let $E = \{\mathbf{w}_1, \cdots, \mathbf{w}_n\}$ and $F = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be two ordered bases for V. The key step is to express each basis vector $\mathbf{w}_j$ as a linear combination of the $\mathbf{v}_i$'s. Let $\mathbf{v} \in V$, $\mathbf{x} = [\mathbf{v}]_E, \mathbf{y} = [\mathbf{v}]_F$ (Page 154)

$$\begin{aligned}
\mathbf{w}_1 &= s_{11}\mathbf{v}_1 + s_{21}\mathbf{v}_2 + \cdots + s_{n1}\mathbf{v}_n \\
\mathbf{w}_2 &= s_{12}\mathbf{v}_1 + s_{22}\mathbf{v}_2 + \cdots + s_{n2}\mathbf{v}_n \\
&\vdots \\
\mathbf{w}_n &= s_{1n}\mathbf{v}_1 + s_{2n}\mathbf{v}_2 + \cdots + s_{nn}\mathbf{v}_n
\end{aligned}$$

$$y_i = \sum_{j=1}^{n} s_{ij}x_j \quad i = 1, \cdots, n$$

$$\mathbf{y} = S\mathbf{x}$$

$$S\mathbf{x} = \mathbf{y} \quad \text{if and only if} \quad x_1\mathbf{w}_1 + \cdots + x_n\mathbf{w}_n = y_1\mathbf{v}_1 + \cdots + y_n\mathbf{v}_n$$

$$S^{-1}\mathbf{y} = \mathbf{x} \quad \text{if and only if} \quad y_1\mathbf{v}_1 + \cdots + y_n\mathbf{v}_n = x_1\mathbf{w}_1 + \cdots + x_n\mathbf{w}_n$$

The matrix S is referred to as the ***transition matrix*** corresponding to the change of basis from $\{\mathbf{w}_1, \cdots, \mathbf{w}_n\}$ to $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$. The matrix S^{-1} is the transition matrix used to change basis from $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ to $\{\mathbf{w}_1, \cdots, \mathbf{w}_n\}$.

We have seen that each transition matrix is nonsingular. Actually, any nonsingular matrix can be thought of as a transition matrix.

EXERCISES

11. Let $E = \{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ and $F = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be two ordered bases for R^n, and set

$$U = (\mathbf{u}_1, \cdots, \mathbf{u}_n), \quad V = (\mathbf{v}_1, \cdots, \mathbf{v}_n)$$

Show that the transition matrix from E to F can be determined by calculating the reduced row echelon form of $(V|U)$.

3.6 Row Space and Column Space

If A is an $m \times n$ matrix, each row of A is an n-tuple of real numbers and hence can be considered as a vector in $\mathrm{R}^{1\times n}$. The m vectors corresponding to the rows of A will be referred to as the *row vectors* of A. Similarly, each column of A can be considered as a vector in R^m, and we can associate n *column vectors* with the matrix A.

> If A is an $m \times n$ matrix, the subspace of $\mathrm{R}^{1\times n}$ **spanned** by the **row vectors of** A is called the **row space** of A. The subspace of R^m spanned by the **column vectors of** A is called the **column space** of A.

Theorem 3.6.1

Two **row** equivalent matrices have the same row space.

> The **rank** of a matrix A, denoted $\mathrm{rank}(A)$, is the dimension of the **row** space of A.

To determine the rank of a matrix, we can reduce the matrix to row echelon form. The **nonzero** rows of the row echelon matrix will form a basis for the row space.

EXAMPLE 2: Let

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

Reducing A to row echelon form, we obtain the matrix

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

Clearly, $(1, -2,3)$ and $(0,1,5)$ form a basis for the row space of U. Since U and A are row equivalent, they have the same row space, and hence the rank of A is 2.

Linear Systems

The concepts of row space and column space are useful in the study of linear systems. A system $A\mathbf{x} = \mathbf{b}$ can be written in the form

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

Theorem 1.3.1, can now be restated in terms of the column space of the matrix.

Theorem 3.6.2 Consistency Theorem for Linear Systems

A linear system $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b}$ is in the **column** space of A.

If $\mathbf{b}$ is replaced by the zero vector, then the above formula becomes

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{0}$$

It follows that the system $A\mathbf{x} = \mathbf{0}$ will have only the trivial solution $\mathbf{x} = \mathbf{0}$ if and only if the column vectors of A are linearly independent.

Theorem 3.6.3

Let A be a $m \times n$ matrix. The linear system $A\mathbf{x} = \mathbf{b}$ is consistent for **every** $\mathbf{b} \in \mathrm{R}^m$ if and only if the column vectors of A span R^m. The system $A\mathbf{x} = \mathbf{b}$ has at most one solution for every $\mathbf{b} \in \mathrm{R}^m$ if and only if the column vectors of A are linearly independent.

Let A be a $m \times n$ matrix. If the column vectors of A span R^m, then n must be greater than or equal to m, since no set of fewer than m vectors could span R^m. If the columns of A are linearly independent, then n must be less than or equal to m, since every set of more than m vectors in R^m is linearly dependent. **Thus**, if the column vectors of A form a basis for R^m, then n must equal m.

Corollary 3.6.4

An $n \times n$ matrix A is nonsingular if and only if the column vectors of A form a basis for R^n.

In general, the rank and the dimension of the null space always add up to the number of columns of the matrix. The dimension of the **null space** of a matrix is called the ***nullity*** of the matrix.

Theorem 3.6.5 The Rank–Nullity Theorem

If A is an $m \times n$ matrix, then the rank of A plus the nullity of A equals n.

Proof:

Let U be the reduced row echelon form of A. The system $A\mathbf{x} = \mathbf{0}$ is equivalent to the system $U\mathbf{x} = \mathbf{0}$. If A has rank r, then U will have r nonzero rows, and consequently the system $U\mathbf{x} = \mathbf{0}$ will involve r lead variables and $n - r$ **free** variables. The dimension of $N(A)$ will equal the number of free variables.

EXAMPLE 3: (Page 159)

The Column Space

The matrices A and U in Example 3 have different column spaces; **however**, their column vectors **satisfy** the same dependency **relations** (Page 160).

In general, if A is an $m \times n$ matrix and U is the row echelon form of A, then, since $A\mathbf{x} = \mathbf{0}$ If and only if $U\mathbf{x} = \mathbf{0}$, their column vectors **satisfy** the **same** dependency relations.

Theorem 3.6.6

If A is an $m \times n$ matrix, the dimension of the row space of A equals the dimension of the column space of A. (**Note**: it's different with the Rank–Nullity Theorem since (1) $\{\mathbf{x} \in \mathrm{R}^n | A\mathbf{x} = \mathbf{0}\}$ and (2) $\mathrm{Span}\{a_1, a_2, \cdots, a_n\}$ is the subspace of R^m)

We can use the row echelon form U of A to find a basis for the column space of A. We **need only** determine the columns of U that correspond to the leading 1's. These **same** columns of A will be linearly independent and form a basis for the column space of A.

Note

The row echelon form U tells us only which columns of A to use to form a basis. We cannot use the column vectors from U, since, in general, U and A have **different** column spaces.

EXAMPLE 5: **Find** the dimension of the subspace of R^4 spanned by

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 2 \\ 5 \\ -3 \\ 2 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 2 \\ 4 \\ -2 \\ 0 \end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix} 3 \\ 8 \\ -5 \\ 4 \end{bmatrix}$$

Solution

The subspace $\mathrm{Span}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)$ is the same as the column space of the matrix

$$X = \begin{bmatrix} 1 & 2 & 2 & 3 \\ 2 & 5 & 4 & 8 \\ -1 & -3 & -2 & -5 \\ 0 & 2 & 0 & 4 \end{bmatrix}$$

The row echelon form of X is

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

The first two columns $\mathbf{x}_1, \mathbf{x}_2$ of X will form a basis for the column space of X. Thus, $\dim \mathrm{Span}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) = 2$.

EXERCISES

6. if $\mathbf{b}$ is in the column space of A and the column vectors of A are linearly dependent, there will have infinitely many solutions of the linear system $A\mathbf{x} = \mathbf{b}$.

11. Let A be an $m \times n$ matrix. Prove that

$$\mathrm{rank}(A) \leq \min(m, n)$$

19. Let A be an $m \times n$ matrix with rank equal to n. If $\mathbf{x} \neq \mathbf{0}$ and $\mathbf{y} = A\mathbf{x}$, then $\mathbf{y} \neq \mathbf{0}$.

CHAPTER 4 Linear Transformations

4.1 Definition and Examples

In the study of vector spaces, the most important types of mappings are linear transformations.

> A mapping L from a vector space V into a vector space W is said to be a **linear transformation** if
>
> $$L(\alpha \mathbf{v}_1 + \beta \mathbf{v}_2) = \alpha\, L(\mathbf{v}_1) + \beta\, L(\mathbf{v}_2)$$
>
> for all $\mathbf{v}_1, \mathbf{v}_2 \in V$ and for all scalars α and β.

Notation

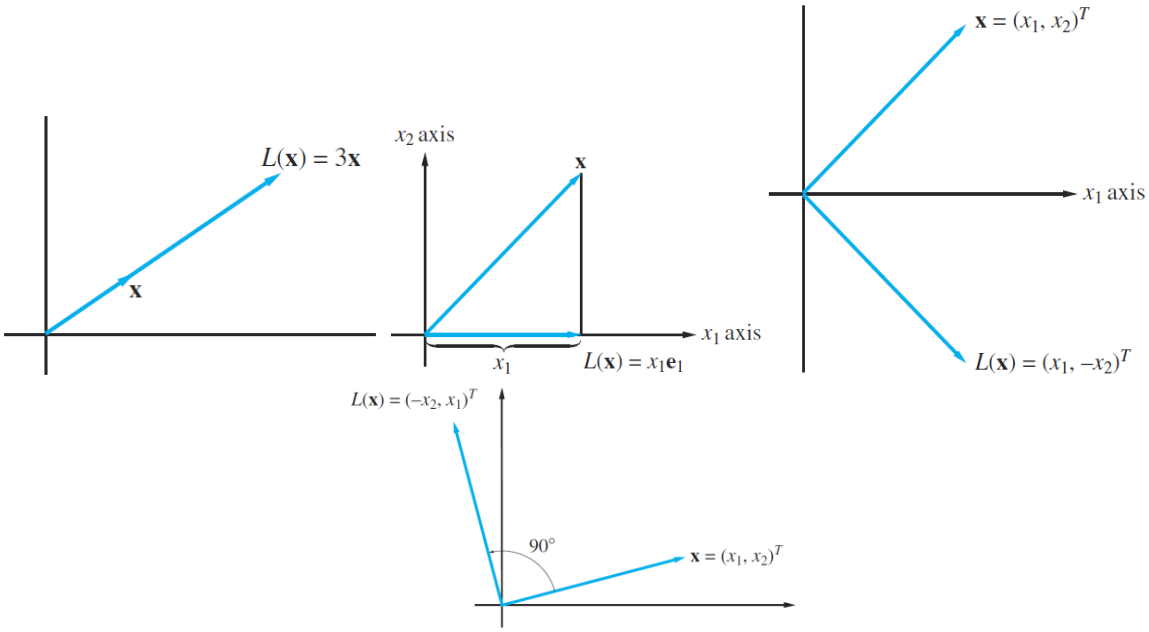
A mapping L from a vector space V into a vector space W will be denoted

$$L: V \to W$$

When the **arrow notation** is used, it will be **assumed** that V and W represent vector spaces.

In the case that the vector spaces V and W are the same, we will refer to a linear transformation $L: V \to V$ as a *linear operator* on V. Thus, a linear operator is a linear transformation that maps a vector space V into itself.

Linear Operators on R^2 (Page 170)



Linear Transformations from R^n to R^m (Page 174)

In general, if A is any $m \times n$ matrix, we can define a linear transformation L_A from R^n to R^m by

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

for each $\mathbf{x} \in \mathrm{R}^n$. The transformation L_A is linear, since

$$L_A(\alpha\mathbf{x} + \beta\mathbf{y}) = A(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha A\mathbf{x} + \beta A\mathbf{y} = \alpha L_A(\mathbf{x}) + \beta L_A(\mathbf{y})$$

Thus, we can think of each $m \times n$ matrix A as defining a linear transformation from R^n to R^m.

Linear Transformations from V **to** W

If L is a linear transformation mapping a vector space V into a vector space W, then

i. $L(\mathbf{0}_V) = \mathbf{0}_W$ (where $\mathbf{0}_V$ and $\mathbf{0}_W$ are the zero vectors in V and W, respectively)

ii. if $\mathbf{v}_1, \cdots, \mathbf{v}_n$ are elements of V and $\alpha_1, \cdots, \alpha_n$ are scalars, then

$$L(\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_n\mathbf{v}_n) = \alpha_1\, L(\mathbf{v}_1) + \alpha_2\, L(\mathbf{v}_2) + \cdots + \alpha_n\, L(\mathbf{v}_n)$$

iii. $L(-\mathbf{v}) = -L(\mathbf{v})$ for all $\mathbf{v} \in V$. ($L(-\mathbf{v})$ is the additive inverse of $L(\mathbf{v})$)

EXAMPLE 8: If V is any vector space, then the identity operator I is defined by

$$I(\mathbf{v}) = \mathbf{v}$$

for all $\mathbf{v} \in V$. Clearly, I is a linear transformation that maps V into itself:

$$I(\alpha\mathbf{v_1} + \beta\mathbf{v_2}) = \alpha\mathbf{v_1} + \beta\mathbf{v_2} = \alpha I(\mathbf{v_1}) + \beta I(\mathbf{v_2})$$

The Image and Kernel

Let $L: V \to W$ be a linear transformation. We close this section by considering the effect that L has on subspaces of V. Of particular importance is the set of vectors in V that get mapped into the zero vector of W.

> Let $L: V \to W$ be a linear transformation. The **kernel** of L, denoted $\ker(L)$, is defined by
> $$\ker(L) = \{\mathbf{v} \in V \mid L(\mathbf{v}) = \mathbf{0}_W\}$$

(use the Rank–Nullity Theorem ↑)

> Let $L: V \to W$ be a linear transformation and let S be a subspace of V. The **image** of S, denoted $L(S)$, is defined by
> $$L(S) = \{\mathbf{w} \in W \mid \mathbf{w} = L(\mathbf{v}) \quad \text{for some } \mathbf{v} \in S\}$$
> The image of the entire vector space, $L(V)$, is called the **range** of L.

(use the Theorem 3.6.6 ↑)

Let $L: V \to W$ be a linear transformation. It is easily seen that $\ker(L)$ is a subspace of V, and if S is any subspace of V, then $L(S)$ is a subspace of W. In particular, $L(V)$ is a subspace of W. Indeed, we have the following theorem:

Theorem 4.1.1

If $L: V \to W$ is a linear transformation and S is a subspace of V, then

i. $\ker(L)$ is a subspace of V
ii. $L(S)$ is a subspace of W

EXERCISES

3. Let $\mathbf{a}$ be a fixed nonzero vector in R^2. A mapping of the form

$$L(\mathbf{x}) = \mathbf{x} + \mathbf{a}$$

is called a *translation*. That a translation is not a linear operator.

20. Let $L: V \to W$ be a linear transformation, and let T be a subspace of W. The *inverse image* of T, denoted $L^{-1}(T)$, is defined by

$$L^{-1}(T) = \{\mathbf{v} \in V | L(\mathbf{v}) \in T\}$$

That $L^{-1}(T)$ is a subspace of V.

21. A linear transformation $L: V \to W$ is said to be *one-to-one* if $L(\mathbf{v}_1) = L(\mathbf{v}_2)$ implies that $\mathbf{v}_1 = \mathbf{v}_2$ (i.e., no two distinct vectors $\mathbf{w} \in W$). Show that L is one-to-one if and only if $\ker(L) = \{\mathbf{0}_V\}$.

24. Let A be a $a \times a$ matrix, and let L_A be the linear operator defined by

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

Show that

(a) L_A maps R^2 onto the column space of A.
(b) if A is nonsingular, then L_A maps R^2 onto R^2.

4.2 Matrix Representations of Linear Transformations

In Section 4.1, it was shown that each $m \times n$ matrix A defines a linear transformation L_A from R^n to R^m, where

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

for each $\mathbf{x} \in \mathrm{R}^n$. In this section, we will see that, for each linear transformation L mapping R^n into R^m, there is an $m \times n$ matrix A such that

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

Theorem 4.2.1

If L is a linear transformation mapping $\mathbf{R^n}$ **into** $\mathbf{R^m}$, there is an $m \times n$ matrix A such that

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

for each $\mathbf{x} \in \mathrm{R}^n$. In fact, the jth column vector of A is given by

$$\mathbf{a}_j = L(\mathbf{e}_j) \quad j = 1,2,\cdots,n$$

Proof:

For $j = 1,\cdots,n$, define

$$\mathbf{a}_j = L(\mathbf{e}_j)$$

and let

$$A = (\mathbf{a}_{ij}) = (\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n)$$

If

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

is an arbitrary element of R^n, then

$$\begin{aligned} L(\mathbf{x}) &= x_1\,L(\mathbf{e}_1) + x_2\,L(\mathbf{e}_2) + \cdots + x_n\,L(\mathbf{e}_n) \\ &= x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n \\ &= (\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n)\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \\ &= A\mathbf{x} \end{aligned}$$

We have established that each linear transformation from R^n into R^m can be represented in terms of an $m \times n$ matrix. Theorem 4.2.1 tells us how to construct the matrix A corresponding to a particular linear transformation L. To get the first column of A, see what L does to the first basis element $\mathbf{e}_1$ of R^n. Set $\mathbf{a}_1 = L(\mathbf{e}_1)$. To get the second column of A, determine the effect of L on $\mathbf{e}_2$ and $\mathbf{a}_2 = L(\mathbf{e}_2)$, and so on. Since the standard basis elements $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$ (the column vectors of the $n \times n$ identity matrix) are used for R^n, and the column vectors of the $m \times m$ identity matrix are being used as a basis for R^m, we refer to A as the *standard matrix representation* of L.

Now that we have seen how matrices are used to represent linear transformations from R^n to R^m, we may ask whether it is possible to find a similar representation for linear transformations from V into W, where V and W are vector spaces of dimension n and m, respectively. To see how this is done, let $E = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ be an ordered basis for V and $F = \{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_m\}$ be an ordered basis for W. Let L be a linear transformation mapping V into W. If $\mathbf{v}$ is any vector in V, then we can express $\mathbf{v}$ in terms of the basis E:

$$\mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$$

We will show that there exists an $m \times n$ matrix A representing the linear transformation L, in the sense that

$$A\mathbf{x} = \mathbf{y} \quad \text{if and only if} \quad L(\mathbf{v}) = y_1\mathbf{w}_1 + y_2\mathbf{w}_2 + \cdots + y_m\mathbf{w}_m$$

The matrix A characterizes the effect of the linear transformation L. If $\mathbf{x}$ is the coordinate vector of $\mathbf{v}$ with respect to E, then the coordinate vector of $L(\mathbf{v})$ with respect to F is given by

$$[L(\mathbf{v})]_F = A\mathbf{x}$$

The procedure for determining the matrix representation A is essentially the same as before. For $j = 1,\cdots,n$, let $\mathbf{a}_j = (a_{1j}, a_{2j}, \cdots, a_{mj})^T$ be the coordinate vector of $L(\mathbf{v}_j)$ with respect to $\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_m\}$; that is,

$$L(\mathbf{v}_j) = a_{1j}\mathbf{w}_1 + a_{2j}\mathbf{w}_2 + \cdots + a_{mj}\mathbf{w}_m \quad 1 \le j \le n$$

Let $A = (a_{ij}) = (\mathbf{a}_1, \cdots, \mathbf{a}_n)$. If

$$\mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_n\mathbf{v}_n$$

then

$$\begin{aligned} L(\mathbf{v}) &= L\left(\sum_{j=1}^{n} x_j\mathbf{v}_j\right) \\ &= \sum_{j=1}^{n} x_j L(\mathbf{v}_j) \\ &= \sum_{j=1}^{n} x_j\left(\sum_{i=1}^{m} a_{ij}\mathbf{w}_i\right) \end{aligned}$$

$$= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j \right) \mathbf{w}_i$$

For $i = 1, \cdots, m$, let

$$y_1 = \sum_{j=1}^{n} a_{ij} x_j$$

Thus,

$$\mathbf{y} = (y_1, y_2, \cdots, y_m)^T = A\mathbf{x}$$

is the coordinate vector of $L(\mathbf{v})$ with respect to $\{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_m\}$. We have established the following theorem:

Theorem 4.2.2 Matrix Representation Theorem

If $E = \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ and $F = \{\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_m\}$ are ordered bases for vector spaces ***V* and *W***, respectively, then, corresponding to each linear transformation $L: V \to W$, there is an $m \times n$ matrix A such that

$$[L(\mathbf{v})]_F = A[\mathbf{v}]_E \quad \text{for each } \mathbf{v} \in V$$

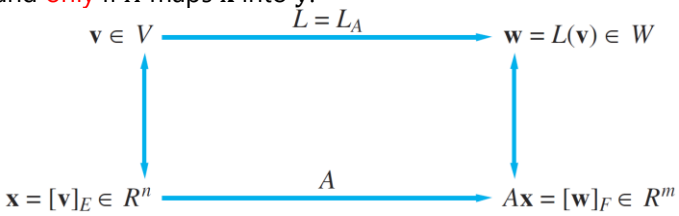
A is the matrix representing L relative to the ordered bases E and F. In fact,

$$\mathbf{a}_j = \left[L(\mathbf{v}_j)\right]_F \quad j = 1,2, \cdots, n$$

Theorem 4.2.2 is illustrated in the below Figure. If A is the matrix representing L with respect to the bases E and F, and if

$$\mathbf{x} = [\mathbf{v}]_E \quad \text{(the coordinate vector of } \mathbf{v} \text{ with respect to } E)$$
$$\mathbf{y} = [\mathbf{w}]_F \quad \text{(the coordinate vector of } \mathbf{w} \text{ with respect to } F)$$

then L maps $\mathbf{v}$ into $\mathbf{w}$ if and only if A maps $\mathbf{x}$ into $\mathbf{y}$.



Theorem 4.2.3

Let $E = \{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ and $F = \{\mathbf{b}_1, \cdots, \mathbf{b}_m\}$ be ordered bases for R^n and R^m, respectively. If $L: \mathrm{R}^n \to \mathrm{R}^m$ is a linear transformation and A is the matrix representing L with respect to E and F, then

$$\mathbf{a}_j = B^{-1} L(\mathbf{u}_j) \quad \text{for } j = 1, \cdots, n$$

where $B = (\mathbf{b}_1, \cdots, \mathbf{b}_m)$.

Corollary 4.2.4

If A is the matrix representing the linear transformation $L: \mathrm{R}^n \to \mathrm{R}^m$ with respect to the bases

$$E = \{\mathbf{u}_1, \cdots, \mathbf{u}_n\} \quad \text{and} \quad F = \{\mathbf{b}_1, \cdots, \mathbf{b}_m\}$$

then the reduced row echelon form of $(\mathbf{b}_1, \cdots, \mathbf{b}_m \mid L(\mathbf{u}_1), \cdots, L(\mathbf{u}_n))$ is $(I \mid A)$.

Note:

Let $\mathbf{v} \in V$, $\mathbf{w} \in W$ and $\mathbf{v} = a\mathbf{v}_1 + b\mathbf{v}_2$, $\mathbf{w} = c\mathbf{w}_1 + d\mathbf{w}_2$. If $\mathbf{w} = L(\mathbf{v})$, then (condition is same as above)

$$c\mathbf{w}_1 + d\mathbf{w}_2 = L(a\mathbf{v}_1 + b\mathbf{v}_2)$$

We can rearrange it and get

$$(\mathbf{w}_1 \quad \mathbf{w}_2) \begin{bmatrix} c \\ d \end{bmatrix} = aL(\mathbf{v}_1) + bL(\mathbf{v}_2)$$
$$= (L(\mathbf{v}_1) \quad L(\mathbf{v}_2)) \begin{bmatrix} a \\ b \end{bmatrix}$$

Moreover,

$$\begin{bmatrix} c \\ d \end{bmatrix} = (\mathbf{w}_1 \quad \mathbf{w}_2)^{-1} (L(\mathbf{v}_1) \quad L(\mathbf{v}_2)) \begin{bmatrix} a \\ b \end{bmatrix}$$

So we get $(\mathbf{w}_1 \quad \mathbf{w}_2)^{-1}(L(\mathbf{v}_1) \quad L(\mathbf{v}_2)) = B^{-1}(L(\mathbf{v}_1) \quad L(\mathbf{v}_2)) = A$.

APPLICATION 1: Computer Graphics and Animation (Page 185)

Homogeneous Coordinates (Page 187)

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \leftrightarrow \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}$$

APPLICATION 2: The Yaw, Pitch, and Roll of an Airplane

The terms *yaw*, *pitch*, and *roll* are commonly used in the aerospace industry to describe the maneuvering of an aircraft. Figure 4.1(a) shows the initial position of a model airplane. In describing *yaw*, *pitch*, and *roll*, the current coordinate system is given in terms of the position of the vehicle. It is always assumed that the craft is situated on the xy-plane with its nose pointing in the direction of the positive x-axis and the left wing pointing in the direction of the positive y-axis. Furthermore, when the plane moves, the three coordinate axes move with the vehicle (see Figure 4.1).

A *yaw* is a rotation in the xy-plane. Figure 4.1(b) illustrates a yaw of $45°$ In this case, the craft has been rotated $45°$. to the right (clockwise). Viewed as a linear transformation in 3-space, a yaw is simply a rotation about the z-axis. Note that if the initial coordinates of the nose of the model plane are represented by the vector $(1,0,0)$, then its xyz coordinates after the yaw transformation will still be $(1,0,0)$, since the coordinate axis rotated with the craft. In the initial position of the airplane, the x, y, and z axes are in the same directions as the front-back, left-right, and top-bottom axes shown in the figure. We will refer to this initial front, left, top axis system as the FLT axis system. After the $45°$ yaw, the position of the nose of the craft with respect to the FLT axis system is $\left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0\right)$.

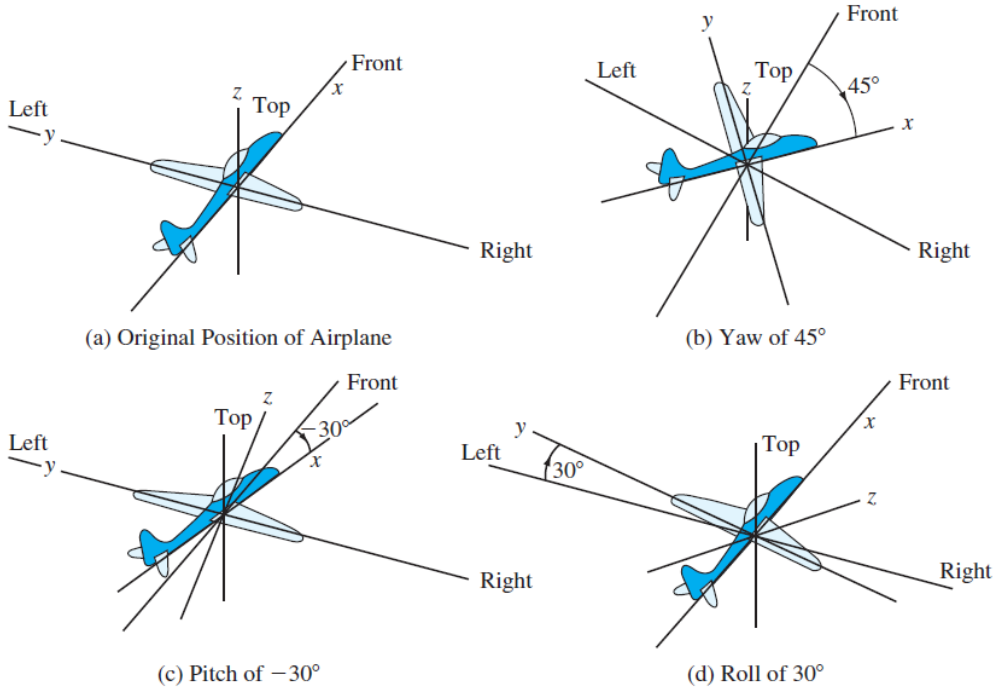


Figure 4.1

If we view a yaw transformation L in terms of the FLT axis system, it is easy to find a matrix representation. If L corresponds to yaw by an angle u, then L will rotate the points $(1,0,0)$ and $(0,1,0)$ to the positions $(\cos u\,, -\sin u\,, 0)$ and $(\sin u\,, \cos u\,, 0)$, respectively. The point $(0,0,1)$ will remained unchanged by the yaw since it is on the axis of rotation. In terms of column vectors, if $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$ are the images of the standard basis vectors for R^3 under L, then

$$\mathbf{y}_1 = L(\mathbf{e}_1) = \begin{bmatrix} \cos u \\ -\sin u \\ 0 \end{bmatrix}, \quad \mathbf{y}_2 = L(\mathbf{e}_2) = \begin{bmatrix} \sin u \\ \cos u \\ 0 \end{bmatrix}, \quad \mathbf{y}_3 = L(\mathbf{e}_3) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Therefore, the matrix representation of the yaw transformation is

$$Y = \begin{bmatrix} \cos u & \sin u & 0 \\ -\sin u & \cos u & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

A *pitch* is a rotation of the aircraft in the xz plane. Figure 4.1(c) illustrates a pitch of $-30°$. Since the angle is negative, the nose of the craft is rotated $30°$ downward, toward the bottom axis of the figure. Viewed as a linear transformation in 3-space, a pitch is simply a rotation about the y-axis. As with the yaw, we can find the matrix for a pitch transformation with respect to the FLT axis system. If L is a pitch transformation with angle of rotation v, the matrix representation of L is given by

$$P = \begin{bmatrix} \cos v & 0 & -\sin v \\ 0 & 1 & 0 \\ \sin v & 0 & \cos v \end{bmatrix}$$

A *roll* is a rotation of the aircraft in the yz plane. Figure 4.1(d) illustrates a roll of $30°$. In this case the left wing is rotated up $30°$ toward the top axis in the figure and the right wing is rotated $30°$ downward toward the bottom axis. Viewed as a linear transformation in 3-space, a roll is simply a rotation about the x-axis. As with the yaw and pitch, we can find the matrix representation for a roll transformation with respect to the FLT axis system. If L is a roll transformation with angle of rotation w, the matrix representation of L is given by

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos w & -\sin w \\ 0 & \sin w & \cos w \end{bmatrix}$$

If we perform a yaw by an angle u and then a pitch by an angle v, the composite transformation is linear; however, its matrix representation is *not* equal to the product PY. The effect of the yaw on the standard basis vectors $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ is to rotate them to the new directions $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$. So the vectors $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$ will define the directions of the x, y, and z axes when we do the pitch. The desired pitch transformation is then a rotation about the new y-axis (i.e., the axis in the direction of the vector $\mathbf{y}_2$). The vectors $\mathbf{y}_1$ and $\mathbf{y}_3$ form a plane, and when the pitch is applied, they are both rotated by an angle v in that plane. The vector $\mathbf{y}_2$ will remain unaffected by the pitch, since it lies on the axis of rotation. Thus, the composite transformation L has the following effect on the standard basis vectors.

$$\mathbf{e}_1 \overset{\text{yaw}}{\rightarrow} \mathbf{y}_1 \overset{\text{pitch}}{\rightarrow} \cos v\, \mathbf{y}_1 + \sin v\, \mathbf{y}_3$$
$$\mathbf{e}_2 \overset{\text{yaw}}{\rightarrow} \mathbf{y}_2 \overset{\text{pitch}}{\rightarrow} \mathbf{y}_2$$
$$\mathbf{e}_3 \overset{\text{yaw}}{\rightarrow} \mathbf{y}_3 \overset{\text{pitch}}{\rightarrow} -\sin v\, \mathbf{y}_1 + \cos v\, \mathbf{y}_3$$

The images of the standard basis vectors form the columns of the matrix representing the composite transformation:

$$(\cos v\, \mathbf{y}_1 + \sin v\, \mathbf{y}_3, \mathbf{y}_2, -\sin v\, \mathbf{y}_1 + \cos v\, \mathbf{y}_3) = (\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3) \begin{bmatrix} \cos v & 0 & -\sin v \\ 0 & 1 & 0 \\ \sin v & 0 & \cos v \end{bmatrix}$$
$$= YP$$

It follows that matrix representation of the composite is a product of the two individual matrices representing the yaw and the pitch, but the product must be taken in the reverse order, with the yaw matrix Y on the left and the pitch matrix P on the right. Similarly, for a composite transformation of a yaw with angle u, followed by a pitch with angle v, and then a roll with angle w, the matrix representation of the composite transformation would be the product YPR.

EXERCISES

12. Let Y, P, and R be the yaw, pitch, and roll matrices given in APPLICATION 2, respectively, and let $Q = YPR$.

a. Y, P, and R all have determinants equal to 1.

b. The matrix Y represents a yaw with angle u. The inverse transformation should be a yaw with angle $-u$. The matrix representation of the inverse transformation is Y^T and that $Y^T = Y^{-1}$.

c. Q is nonsingular and Q^{-1} can be expressed in terms of the transposes of Y, P, and R.

20. Let V and W be vector spaces with ordered bases E and F, respectively. If L: $V \to W$ is a linear transformation and A is the matrix representing L relative to E and F, show that

a. $\mathbf{v} \in \ker(L)$ if and only if $[\mathbf{v}]_E \in N(A)$.

b. $\mathbf{w} \in L(V)$ if and only if $[\mathbf{w}]_F$ is in the column space of A.

4.3 Similarity

If L is a linear operator on an n-dimensional vector space V, the matrix representation of L will depend on the ordered basis chosen for V. By using different bases, it is possible to represent L by different $n \times n$ matrices.

Let us begin by considering an example in R^2. Let L be the linear transformation mapping R^2 into itself defined by

$$L(\mathbf{x}) = (2x_1, x_1 + x_2)^T$$

Since

$$L(\mathbf{e}_1) = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad \text{and} \quad L(\mathbf{e}_2) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

it follows that the matrix representing L with respect to $\{\mathbf{e}_1, \mathbf{e}_2\}$ is

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

If we use a different basis for R^2, the matrix representation of L will change. If, for example, we use

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{u}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

for a basis, then to determine the matrix representation of L with respect to $\{\mathbf{u}_1, \mathbf{u}_2\}$ we must determine $L(\mathbf{u}_1)$ and $L(\mathbf{u}_2)$ and express these vectors as linear combinations of $\mathbf{u}_1$ and $\mathbf{u}_2$. We can use the matrix A to determine $L(\mathbf{u}_1)$ and $L(\mathbf{u}_2)$:

$$L(\mathbf{u}_1) = A\mathbf{u}_1 = \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$L(\mathbf{u}_2) = A\mathbf{u}_2 = \begin{bmatrix} 2 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$$

To express these vectors in terms of $\mathbf{u}_1$ and $\mathbf{u}_2$, we use a transition matrix to change from the ordered basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ to $\{\mathbf{u}_1, \mathbf{u}_2\}$. Let us first compute the transition matrix **from** $\{\mathbf{u}_1, \mathbf{u}_2\}$ **to** $\{\mathbf{e}_1, \mathbf{e}_2\}$. This is simply

$$U = (\mathbf{u}_1, \mathbf{u}_2) = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

The transition matrix from $\{\mathbf{e}_1, \mathbf{e}_2\}$ to $\{\mathbf{u}_1, \mathbf{u}_2\}$ will then be

$$U^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

To determine the coordinates of $L(\mathbf{u}_1)$ and $L(\mathbf{u}_2)$ with respect to $\{\mathbf{u}_1, \mathbf{u}_2\}$, we multiply the vectors by U^{-1}:

$$U^{-1}L(\mathbf{u}_1) = U^{-1}A\mathbf{u}_1 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$U^{-1}L(\mathbf{u}_2) = U^{-1}A\mathbf{u}_2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} -2 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Thus,

$$L(\mathbf{u}_1) = 2\mathbf{u}_1 + 0\mathbf{u}_2$$
$$L(\mathbf{u}_2) = -1\mathbf{u}_1 + 1\mathbf{u}_2$$

and the matrix representing L with respect to $\{\mathbf{u}_1, \mathbf{u}_2\}$ is

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

How are A and B related? Note that the columns of B are

$$\begin{bmatrix}2\\0\end{bmatrix} = U^{-1}A\mathbf{u}_1 \quad \text{and} \quad \begin{bmatrix}-1\\1\end{bmatrix} = U^{-1}A\mathbf{u}_2$$

Hence,

$$B = (U^{-1}A\mathbf{u}_1, U^{-1}A\mathbf{u}_2) = U^{-1}A(\mathbf{u}_1, \mathbf{u}_2) = U^{-1}AU$$

Thus, if

i. B is the matrix representing L with respect to $\{\mathbf{u}_1, \mathbf{u}_2\}$
ii. A is the matrix representing L with respect to $\{\mathbf{e}_1, \mathbf{e}_2\}$
iii. U is the transition matrix corresponding to the change of basis from $\{\mathbf{u}_1, \mathbf{u}_2\}$ to $\{\mathbf{e}_1, \mathbf{e}_2\}$

then

$$B = U^{-1}AU$$

We will show next that the same sort of relationship above will hold for any two matrix representations of a linear operator that maps an n-dimensional vector space into itself.

Note:

Let $\mathbf{v}, \mathbf{w} \in V$, and $\mathbf{v} = a_1\mathbf{u}_1 + b_1\mathbf{u}_2 = a_2\mathbf{e}_1 + b_2\mathbf{e}_2$, $\mathbf{w} = c\mathbf{u}_1 + d\mathbf{u}_2$. If $\mathbf{w} = L(\mathbf{v})$, then

$$c\mathbf{u}_1 + d\mathbf{u}_2 = L(a_1\mathbf{u}_1 + b_1\mathbf{u}_2)$$

We can rearrange it and get

$$\begin{aligned}(\mathbf{u}_1 \quad \mathbf{u}_2)\begin{bmatrix}c\\d\end{bmatrix} &= a_1L(\mathbf{u}_1) + b_1L(\mathbf{u}_2)\\ &= (L(\mathbf{u}_1) \quad L(\mathbf{u}_2))\begin{bmatrix}a_1\\b_1\end{bmatrix}\end{aligned}$$

Moreover,

$$\begin{bmatrix}c\\d\end{bmatrix} = (\mathbf{u}_1 \quad \mathbf{u}_2)^{-1}(L(\mathbf{u}_1) \quad L(\mathbf{u}_2))\begin{bmatrix}a_1\\b_1\end{bmatrix}$$

Since, we know

$$(\mathbf{u}_1 \quad \mathbf{u}_2)\begin{bmatrix}a_1\\b_1\end{bmatrix} = (\mathbf{e}_1 \quad \mathbf{e}_2)\begin{bmatrix}a_2\\b_2\end{bmatrix} = \begin{bmatrix}a_2\\b_2\end{bmatrix}$$

Similarly,

$$\begin{bmatrix}c\\d\end{bmatrix} = (\mathbf{u}_1 \quad \mathbf{u}_2)^{-1}(L(\mathbf{e}_1) \quad L(\mathbf{e}_2))\begin{bmatrix}a_2\\b_2\end{bmatrix}$$

$$= (\mathbf{u}_1 \quad \mathbf{u}_2)^{-1}(L(\mathbf{e}_1) \quad L(\mathbf{e}_2))(\mathbf{u}_1 \quad \mathbf{u}_2)\begin{bmatrix}a_1\\b_1\end{bmatrix}$$

Now we get $(\mathbf{u}_1 \quad \mathbf{u}_2)^{-1}(L(\mathbf{e}_1) \quad L(\mathbf{e}_2))(\mathbf{u}_1 \quad \mathbf{u}_2) = B = U^{-1}AU$.

Theorem 4.3.1

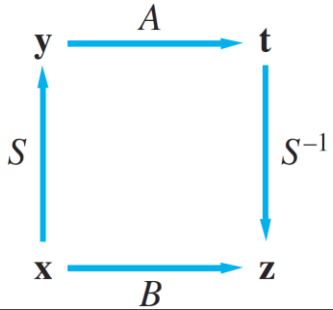
Let $E = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ and $F = \{\mathbf{w}_1, \cdots, \mathbf{w}_n\}$ be two ordered bases for **a** vector space V, and let L be a linear operator on V. Let S be the transition matrix representing the change from F to E. If A is the matrix representing L with respect to E, and B is the matrix representing L with respect to F, then $B = S^{-1}AS$.

Let $\mathbf{x}$ be any vector in R^n and let (Page 194)

$$\mathbf{v} = x_1\mathbf{w}_1 + x_2\mathbf{w}_2 + \cdots + x_n\mathbf{w}_n$$

Let

$$\mathbf{y} = S\mathbf{x}, \quad \mathbf{t} = A\mathbf{y}, \quad \mathbf{z} = B\mathbf{x}$$



Let A and B be $n \times n$ matrices. B is said to be **similar** to A if there exists a nonsingular matrix S such that $B = S^{-1}AS$.

Note that if B is similar to A, then $A = (S^{-1})^{-1}BS^{-1}$ is similar to B. Thus, we may simply say that A and B

are similar matrices.

EXAMPLE 1 (Page 195)

EXERCISES

11. if A and B are similar matrices, then $\det(A) = \det(B)$.

15. The *trace* of an $n \times n$ matrix A, denoted $\text{tr}(A)$, is the sum of its diagonal entries; that is,

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

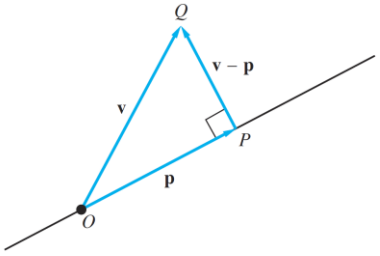
a. $\text{tr}(AB) = \text{tr}(BA)$

b. if A is similar to B, then $\text{tr}(A) = \text{tr}(B)$.

CHAPTER 5 Orthogonality

We **can add** to the **structure** of **a vector space** by **defining** a scalar or inner product. Such a product is not a true vector multiplication, **since** to every pair of vectors it associates **a scalar** rather than a third vector. For example, in R^2, we can define the scalar product of two vectors $\mathbf{x}$ and $\mathbf{y}$ to be $\mathbf{x}^T\mathbf{y}$.

We can think of **orthogonality** as a **generalization** of the concept of *perpendicularity* to **any vector space** with **an inner product**. To see the significance of this, consider the following problem: Let l be a line passing through the origin, and let Q be a point not on l. **Find** the point P on l that is **closest** to Q. The solution P to this problem is characterized by the condition that QP is perpendicular to OP (see below figure). If we think of the line l as corresponding to a subspace of R^2 and $\mathbf{v} = OQ$ as a vector in R^2, then the problem is to find a vector in the subspace that is "closest" to $\mathbf{v}$. The solution $\mathbf{p}$ will then be characterized by the property that $\mathbf{p}$ is orthogonal to $\mathbf{v} - \mathbf{p}$ (see below figure). In the setting of a vector space with an inner product, we are able to consider general *least squares problems*.



5.1 The Scalar Product in R^n

Two vectors $\mathbf{x}$ and $\mathbf{y}$ in R^n may be regarded as $n \times 1$ matrices. We can then form the matrix product $\mathbf{x}^T\mathbf{y}$. This product is a 1×1 matrix that may be regarded as a vector in R^1 or, more simply, as a real number. The product $\mathbf{x}^T\mathbf{y}$ is called the *scalar product* of $\mathbf{x}$ and $\mathbf{y}$. In particular, if $\mathbf{x} = (x_1, \cdots, x_n)^T$ and $\mathbf{y} = (y_1, \cdots, y_n)^T$, then

$$\mathbf{x}^T\mathbf{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n$$

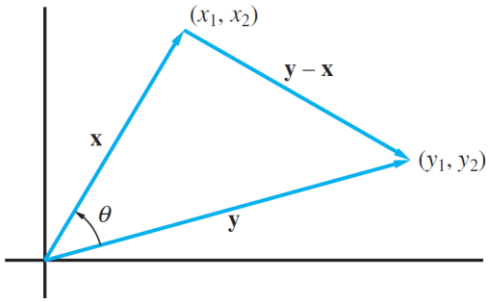
The Scalar Product in R^2 and R^3

In order to see the geometric significance of the scalar product, let us begin by restricting our attention to R^2 and R^3. Vectors in R^2 and R^3 can be represented by directed line segments. Given a vector $\mathbf{x}$ in either R^2 or R^3, its *Euclidean length* can be defined in terms of the scalar product

$$\|\mathbf{x}\| = (\mathbf{x}^T\mathbf{x})^{1/2} = \begin{cases} \sqrt{x_1^2 + x_2^2} & \text{if } \mathbf{x} \in \mathrm{R}^2 \\ \sqrt{x_1^2 + x_2^2 + x_3^2} & \text{if } \mathbf{x} \in \mathrm{R}^3 \end{cases}$$

Given two nonzero vectors $\mathbf{x}$ and $\mathbf{y}$, we can think of them as directed line segments starting at the same point. The angle between the two vectors is then defined as the angle θ between the line segments. We can measure the distance between the vectors by measuring the length of the vector joining the terminal point of $\mathbf{x}$ to the terminal point of $\mathbf{y}$. Thus we have the following definition

> Let $\mathbf{x}$ and $\mathbf{y}$ be vectors in either R^2 or R^3. The distance between $\mathbf{x}$ and $\mathbf{y}$ is defined to be the number $\|\mathbf{x} - \mathbf{y}\|$.



The angle between two vectors can be computed using the following theorem.

Theorem 5.1.1

If $\mathbf{x}$ and $\mathbf{y}$ are two nonzero vectors in either R^2 or R^3 and θ is the angle between them, then

$$\mathbf{x}^T\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\| \cos\theta$$

Proof

The vectors $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{y} - \mathbf{x}$ may be used to form a triangle as in the above figure. By the law of cosines, we have

$$\|\mathbf{y} - \mathbf{x}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\|\mathbf{x}\|\|\mathbf{y}\| \cos\theta$$

and hence it follows that

$$\begin{aligned}\|\mathbf{x}\|\|\mathbf{y}\| \cos\theta &= \frac{1}{2}(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - \|\mathbf{y} - \mathbf{x}\|^2) \\ &= \frac{1}{2}\left(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - (\mathbf{y} - \mathbf{x})^T(\mathbf{y} - \mathbf{x})\right) \\ &= \frac{1}{2}\left(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - (\mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{x} - \mathbf{x}^T\mathbf{y} + \mathbf{x}^T\mathbf{x})\right) \\ &= \mathbf{x}^T\mathbf{y}\end{aligned}$$

If $\mathbf{x}$ and $\mathbf{y}$ are nonzero vectors, then we can specify their directions by forming unit vectors

$$\mathbf{u} = \frac{1}{\|\mathbf{x}\|}\mathbf{x} \quad \text{and} \quad \mathbf{v} = \frac{1}{\|\mathbf{y}\|}\mathbf{y}$$

If θ is the angle between $\mathbf{x}$ and $\mathbf{y}$, then

$$\cos\theta = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \mathbf{u}^T\mathbf{v}$$

The cosine of the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$ is simply the scalar product of the corresponding direction vectors $\mathbf{u}$ and $\mathbf{v}$.

Corollary 5.1.2 Cauchy–Schwarz Inequality

If $\mathbf{x}$ and $\mathbf{y}$ are vectors in either R^2 or R^3, then

$$|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\|$$

with equality holding if and only if one of the vectors is $\mathbf{0}$ or one vector is a multiple of the other.

If $\mathbf{x}^T\mathbf{y} = \mathbf{0}$, it follows from Theorem 5.1.1 that either one of the vectors is the zero vector or $\cos\theta = 0$. If $\cos\theta = 0$, the angle between the vectors is a right angle.

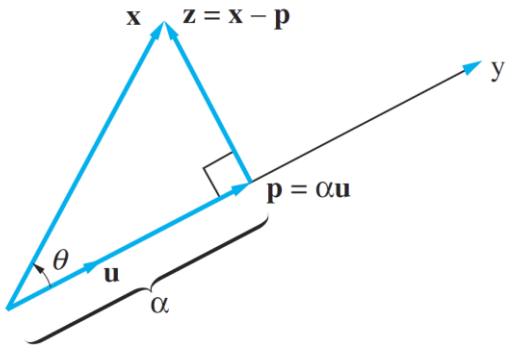
> The vectors $\mathbf{x}$ and $\mathbf{y}$ in R^2 (or R^3) are said to be **orthogonal** if $\mathbf{x}^T\mathbf{y} = 0$.

Scalar and Vector Projections

The scalar product can be used to find the component of one vector in the direction of another. Let $\mathbf{x}$ and $\mathbf{y}$ be nonzero vectors in either R^2 or R^3. We would like to write $\mathbf{x}$ as a sum of the form $\mathbf{p} + \mathbf{z}$, where $\mathbf{p}$ is in the direction of $\mathbf{y}$ and $\mathbf{z}$ is orthogonal to $\mathbf{p}$ (see below figure). To do this, let $\mathbf{u} = (1/\|\mathbf{y}\|)\mathbf{y}$. Thus $\mathbf{u}$ is a unit vector (length 1) in the direction of $\mathbf{y}$. We wish to find α such that $\mathbf{p} = \alpha\mathbf{u}$ is orthogonal to $\mathbf{z} = \mathbf{x} - \alpha\mathbf{u}$. For $\mathbf{p}$ and $\mathbf{z}$ to be orthogonal, the scalar α must satisfy $(\mathbf{p}^T\mathbf{z} = (\alpha\mathbf{u})^T(\mathbf{x} - \alpha\mathbf{u}) = 0)$

$$\begin{aligned}\alpha &= \|\mathbf{x}\| \cos\theta \\ &= \frac{\|\mathbf{x}\|\|\mathbf{y}\| \cos\theta}{\|\mathbf{y}\|}\end{aligned}$$

$$= \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{y}\|}$$



The scalar α is called the *scalar projection* of $\mathbf{x}$ onto $\mathbf{y}$, and the vector $\mathbf{p}$ is called the *vector projection* of $\mathbf{x}$ onto $\mathbf{y}$.

Scalar projection of $\mathbf{x}$ onto $\mathbf{y}$:

$$\alpha = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{y}\|}$$

Vector projection of $\mathbf{x}$ onto $\mathbf{y}$:

$$\mathbf{p} = \alpha\mathbf{u} = \alpha\frac{1}{\|\mathbf{y}\|}\mathbf{y} = \frac{\mathbf{x}^T\mathbf{y}}{\mathbf{y}^T\mathbf{y}}\mathbf{y}$$

Notation

If P_1 and P_2 are two points in 3-space, we will denote the vector from P_1 to P_2 by $\overrightarrow{P_1P_2}$

If $\mathbf{N}$ is a nonzero vector and P_0 is a fixed point, the set of points P such that $\overrightarrow{P_0P}$ is orthogonal to $\mathbf{N}$ forms a plane π in 3-space that passes through P_0. The vector $\mathbf{N}$ and the plane π are said to be ***normal*** to each other. A point $P = (x, y, z)$ will lie on π if and only if

$$\left(\overrightarrow{P_0P}\right)^T\mathbf{N} = 0$$

If $\mathbf{N} = (a, b, c)^T$ and $P_0 = (x_0, y_0, z_0)$, this equation can be written in the form

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

The span of two linearly independent vectors $\mathbf{x}$ and $\mathbf{y}$ in R^3 corresponds to a plane through the origin in 3-space. To determine the equation of the plane we must find a vector normal to the plane. In Section 3 of Chapter 2, it was shown that the cross product of the two vectors is orthogonal to each vector. If we take $\mathbf{N} = \mathbf{x} \times \mathbf{y} = (n_1, n_2, n_3)$ (Be careful for the sign) as our normal vector, then the equation of the plane is given by

$$n_1x + n_2y + n_3z = 0$$

EXAMPLE 7: Find the equation of the plane that passes through the points (Page 206)

$$P_1 = (1,1,2), \quad P_2 = (2,3,3), \quad P_3 = (3,-3,3)$$

EXAMPLE 8: Find the distance from the point $(2,0,0)$ to the plane $x + 2y + 2z = 0$. (Page 207)

If $\mathbf{x}$ and $\mathbf{y}$ are nonzero vectors in R^3 and θ is the angle between the vectors, then

$$\cos\theta = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}$$

It then follows that

$$\sin\theta = \sqrt{1 - \cos^2\theta} = \sqrt{1 - \frac{(\mathbf{x}^T\mathbf{y})^2}{\|\mathbf{x}\|^2\|\mathbf{y}\|^2}} = \frac{\sqrt{\|\mathbf{x}\|^2\|\mathbf{y}\|^2 - (\mathbf{x}^T\mathbf{y})^2}}{\|\mathbf{x}\|\|\mathbf{y}\|}$$

and hence

$$\|\mathbf{x}\|\|\mathbf{y}\|\sin\theta = \sqrt{\|\mathbf{x}\|^2\|\mathbf{y}\|^2 - (\mathbf{x}^T\mathbf{y})^2}$$
$$= \sqrt{(x_2y_3 - x_3y_2)^2 + (x_3y_1 - x_1y_3)^2 + (x_1y_2 - x_2y_1)^2}$$
$$= \|\mathbf{x} \times \mathbf{y}\|$$

Thus, we have, for any nonzero vectors $\mathbf{x}$ and $\mathbf{y}$ in R^3,

$$\|\mathbf{x} \times \mathbf{y}\| = \|\mathbf{x}\|\|\mathbf{y}\| \sin\theta$$

If either $\mathbf{x}$ or $\mathbf{y}$ is the zero vector then $\mathbf{x} \times \mathbf{y} = \mathbf{0}$ and hence the norm of $\mathbf{x} \times \mathbf{y}$ will be 0.

Orthogonality in R^n

If $\mathbf{x} \in \mathrm{R}^n$, then the *Euclidean length* of $\mathbf{x}$ is defined by

$$\|\mathbf{x}\| = (\mathbf{x}^T\mathbf{x})^{1/2} = (x_1^2 + x_2^2 + \cdots + x_n^2)^{1/2}$$

If $\mathbf{x}$ and $\mathbf{y}$ are two vectors in R^n, then the distance between the vectors is $\|\mathbf{y} - \mathbf{x}\|$.

The Cauchy–Schwarz inequality holds in R^n. Consequently,

$$-1 \le \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} \le 1$$

for any nonzero vectors $\mathbf{x}$ and $\mathbf{y}$ in R^n. The angle θ between two nonzero vectors $\mathbf{x}$ and $\mathbf{y}$ in R^n is given by

$$\cos\theta = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}, \quad 0 \le \theta \le \pi$$

In talking about angles between vectors it is usually more convenient to scale the vectors so as to make them unit vectors. If we set

$$\mathbf{u} = \frac{1}{\|\mathbf{x}\|}\mathbf{x} \quad \text{and} \quad \mathbf{v} = \frac{1}{\|\mathbf{y}\|}\mathbf{y}$$

then the angle θ between $\mathbf{u}$ and $\mathbf{v}$ is clearly the same as the angle between $\mathbf{x}$ and $\mathbf{y}$, and its cosine can be computed simply by taking the scalar product of the two unit vectors:

$$\cos\theta = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \mathbf{u}^T\mathbf{v}$$

The vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be ***orthogonal*** if $\mathbf{x}^T\mathbf{y} = 0$. Often the symbol $\perp$ is used to indicate orthogonality. Thus, if $\mathbf{x}$ and $\mathbf{y}$ are orthogonal, we will write $\mathbf{x} \perp \mathbf{y}$. Vector and scalar projections are defined in R^n in the same way that they were defined for R^2.

If $\mathbf{x}$ and $\mathbf{y}$ are vectors in R^n, then

$$\|\mathbf{x} + \mathbf{y}\|^2 = (\mathbf{x} + \mathbf{y})^T(\mathbf{x} + \mathbf{y}) = \|\mathbf{x}\|^2 + 2\mathbf{x}^T\mathbf{y} + \|\mathbf{y}\|^2$$

In the case that $\mathbf{x}$ and $\mathbf{y}$ are orthogonal, above equation becomes the *Pythagorean law*

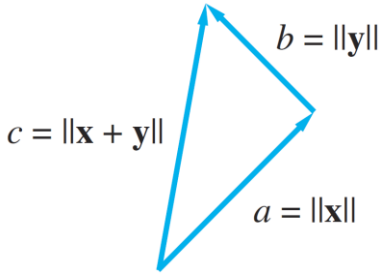
$$\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$$

The Pythagorean Law is a generalization of the Pythagorean theorem. When $\mathbf{x}$ and $\mathbf{y}$ are nonzero orthogonal vectors in R^2, we can use these vectors and their sum $\mathbf{x} + \mathbf{y}$ to form a right triangle. The Pythagorean law relates the lengths of the sides of the triangle. Indeed, if we set

$$a = \|\mathbf{x}\|, \quad b = \|\mathbf{y}\|, \quad c = \|\mathbf{x} + \mathbf{y}\|$$

then

$$c^2 = a^2 + b^2 \quad \text{(the famous Pythagorean theorem)}$$



Note:

$$\|\mathbf{x} + \mathbf{y}\|^2 = (\mathbf{x} + \mathbf{y})^T(\mathbf{x} + \mathbf{y}) = \|\mathbf{x}\|^2 + 2\mathbf{x}^T\mathbf{y} + \|\mathbf{y}\|^2 \le \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 = (\|\mathbf{x}\| + \|\mathbf{y}\|)^2$$

$$\|\mathbf{x} + \mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|$$

This inequality is also called the **triangle inequality**. The equality would hold when $\mathbf{x}^T\mathbf{y} = 0$, meaning when vectors $\mathbf{x}$ and $\mathbf{y}$ are colinear ($\cos\theta = 1$ or $\theta = 0$). **Geometric interpretation**: length of the sum of vectors is less or equal to the sum of lengths of individual vectors.

Moreover,

$$\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 - 2\mathbf{x}^T\mathbf{y} + \|\mathbf{y}\|^2$$

APPLICATION 1 Information Retrieval Revisited

APPLICATION 2 Statistics—Correlation and Covariance Matrices

To illustrate the importance of the correlation and covariance matrices, we will consider an application to the field of psychology

APPLICATION 3 Psychology—Factor Analysis and Principal Component Analysis

EXERCISES

4. Let $\mathbf{x}$ and $\mathbf{y}$ be linearly independent vectors in R^2. If $\|\mathbf{x}\| = 2$ and $\|\mathbf{y}\| = 3$, what, if anything, can we conclude about the possible values of $|\mathbf{x}^T\mathbf{y}|$?

Solution:

Since $\mathbf{x}$ and $\mathbf{y}$ be linearly independent vectors then $|\cos\theta| < 1$ and hence

$$|\mathbf{x}^T\mathbf{y}| = |\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta| < \|\mathbf{x}\|\|\mathbf{y}\| = 6$$

16. If $\mathbf{x}$ and $\mathbf{y}$ are linearly independent vectors in R^3, then they can be used to form a parallelogram P in the plane through the origin corresponding to $\mathrm{Span}(\mathbf{x},\mathbf{y})$. Show that

$$\text{Area of } P = \|\mathbf{x} \times \mathbf{y}\|$$

5.2 Orthogonal Subspaces

Let A be an $m \times n$ matrix and let $\mathbf{x} \in N(A)$, the null space of A. Since $A\mathbf{x} = \mathbf{0}$, we have

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = 0$$

for $i = 1, \cdots, m$. The above equation says that $\mathbf{x}$ is orthogonal to the ith column vector of A^T for $i = 1, \cdots, m$. Since $\mathbf{x}$ is orthogonal to each column vector of A^T, it is orthogonal to any linear combination of the column vectors of A^T. So if $\mathbf{y}$ is any vector in the column space of A^T, then $\mathbf{x}^T\mathbf{y} = 0$. Thus each vector in $N(A)$ is orthogonal to every vector in the column space of A^T. When two subspaces of R^n have this property, we say that they are orthogonal.

> Two subspaces X and Y of R^n are said to be **orthogonal** if $\mathbf{x}^T\mathbf{y} = 0$ for every $\mathbf{x} \in X$ and every $\mathbf{y} \in Y$. If X and Y are orthogonal, we write $X \perp Y$.

The concept of orthogonal subspaces does not always agree with our intuitive idea of perpendicularity. (Page 217)

> Let Y be a subspace of R^n. The set of all vectors in R^n that are orthogonal to every vector in Y will be denoted $Y^\perp$. Thus,
>
> $$Y^\perp = \{\mathbf{x} \in \mathrm{R}^n | \ \mathbf{x}^T\mathbf{y} = 0 \quad \text{for every} \quad \mathbf{y} \in Y\}$$
>
> The set $Y^\perp$ is called the **orthogonal complement** of Y.

Remarks

1. If X and Y are orthogonal subspaces of R^n, then $X \cap Y = \{\mathbf{0}\}$
2. If Y is a subspace of R^n, then $Y^\perp$ is also a subspace of R^n

Fundamental Subspaces

Let A be an $m \times n$ matrix. We saw in Chapter 3 that a vector $\mathbf{b} \in \mathrm{R}^m$ is in the column space of A if and only if $\mathbf{b} = A\mathbf{x}$ for some $\mathbf{x} \in \mathrm{R}^n$. If we think of A as a linear transformation mapping R^n into R^m, then the column space of A is the same as the range of A. Let us denote the range of A by $R(A)$. Thus

$$R(A) = \{\mathbf{b} \in \mathrm{R}^m \mid \mathbf{b} = A\mathbf{x} \quad \text{for some} \quad \mathbf{x} \in \mathrm{R}^n\}$$
$$= \text{the column space of } A$$

The column space of A^T, $R(A^T)$, is a subspace of R^n:

$$R(A^T) = \{\mathbf{y} \in \mathrm{R}^n \mid \mathbf{y} = A^T\mathbf{x} \quad \text{for some} \quad \mathbf{x} \in \mathrm{R}^m\}$$

The column space of $R(A^T)$ is **essentially** the **same** as the row space of A, except that it consists of vectors in R^n ($n \times 1$ matrices) rather than n-tuples. Thus, $\mathbf{y} \in R(A^T)$ if and only if $\mathbf{y}^T$ is in the row space of A. We have seen that $R(A^T) \perp N(A)$.

Theorem 5.2.1 Fundamental Subspaces Theorem

If A is an $m \times n$ matrix, then $N(A) = R(A^T)^\perp$ and $N(A^T) = R(A)^\perp$.

Theorem 5.2.2

If S is a subspace of R^n, then $\dim S + \dim S^{\perp} = n$. Furthermore, if $\{\mathbf{x}_1, \cdots, \mathbf{x}_r\}$ is a basis for S and $\{\mathbf{x}_{r+1}, \cdots, \mathbf{x}_n\}$ is a basis for $S^{\perp}$, then $\{\mathbf{x}_1, \cdots, \mathbf{x}_r, \mathbf{x}_{r+1}, \cdots, \mathbf{x}_n\}$ is a basis for R^n.

Given a subspace S of R^n, we will use Theorem 5.2.2 to prove that each $\mathbf{x} \in \mathrm{R}^n$ can be expressed uniquely as a sum $\mathbf{y} + \mathbf{z}$, where $\mathbf{y} \in S$ and $\mathbf{z} \in S^{\perp}$.

> If U and V are subspaces of a vector space W and each $\mathbf{w} \in W$ can be written uniquely as a sum $\mathbf{u} + \mathbf{v}$, where $\mathbf{u} \in U$ and $\mathbf{v} \in V$, then we say that W is a **direct sum** of U and V, and we write $W = U \oplus V$.

If S is a subspace of R^n, then

$$\mathrm{R}^n = S \oplus S^{\perp}$$

Theorem 5.2.4

If S is a subspace of R^n, then $(S^{\perp})^{\perp} = S$.

It follows from Theorem 5.2.4 that if T is the orthogonal complement of a subspace S, then S is the orthogonal complement of T, and we may say simply that S and T are orthogonal complements. In particular, it follows from Theorem 5.2.1 that $N(A)$ and $R(A^T)$ are orthogonal complements of each other and that $N(A^T)$ and $R(A)$ are orthogonal complements. Hence we may write

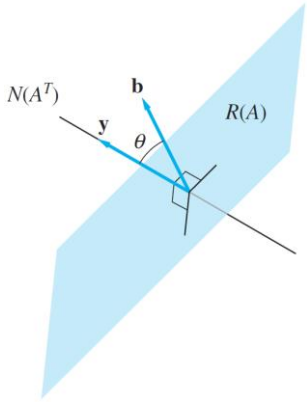
$$N(A)^{\perp} = R(A^T) \quad \text{and} \quad N(A^T)^{\perp} = R(A)$$

Recall that the system $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b} \in R(A)$. Since $R(A) = N(A^T)^{\perp}$, we have the following result,

Corollary 5.2.5

If A is an $m \times n$ matrix and $\mathbf{b} \in \mathrm{R}^m$, then **either** there is a vector $\mathbf{x} \in \mathrm{R}^n$ such that $A\mathbf{x} = \mathbf{b}$ **or** there is a vector $\mathbf{y} \in \mathrm{R}^m$ such that $A^T\mathbf{y} = \mathbf{0}$ and $\mathbf{y}^T\mathbf{b} \neq 0$.

The angle θ in the figure will be a right angle if and only if $\mathbf{b} \in R(A)$.



EXAMPLE 4: Let

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

Find the bases for $N(A), R(A^T), N(A^T)$, and $R(A)$. (Page 222)

We saw in Chapter 3 that the row space and the column space have the same dimension. If A has rank r, then

$$\dim R(A) = \dim R(A^T) = r$$

Actually, A can be used to establish a one-to-one correspondence between $R(A^T)$ and $R(A)$.

5.3 Least Squares Problems

A standard technique in mathematical and statistical modeling is to find a ***least squares*** fit to a set of data points in the plane. The least squares curve is usually the graph of a standard type of function, such as a linear function, a polynomial, or a trigonometric polynomial. Since the data may include errors in

measurement or experiment-related inaccuracies, we do not require the curve to pass through all the data points. Instead, we require the curve to provide an optimal approximation in the sense that the sum of squares of errors between the y values of the data points and the corresponding y values of the approximating curve are minimized.

APPLICATION 1 Astronomy—The Ceres Orbit of Gauss.

Least Squares Solutions of Overdetermined Systems

A least squares problem can generally be formulated as an overdetermined linear system of equations. Recall that an overdetermined system is one involving more equations than unknowns. Such systems are usually inconsistent. Thus, given an $m \times n$ system $A\mathbf{x} = \mathbf{b}$ with $m > n$, we cannot expect in general to find a vector $\mathbf{x} \in \mathrm{R}^n$ for which $A\mathbf{x}$ equals $\mathbf{b}$. Instead, we can look for a vector $\mathbf{x}$ for which $A\mathbf{x}$ is "closest" to $\mathbf{b}$. As you might expect, orthogonality plays an important role in finding such an $\mathbf{x}$.

If we are given a system of equations $A\mathbf{x} = \mathbf{b}$, where A is an $m \times n$ matrix with $m > n$ and $\mathbf{b} \in \mathrm{R}^m$, then, for each $\mathbf{x} \in \mathrm{R}^n$, we can form a ***residual***

$$r(\mathbf{x}) = \mathbf{b} - A\mathbf{x}$$

The distance between $\mathbf{b}$ and $A\mathbf{x}$ is given by

$$\|\mathbf{b} - A\mathbf{x}\| = \|r(\mathbf{x})\|$$

We wish to find a vector $\mathbf{x} \in \mathrm{R}^n$ for which $\|r(\mathbf{x})\|$ will be a minimum. Minimizing $\|r(\mathbf{x})\|$ is equivalent to minimizing $\|r(\mathbf{x})\|^2$. A vector $\hat{\mathbf{x}}$ that accomplishes this is said to be a *least squares solution* of the system $A\mathbf{x} = \mathbf{b}$.

If $\hat{\mathbf{x}}$ is a least squares solution of the system $A\mathbf{x} = \mathbf{b}$ and $\mathbf{p} = A\hat{\mathbf{x}}$, then $\mathbf{p}$ is a vector in the column space of A that is closest to $\mathbf{b}$. The next theorem guarantees that such a closest vector $\mathbf{p}$ not only exists, but is unique. Additionally, it provides an important characterization of the closest vector

Theorem 5.3.1

Let S be a subspace of R^m. For each $\mathbf{b} \in \mathrm{R}^m$, there is a unique element $\mathbf{p}$ of S that is closest to $\mathbf{b}$; that is,

$$\|\mathbf{b} - \mathbf{y}\| > \|\mathbf{b} - \mathbf{p}\|$$

for any $\mathbf{y} \neq \mathbf{p}$ in S. Furthermore, a given vector $\mathbf{p}$ in S will be closest to a given vector $\mathbf{b} \in \mathrm{R}^m$ if and only if $\mathbf{b} - \mathbf{p} \in S^{\perp}$.

In the special case that $\mathbf{b}$ is in the subspace S to begin with, we have

$$\mathbf{b} = \mathbf{p} + \mathbf{z}, \quad \mathbf{p} \in S, \quad \mathbf{z} \in S^{\perp}$$

and

$$\mathbf{b} = \mathbf{b} + \mathbf{0}$$

By the uniqueness of the direct sum representation,

$$\mathbf{p} = \mathbf{b} \quad \text{and} \quad \mathbf{z} = \mathbf{0}$$

A vector $\hat{\mathbf{x}}$ will be a solution of the least squares problem $A\mathbf{x} = \mathbf{b}$ if and only if $\mathbf{p} = A\hat{\mathbf{x}}$ is the vector in $R(A)$ that is closest to $\mathbf{b}$. The vector $\mathbf{p}$ is said to be the *projection of* $\mathbf{b}$ onto $R(A)$. It follows from Theorem 5.3.1 that

$$\mathbf{b} - \mathbf{p} = \mathbf{b} - A\hat{\mathbf{x}} = r(\hat{\mathbf{x}})$$

must be an element of $R(A)^{\perp}$. Thus, $\hat{\mathbf{x}}$ is a solution of the least squares problem if and only if

$$r(\hat{\mathbf{x}}) \in R(A)^{\perp}$$

(a) $\mathbf{b} \in R^2$ and A is a 2×1 matrix of rank 1. (b) $\mathbf{b} \in R^3$ and A is a 3×2 matrix of rank 2.

How do we find a vector $\hat{\mathbf{x}}$ satisfying this? The key to solving the least squares problem is provided by Theorem 5.2.1, which states that

$$R(A)^{\perp} = N(A^T)$$

A vector $\hat{\mathbf{x}}$ will be a least squares solution to the system $A\mathbf{x} = \mathbf{b}$ if and only if

$$r(\hat{\mathbf{x}}) \in N(A^T)$$

or, equivalently,

$$\mathbf{0} = A^T r(\hat{\mathbf{x}}) = A^T(\mathbf{b} - A\hat{\mathbf{x}})$$

Thus, to solve the least squares problem $A\mathbf{x} = \mathbf{b}$, we **must** solve

$$A^T A\mathbf{x} = A^T\mathbf{b}$$

which represents an $n \times n$ system of linear equations. These equations are called the ***normal equations***. In general, it is possible to have more than one solution of the normal equations; however, if $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are both solutions, then, since the projection $\mathbf{p}$ of $\mathbf{b}$ onto $R(A)$ is unique,

$$A\hat{\mathbf{x}} = A\hat{\mathbf{y}} = \mathbf{p}$$

The following theorem characterizes the conditions under which the least squares problem $A\mathbf{x} = \mathbf{b}$ will have a unique solution:

Theorem 5.3.2

If A is an $m \times n$ matrix of rank n, the normal equations

$$A^T A\mathbf{x} = A^T\mathbf{b}$$

have a unique solution

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T\mathbf{b}$$

and $\hat{\mathbf{x}}$ is the unique least squares solution of the system $A\mathbf{x} = \mathbf{b}$. ($(A^T A)$ is nonsingular)

The projection vector

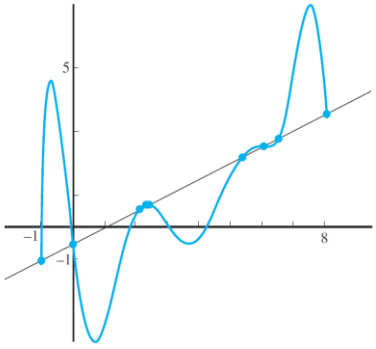
$$\mathbf{p} = A\hat{\mathbf{x}} = A(A^T A)^{-1} A^T\mathbf{b}$$

is the element of $R(A)$ that is closest to $\mathbf{b}$ in the least squares sense. The matrix $P = A(A^T A)^{-1} A^T$ is called the *projection matrix*.

APPLICATION 2 Spring Constants

EXAMPLE 1 (Page 229)

If the data consist of $n + 1$ points in the plane, it is possible to find a polynomial of degree n or less passing through all the points. Such a polynomial is called an *interpolating polynomial*. Actually, since the data usually involve experimental error, there is no reason to require that the function pass through all the points. Indeed, lower degree polynomials that do not pass through the points exactly usually give a truer description of the relationship between the variables



Given a table of data

x	x_1	x_2	$\cdots$	x_m
y	y_1	y_2	$\cdots$	y_m

we wish to find a linear function

$$y = c_0 + c_1 x$$

that best fits the data in the least squares sense. If we require that

$$y_i = c_0 + c_1 x_i \quad \text{for} \quad i = 1, \cdots, m$$

we get a system of m equations in two unknowns.

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

The linear function whose coefficients are the least squares solution above is said to be the best least squares fit to the data by a linear function.

EXAMPLE 2 (Page 230)

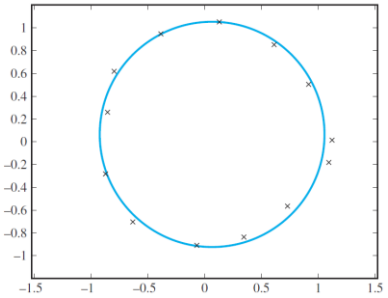
If the data do not resemble a linear function, we could use a higher degree polynomial. To find the coefficients $c_0, c_1, \cdots, c_n$ of the best least squares fit to the data

x	x_1	x_2	$\cdots$	x_m
y	y_1	y_2	$\cdots$	y_m

by a polynomial of degree n, we must find the least squares solution the system

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & & & & \\ 1 & x_m & x_m^2 & \cdots & x_m^n \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

APPLICATION 3 Coordinate Metrology



APPLICATION 4 Management Science: The Analytic Hierarchy Process Revisited

5.4 Inner Product Spaces

Scalar products are useful not only in R^n, but in a wide variety of contexts.

Definition and Examples

An **inner product** on a vector space V is an operation on V that assigns, to each pair of vectors $\mathbf{x}$ and $\mathbf{y}$ in V, a real number $\langle \mathbf{x}, \mathbf{y} \rangle$ satisfying the following conditions:

I. $\langle \mathbf{x}, \mathbf{y} \rangle \geq 0$ with equality if and only if $\mathbf{x} = \mathbf{0}$

II. $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ for all $\mathbf{x}$ and $\mathbf{y}$ in V

III. $\langle \alpha\mathbf{x} + \beta\mathbf{y}, \mathbf{z} \rangle = \alpha\langle \mathbf{x}, \mathbf{z} \rangle + \beta\langle \mathbf{y}, \mathbf{z} \rangle$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z}$ in V and all scalars α and β

A vector space V with an inner product is called an **inner product space**.

The Vector Space R^n

The standard inner product for R^n is the scalar product

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T\mathbf{y}$$

Given a vector $\mathbf{w}$ with positive entries, we could also define an inner product on R^n by

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i y_i w_i$$

The entries w_i are referred to as *weights*.

The Vector Space $\mathrm{R}^{m\times n}$

Given A and B in $\mathrm{R}^{m\times n}$, we can define an inner product by

$$\langle A, B \rangle = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ij}$$

The Vector Space $C[a,b]$

We may define an inner product on $C[a,b]$ by

$$\langle f,g\rangle = \int_a^b f(x)\,g(x)\,dx$$

Note that

$$\langle f,f\rangle = \int_a^b (f(x))^2\,dx \geq 0$$

If $\langle f,f\rangle = 0$, then $f(x)$ must be identically zero on $[a,b]$.

If $w(x)$ is a positive continuous function on $[a,b]$, then

$$\langle f,g\rangle = \int_a^b f(x)\,g(x)\,w(x)\,dx$$

also defines an inner product on $C[a,b]$. The function $w(x)$ is called a *weight function*.

The Vector Space P_n

Let $x_1, x_2, \cdots, x_n$ be distinct real numbers. For each pair of polynomials in P_n, define

$$\langle p,q\rangle = \sum_{i=1}^{n} p(x_i)\,q(x_i)$$

Note that

$$\langle p,p\rangle = \sum_{i=1}^{n} (p(x_i))^2 \geq 0$$

If $\langle p,p\rangle = 0$, then $x_1, x_2, \cdots, x_n$ must be roots of $p(x) = 0$. Since $p(x)$ is of degree less than n, it must be the zero polynomial.

If $w(x)$ is a positive function, then

$$\langle p,q\rangle = \sum_{i=1}^{n} p(x_i)\,q(x_i)\,w(x_i)$$

also defines an inner product on P_n.

Basic Properties of Inner Product Spaces

The results presented in Section 5.1 for scalar products in R^n all generalize to inner product spaces. In particular, if $\mathbf{v}$ is a vector in an inner product space V, the *length*, or ***norm*** of $\mathbf{v}$ is given by

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v},\mathbf{v}\rangle}$$

Two vectors $\mathbf{u}$ and $\mathbf{v}$ are said to be *orthogonal* if $\langle \mathbf{u},\mathbf{v}\rangle = 0$. As in R^n, a pair of orthogonal vectors will satisfy the Pythagorean law.

Theorem 5.4.1 The Pythagorean Law

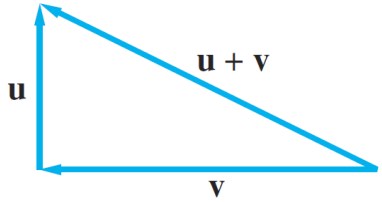
If $\mathbf{u}$ and $\mathbf{v}$ are orthogonal vectors in an inner product space V, then

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

Proof

$$\begin{aligned}\|\mathbf{u}+\mathbf{v}\|^2 &= \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v}\rangle \\ &= \langle \mathbf{u},\mathbf{u}\rangle + 2\langle \mathbf{u},\mathbf{v}\rangle + \langle \mathbf{v},\mathbf{v}\rangle \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2\end{aligned}$$

Interpreted in R^2, this is just the familiar Pythagorean theorem as shown below:



EXAMPLE 2: (Page 241)

For the vector space $C[-\pi,\pi]$, if we use a constant weight function $w(x) = 1/\pi$ to define an inner product

$$\langle f, g\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\, g(x)\, dx$$

This inner product plays a key role in Fourier analysis applications involving trigonometric approximation of functions.

For the vector space $\mathrm{R}^{m\times n}$ the norm derived from the inner product above is called the *Frobenius norm* and is denoted by $\|\cdot\|_F$. Thus, if $A \in \mathrm{R}^{m\times n}$, then

$$\|A\|_F = (\langle A, A\rangle)^{1/2} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2\right)^{1/2}$$

If $\mathbf{u}$ and $\mathbf{v}$ are vectors in an inner product space V and $\mathbf{v} \neq \mathbf{0}$, then the **scalar projection** of $\mathbf{u}$ onto $\mathbf{v}$ is given by

$$\alpha = \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{v}\|}$$

and the **vector projection** of $\mathbf{u}$ onto $\mathbf{v}$ is given by

$$\mathbf{p} = \alpha\left(\frac{1}{\|\mathbf{v}\|}\mathbf{v}\right) = \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}\mathbf{v}$$

Observations

If $\mathbf{v} \neq \mathbf{0}$ and $\mathbf{p}$ is the vector projection of $\mathbf{u}$ onto $\mathbf{v}$, then

I. $\mathbf{u} - \mathbf{p}$ and $\mathbf{p}$ are orthogonal.

II. $\mathbf{u} = \mathbf{p}$ if and only if $\mathbf{u}$ is a scalar multiple of $\mathbf{v}$.

Theorem 5.4.2 The Cauchy–Schwarz Inequality

If $\mathbf{u}$ and $\mathbf{v}$ are any two vectors in an inner product space V, then

$$|\langle \mathbf{u}, \mathbf{v}\rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\|$$

Equality holds if and only if $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent ($\mathbf{v} = \mathbf{0}$ or $\mathbf{u}$ is a multiple of $\mathbf{v}$).

One consequence of the Cauchy–Schwarz inequality is that if $\mathbf{u}$ and $\mathbf{v}$ are nonzero vectors, then

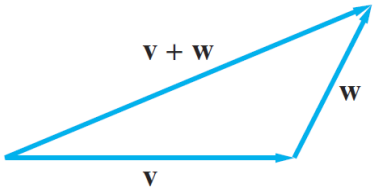
$$-1 \leq \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\|\|\mathbf{v}\|} \leq 1$$

and hence there is a unique angle θ in $[0, \pi]$ such that

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\|\|\mathbf{v}\|}$$

Norms

The word *norm* in mathematics has its own meaning that is independent of an inner product and its use here should be justified



A vector space V is said to be a **normed linear space** if, to each vector $\mathbf{v} \in V$, there is associated a real number $\|\mathbf{v}\|$, called the **norm** of $\mathbf{v}$, satisfying

I. $\|\mathbf{v}\| \geq 0$ with equality if and only if $\mathbf{v} = \mathbf{0}$.

II. $\|\alpha\mathbf{v}\| = |\alpha|\|\mathbf{v}\|$ for any scalar α.

III. $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$ for all $\mathbf{v}, \mathbf{w} \in V$.

The third condition is called the *triangle inequality*.

Theorem 5.4.3

If V is an inner product space, then the equation

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} \quad \text{for all} \quad \mathbf{v} \in V$$

defines a **norm** on V.

It is possible to define many different norms on a given vector space. For example, in R^n we could define

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$$

for every $\mathbf{x} = (x_1, x_2, \cdots, x_n)^T$. It is easily verified that $\|\cdot\|_1$ defines a norm on R^n. Another important norm on R^n is the *uniform norm* or *infinity norm*, which is defined by

$$\|\mathbf{x}\|_\infty = \max_{1 \le i \le n} |x_i|$$

More generally, we could define a norm on R^n by

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

for any real number $p \ge 1$. In particular, if $p = 2$, then

$$\|\mathbf{x}\|_2 = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2} = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$$

The norm $\|\cdot\|_2$ is the norm on R^n derived from the inner product. If $p \neq 2$, $\|\cdot\|_p$ does not correspond to any inner product. In the case of a norm that is not derived from an inner product, the Pythagorean law will not hold.

In general, a norm provides a way of measuring the distance between vectors.

> Let $\mathbf{x}$ and $\mathbf{y}$ be vectors in a normed linear space. The distance between $\mathbf{x}$ and $\mathbf{y}$ is defined to be the number $\|\mathbf{y} - \mathbf{x}\|$

Many applications involve finding a unique closest vector in a subspace S to a given vector $\mathbf{v}$ in a vector space V. If the norm used for V is derived from an inner product, then the closest vector can be computed as a vector projection of $\mathbf{v}$ onto the subspace S.

EXERCISES

9. In $C[-\pi, \pi]$ with inner product defined by $\langle f, g \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\, g(x)\, dx$ show that $\cos mx$ and $\sin nx$ are orthogonal and that both are unit vectors.

5.5 Orthonormal Sets

In R^2, it is generally more convenient to use the standard basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ than to use some other basis, such as $\{(2,1)^T, (3,5)^T\}$. For example, it would be easier to find the coordinates of $(x_1, x_2)^T$ with respect to the standard basis. The elements of the standard basis are orthogonal unit vectors. In working with an inner product space V, it is generally desirable to have a basis of mutually orthogonal unit vectors. Such a basis is convenient not only in finding coordinates of vectors, but also in solving least squares problems.

> Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ be nonzero vectors in an inner product space V. If $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0$ whenever $i \neq j$, then $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is said to be an **orthogonal set** of vectors.

Theorem 5.5.1

If $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is an orthogonal set of nonzero vectors in an inner product space V, then $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ are linearly independent.

Proof:

Suppose that $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ are mutually orthogonal nonzero vectors and

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$$

If $1 \le j \le n$, then, taking the inner product of $\mathbf{v}_j$ with both sides of equation above, we see that

$$c_1\langle \mathbf{v}_j, \mathbf{v}_1 \rangle + c_2\langle \mathbf{v}_j, \mathbf{v}_2 \rangle + \cdots + c_n\langle \mathbf{v}_j, \mathbf{v}_n \rangle = \mathbf{0}$$

$$c_j \|\mathbf{v}_j\|^2 = 0$$

and hence all the scalars $c_1, c_2, \cdots, c_n$ must be 0.

> An **orthonormal** set of vectors is an orthogonal set of unit vectors.

The set $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$ will be orthonormal if and only if

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \delta_{ij}$$

where

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

Given any orthogonal set of nonzero vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$, it is possible to form an orthonormal set by defining

$$\mathbf{u}_i = \left(\frac{1}{\|\mathbf{v}_i\|}\right)\mathbf{v}_i \quad \text{for } i = 1,2, \cdots, n$$

It follows from Theorem 5.5.1 that if $B = \{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k\}$ is an orthonormal set in an inner product space V, then B is a basis for the subspace $S = \text{Span}(\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k)$. We say that B is an ***orthonormal basis*** for S. It is generally much easier to work with an orthonormal basis than with an ordinary basis. In particular, it is much easier to calculate the coordinates of a given vector $\mathbf{v}$ with respect to an orthonormal basis. Once these coordinates have been determined, they can be used to compute $\|\mathbf{v}\|$.

Theorem 5.5.2

Let $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$ be an orthonormal basis for an inner product space V. If $\mathbf{v} = \sum_{i=1}^{n} c_i \mathbf{u}_i$, then $c_i = \langle \mathbf{v}, \mathbf{u}_i \rangle$.

Corollary 5.5.3

Let $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$ be an orthonormal basis for an inner product space V. If $\mathbf{u} = \sum_{i=1}^{n} a_i \mathbf{u}_i$ and $\mathbf{v} = \sum_{i=1}^{n} b_i \mathbf{u}_i$, then

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{n} a_i b_i$$

Corollary 5.5.4 Parseval's Formula

If $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ be an orthonormal basis for an inner product space V and $\mathbf{v} = \sum_{i=1}^{n} c_i \mathbf{u}_i$, then

$$\|\mathbf{v}\|^2 = \sum_{i=1}^{n} c_i^2 \left(= \sum_{i=1}^{n} \mathbf{v}_i^2 \right)$$

Orthogonal Matrices

Of particular importance are $n \times n$ matrices whose column vectors form an orthonormal set in R^n.

> An $n \times n$ matrix Q is said to be an **orthogonal matrix** if the column vectors of Q form an **orthonormal** set in R^n.

Theorem 5.5.5

An $n \times n$ matrix Q is orthogonal if and only if $Q^T Q = I$.

It follows from the theorem that if Q is an orthogonal matrix then Q is invertible and $Q^{-1} = Q^T$.

EXAMPLE 6: For any fixed θ, the matrix

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

is orthogonal and

$$Q^{-1} = Q^T = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

The matrix Q in Example 6 can be thought of as a linear transformation from R^2 onto R^2 that has the effect of rotating each vector by an angle θ while leaving the length of the vector unchanged. Similarly, Q^{-1} can be thought of as a rotation by the angle $-\theta$ (see Figure 5.1).

In general, inner products are preserved under multiplication by an orthogonal matrix [i.e., $\langle \mathbf{x}, \mathbf{y} \rangle = \langle Q\mathbf{x}, Q\mathbf{y} \rangle$]. Indeed,

$$\langle Q\mathbf{x}, Q\mathbf{y} \rangle = (Q\mathbf{y})^T Q\mathbf{x} = \mathbf{y}^T Q^T Q\mathbf{x} = \mathbf{y}^T\mathbf{x} = \langle \mathbf{x}, \mathbf{y} \rangle$$

In particular, if $\mathbf{x} = \mathbf{y}$, then $\|Q\mathbf{x}\|^2 = \|\mathbf{x}\|^2$ and hence $\|Q\mathbf{x}\| = \|\mathbf{x}\|$. Multiplication by an orthogonal matrix

preserves the lengths of vectors.

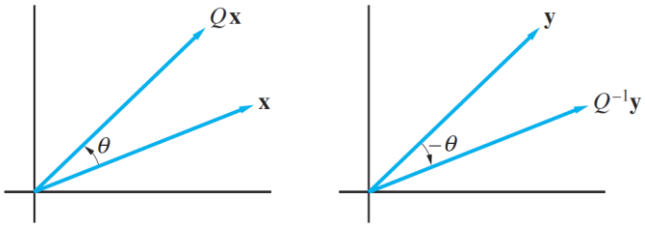


Figure 5. 1

Properties of Orthogonal Matrices
If Q is an $n \times n$ orthogonal matrix, then
(a) the column vectors of Q form an orthonormal basis for R^n.
(b) $Q^T Q = I$
(c) $Q^T = Q^{-1}$
(d) $\langle Q\mathbf{x}, Q\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$
(e) $\|Q\mathbf{x}\|_2 = \|\mathbf{x}\|_2$

Permutation Matrices

A *permutation matrix* is a matrix formed from the identity matrix by reordering its columns. Clearly, then, permutation matrices are orthogonal matrices. If P is the permutation matrix formed by reordering the columns of I in the order $(k_1, \cdots, k_n)$, then $P = \{\mathbf{e}_{k_1}, \cdots, \mathbf{e}_{k_n}\}$. If A is an $m \times n$ matrix, then

$$AP = (A\mathbf{e}_{k_1}, \cdots, A\mathbf{e}_{k_n}) = (\mathbf{a}_{k_1}, \cdots, \mathbf{a}_{k_n})$$

Postmultiplication of A by P reorders the columns of A in the order $(k_1, \cdots, k_n)$.
Since $P = \{\mathbf{e}_{k_1}, \cdots, \mathbf{e}_{k_n}\}$ is orthogonal, it follows that

$$P^{-1} = P^T = \begin{bmatrix} \mathbf{e}_{k_1}^T \\ \vdots \\ \mathbf{e}_{k_n}^T \end{bmatrix}$$

The k_1 column of P^T will be $\mathbf{e}_1$, the k_2 column will be $\mathbf{e}_2$, and so on. Thus, P^T is a permutation matrix. The matrix P^T can be formed directly from I by reordering its rows in the order $(k_1, k_2, \cdots, k_n)$. In general, a permutation matrix can be formed from I by reordering either its rows or its columns.
In general, if P is an $n \times n$ permutation matrix, premultiplication of an $n \times r$ matrix B by P reorders the rows of B and postmultiplication of an $m \times n$ matrix A by P reorders the columns of A.

Orthonormal Sets and Least Squares

Orthogonality plays an important role in solving least squares problems. Recall that if A is an $m \times n$ matrix of rank n, then the least squares problem $A\mathbf{x} = \mathbf{b}$ has a unique solution $\hat{\mathbf{x}}$ that is determined by solving the normal equations $A^T A\mathbf{x} = A^T\mathbf{b}$. The projection $\mathbf{p} = A\hat{\mathbf{x}}$ is the vector in $R(A)$ that is closest to $\mathbf{b}$. The least squares problem is especially easy to solve in the case where the column vectors of A form an orthonormal set in R^m.

Theorem 5.5.6

If the column vectors of A form an orthonormal **set** of vectors in R^m, then $A^T A = I$ and the solution to the least squares problem is

$$\hat{\mathbf{x}} = A^T\mathbf{b}$$

What if the columns of A are not orthonormal? In the next section we will learn a method for finding an orthonormal basis for $R(A)$. From this method we will obtain a factorization of A into a product QR, where Q has an orthonormal set of column vectors and R is upper triangular. With this factorization, the least squares problem is easily solved.
If we have an orthonormal basis for $R(A)$, the projection $\mathbf{p} = A\hat{\mathbf{x}}$ can be determined in terms of the basis elements. Indeed, this is a special case of the more general least squares problem of finding the element $\mathbf{p}$

in a subspace S of an inner product space V that is closest to a given element $\mathbf{x}$ in V. This problem is easily solved if S has an orthonormal basis.

Theorem 5.5.7

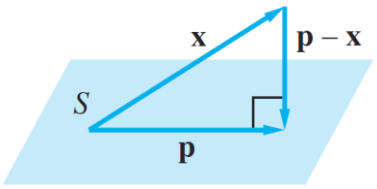
Let S be a subspace of an inner product space V and let $\mathbf{x} \in V$. Let $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$ be an orthonormal basis for S. If

$$\mathbf{p} = \sum_{i=1}^{n} c_i \mathbf{u}_i$$

where

$$c_i = \langle \mathbf{x}, \mathbf{u}_i \rangle \quad \text{for each } i$$

then $\mathbf{p} - \mathbf{x} \in S^{\perp}$.



Theorem 5.5.8

Under the hypothesis of Theorem 5.5.7, $\mathbf{p}$ is the element of S that is closest to $\mathbf{x}$; that is,

$$\|\mathbf{y} - \mathbf{x}\| > \|\mathbf{p} - \mathbf{x}\|$$

for any $\mathbf{y} \neq \mathbf{p}$ in S.

The vector $\mathbf{p}$ defined by Theorem 5.5.7 is said to be the *projection of* $\mathbf{x}$ *onto* S.

Corollary 5.5.9

Let S be a nonzero subspace of R^m and let $\mathbf{b} \in \mathrm{R}^m$. If $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k\}$ is an orthonormal *basis for* S *and* $U = (\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k)$, then the projection $\mathbf{p}$ of $\mathbf{b}$ onto S is given by

$$\mathbf{p} = UU^T\mathbf{b}$$

(For the least squares problem: $A^TA\mathbf{x} = A^T\mathbf{b} \rightarrow I\mathbf{x} = A^T\mathbf{b} \rightarrow A\mathbf{x} = AA^T\mathbf{b} \rightarrow \mathbf{p} = AA^T\mathbf{b}$)

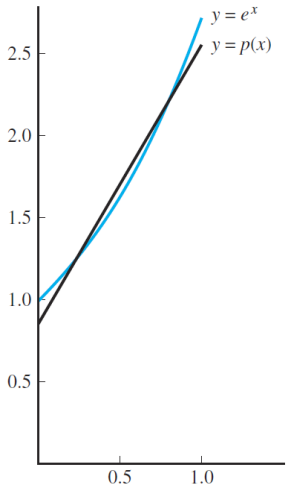
The matrix UU^T in Corollary 5.5.9 is the **projection matrix** corresponding to the subspace S of R^m. To project any vector $\mathbf{b} \in \mathrm{R}^m$ onto S, we need only find an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k\}$ for S, form the matrix UU^T, and then multiply UU^T times $\mathbf{b}$.

If P is a projection matrix corresponding to a subspace S of R^m, then, for any $\mathbf{b} \in \mathrm{R}^m$, the projection $\mathbf{p}$ of $\mathbf{b}$ onto S is unique.

Approximation of Functions

In many applications, it is necessary to approximate a continuous function in terms of functions from some special type of approximating set. Most commonly, we approximate by a polynomial of degree n or less. We can use Theorem 5.5.8 to obtain the best least squares approximation.

EXAMPLE 8: (Page 256)



Approximation by Trigonometric Polynomials

Trigonometric polynomials are used to approximate periodic functions. By a *trigonometric polynomial* of degree n we mean a function of the form

$$t_n(x) = \frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx)$$

We have already seen that the collection of functions (Page 249)

$$\frac{1}{\sqrt{2}}, \cos x, \cos 2x, \cdots, \cos nx$$

forms an orthonormal set with respect to the inner product (Page 249), and if the functions

$$\sin x, \sin 2x, \cdots, \sin nx$$

are **added** to the collection, it will still be an **orthonormal set**. Thus, we can use Theorem 5.5.8 to find the best least squares approximation to a continuous 2π periodic function $f(x)$ by a trigonometric polynomial of degree n or less. Note that

$$\langle f, \frac{1}{\sqrt{2}} \rangle \frac{1}{\sqrt{2}} = \langle f, 1 \rangle \frac{1}{2}$$

so that if

$$a_0 = \langle f, 1 \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\, dx$$

and

$$a_k = \langle f, \cos kx \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx\, dx$$

$$b_k = \langle f, \sin kx \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx\, dx$$

for $k = 1,2,\cdots,n$, then these coefficients determine the best least squares approximation to f. The a_k's and the b_k's turn out to be the well-known *Fourier coefficients* that occur in many applications involving trigonometric series approximations of functions.

Let us think of $f(x)$ as representing the position at time x of an object moving along a line, and let t_n be the Fourier approximation of degree n to f. If we set

$$r_k = \sqrt{a_k^2 + b_k^2} \quad \text{and} \quad \theta_k = \text{Tan}^{-1}\left(\frac{b_k}{a_k}\right)$$

then

$$a_k \cos kx + b_k \sin kx = r_k \left(\frac{a_k}{r_k} \cos kx + \frac{b_k}{r_k} \sin kx\right)$$
$$= r_k \cos(kx - \theta_k)$$

Thus, the motion $f(x)$ is being represented as a sum of simple harmonic motions.

For signal-processing applications, it is useful to express the trigonometric approximation in complex form. To this end, we define **complex** Fourier coefficients c_k in terms of the **real** Fourier coefficients a_k and b_k: (Can't represent them as an orthonormal set)

$$c_k = \frac{1}{2}(a_k - ib_k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,(\cos kx - i\sin kx)dx$$
$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\,e^{-ikx}dx \quad (k \geq 0)$$

The latter equality follows from the identity

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

We also define the coefficient c_{-k} to be the complex conjugate of c_k. Thus

$$c_{-k} = \overline{c_k} = \frac{1}{2}(a_k + ib_k) \quad (k \geq 0)$$

Alternatively, if we solve for a_k and b_k, then

$$a_k = c_k + c_{-k} \quad \text{and} \quad b_k = i(c_k - c_{-k})$$

From these identities, it follows that

$$c_k e^{ikx} + c_{-k}e^{-ikx} = (c_k + c_{-k})\cos kx + i(c_k - c_{-k})\sin kx$$
$$= a_k \cos kx + b_k \sin kx$$

and hence the trigonometric polynomial

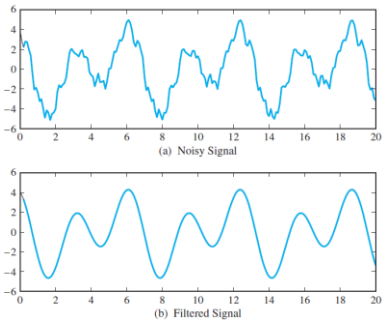
$$t_n(x) = \frac{a_0}{2} + \sum_{k=1}^{n}(a_k \cos kx + b_k \sin kx)$$

can be **rewritten** in complex form as

$$t_n(x) = \sum_{k=-n}^{n} c_k e^{ikx}$$

APPLICATION 1 Signal Processing

The Discrete Fourier Transform



The Fast Fourier Transform

EXERCISES

13. Let Q be an $n \times n$ orthogonal matrix. Use mathematical induction to prove each of the following.
    (a) $(Q^m)^{-1} = (Q^T)^m = (Q^m)^T$ for any positive integer m.
    (b) $\|Q^m\mathbf{x}\| = \|\mathbf{x}\|$ for any $\mathbf{x} \in R^n$.

14. Let $\mathbf{u}$ be a unit vector in R^n and let $H = I - 2\mathbf{u}\mathbf{u}^T$. Show that H is both orthogonal and symmetric and hence is its own inverse.

15. Let Q be an orthogonal matrix and let $d = \det(Q)$. Show that $|d| = 1$. (Use $\det(Q) = \det(Q^T)$)

5.6 The Gram–Schmidt Orthogonalization Process

In this section we learn a process for constructing an orthonormal basis for an n-dimensional inner product space V. The method involves **using projections** to transform an **ordinary** basis $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$ into an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$.

We will construct the $\mathbf{u}_i$'s so that

$$\text{Span}(\mathbf{u}_1, \cdots, \mathbf{u}_k) = \text{Span}(\mathbf{x}_1, \cdots, \mathbf{x}_k)$$

for $k = 1, \cdots, n$.

Theorem 5.6.1 The Gram–Schmidt Process

Let $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$ be a basis for the inner product space V. Let

$$\mathbf{u}_1 = \left(\frac{1}{\|\mathbf{x}_1\|}\right)\mathbf{x}_1$$

and define $\mathbf{u}_2, \cdots, \mathbf{u}_n$ recursively by

$$\mathbf{u}_{k+1} = \frac{1}{\|\mathbf{x}_{k+1} - \mathbf{p}_k\|}(\mathbf{x}_{k+1} - \mathbf{p}_k) \quad \text{for } k = 1, \cdots, n-1$$

where

$$\mathbf{p}_k = \langle \mathbf{x}_{k+1}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{x}_{k+1}, \mathbf{u}_2 \rangle \mathbf{u}_2 + \cdots + \langle \mathbf{x}_{k+1}, \mathbf{u}_k \rangle \mathbf{u}_k$$

is the projection of $\mathbf{x}_{k+1}$ onto $\text{Span}(\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k)$. Then the set

$$\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n\}$$

is an orthonormal basis for V.

EXAMPLE 2: (Page 268)

Theorem 5.6.2 Gram–Schmidt QR Factorization

If A is an $m \times n$ matrix of rank n, then A can be factored into a product QR, where Q is an $m \times n$ matrix with orthonormal column vectors and R is an upper triangular $n \times n$ matrix whose diagonal entries are all positive. [Note: R must be nonsingular since $\det(R) > 0$.] (Page 269)

$$A = (\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n)$$

$$Q = (\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n)$$

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{11} \\ \vdots & & & \\ 0 & 0 & \cdots & r_{nn} \end{bmatrix}$$

where $\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n\}$ be the orthonormal basis of $R(A)$ derived from the Gram–Schmidt process.

$$r_{ik} = \mathbf{q}_i^T \mathbf{a}_k = \langle \mathbf{q}_i, \mathbf{a}_k \rangle \quad \text{for } i = 1, \cdots, k \quad \text{and} \quad k = 1, \cdots, n$$

We saw in Section 5.5 that if the columns of an $m \times n$ matrix A form an orthonormal set, then the least squares solution of $A\mathbf{x} = \mathbf{b}$ is simply $\hat{\mathbf{x}} = A^T\mathbf{b}$. If A has rank n, but its column vectors do not form an orthonormal set in R^m, then the QR factorization can be used to solve the least squares problem.

Theorem 5.6.3

If A is an $m \times n$ matrix of rank n, then the least squares solution of $A\mathbf{x} = \mathbf{b}$ is given by $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$, where Q and R are the matrices obtained from the factorization given in Theorem 5.6.2. The solution $\hat{\mathbf{x}}$ may be obtained by using back substitution to solve $R\hat{\mathbf{x}} = Q^T\mathbf{b}$.

The Modified Gram–Schmidt Process

In Chapter 7 we will consider computer methods for solving least squares problems. The QR factorization method of Example 4 does not in general produce accurate results when carried out with finite-precision arithmetic. In practice, there may be a loss of orthogonality due to round off error in computing $\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n$. We can achieve better numerical accuracy using a modified version of the Gram–Schmidt method. (Page 273)

Algorithm 5.6.1 Modified Gram–Schmidt Process

For $k = 1, 2, \ldots, n$ *set*

$r_{kk} = \|\mathbf{a}_k\|$

$\mathbf{q}_k = \frac{1}{r_{kk}}\mathbf{a}_k$

For $j = k+1, k+2, \ldots, n$, *set*

$r_{kj} = \mathbf{q}_k^T\mathbf{a}_j$

$\mathbf{a}_j = \mathbf{a}_j - r_{kj}\mathbf{q}_k$

End for loop

End for loop

5.7 Orthogonal Polynomials (Unfinished)

Orthogonal Sequences

From the sequence $1, x, x^2, \cdots$, it is possible to construct an *orthonormal sequence* $p_0(x), p_1(x), \cdots$.

Let P be the vector space of all polynomials and define the inner product $\langle , \rangle$ on P by

$$\langle p, q \rangle = \int_a^b p(x)\, q(x)\, w(x)\, dx$$

where $w(x)$ is a positive continuous function. The interval can be taken as either open or closed and may be finite or infinite. If, however,

$$\int_a^b p(x)\, w(x)\, dx$$

is improper, we require that it converge for every $p \in P$.

> Let $p_0(x), p_1(x), \cdots$ be a sequence of polynomials with deg $p_i(x) = i$ for each i. If $\langle p_i(x), p_j(x) \rangle = 0$ whenever $i \neq j$, then $\{p_n(x)\}$ is said to be a **sequence of orthogonal polynomials**. If $\langle p_i, p_j \rangle = \delta_{ij}$, then $\{p_n(x)\}$ is said to be a **sequence of orthonormal polynomials**.

Theorem 5.7.1

If $p_0, p_1, \cdots$ is a sequence of orthogonal polynomials, then

I. $p_0, \cdots, p_{n-1}$ form a basis for P_n

II. $p_n \in P_n^{\perp}$ (i.e., p_n is orthogonal to every polynomial of degree less than n)

Similarly, if $f \in C[a, b]$, then the best least squares approximation to f by the elements of P_n is given by

$$p = \sum_{i=0}^{n-1} \frac{\langle f, p_i \rangle}{\langle p_i, p_i \rangle} p_i$$

where $p_0, p_1, \cdots, p_{n-1}$ are orthogonal polynomials

Theorem 5.7.2

Let $p_0, p_1, \cdots$ be a sequence of orthogonal polynomials. Let a_i denote the lead coefficient of p_i for each i, and define $p_{-1}(x)$ to be the zero polynomial. Then

$$\alpha_{n+1}\, p_{n+1}(x) = (x - \beta_{n+1})\, p_n(x) - \alpha_n \gamma_n\, p_{n-1}(x) \quad (n \geq 0)$$

where $\alpha_0 = \gamma_0 = 1$ and

$$\alpha_n = \frac{a_{n-1}}{a_n}, \quad \beta_n = \frac{\langle p_{n-1}, x p_{n-1} \rangle}{\langle p_{n-1}, p_{n-1} \rangle}, \quad \gamma_n = \frac{\langle p_n, p_n \rangle}{\langle p_{n-1}, p_{n-1} \rangle} \quad (n \geq 1)$$

Classical Orthogonal Polynomials (hard! Need second learn at Page 278)

CHAPTER 6 Eigenvalues

We will be concerned with the equation $A\mathbf{x} = \lambda\mathbf{x}$. This equation occurs in many applications of linear algebra. If the equation has a nonzero solution $\mathbf{x}$, then λ is said to be an *eigenvalue* of A and $\mathbf{x}$ is said to be an *eigenvector* belonging to λ.

In general, we can view eigenvalues as natural frequencies associated with linear operators. If A is an $n \times n$ matrix, we can think of A as representing a linear operator on R^n. Eigenvalues and eigenvectors provide the key to understanding how the operator works. For example, if $\lambda > 0$, the effect of the operator on any eigenvector belonging to λ is simply a stretching or shrinking by a constant factor. Indeed, the effect of the operator is easily determined on any linear combination of eigenvectors. In particular, if it is possible to find a basis of eigenvectors for R^n, the operator can be represented by a diagonal matrix D with respect to that basis and the matrix A can be factored into a product XDX^{-1}.

6.1 Eigenvalues and Eigenvectors

Many application problems involve applying a linear transformation repeatedly to a given vector. The key to solving these problems is to choose a coordinate system or basis that is in some sense natural for the operator and for which it will be simpler to do calculations involving the operator. With respect to these new basis vectors (*eigenvectors*) we associate scaling factors (*eigenvalues*) that represent the natural frequencies of the operator.

EXAMPLE 1: (Page 288)

In general, if a linear transformation is represented by an $n \times n$ matrix A and we can find a nonzero vector $\mathbf{x}$ so that $A\mathbf{x} = \lambda\mathbf{x}$, for some scalar λ, then, for this transformation, $\mathbf{x}$ is a natural choice to use as **a basis vector** for R^n and the scalar λ defines a natural frequency corresponding to that basis vector.

> Let A be an $n \times n$ matrix. A scalar λ is said to be an **eigenvalue** or a **characteristic value** of A if there exists a **nonzero** vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$. The vector $\mathbf{x}$ is said to be an **eigenvector** or a **characteristic vector** belonging to λ.

Actually, any nonzero multiple of $\mathbf{x}$ will be an eigenvector, because

$$A(\alpha\mathbf{x}) = \alpha A\mathbf{x} = \alpha\lambda\mathbf{x} = \lambda(\alpha\mathbf{x})$$

The equation $A\mathbf{x} = \lambda\mathbf{x}$ can be written in the form

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

Thus, λ is an eigenvalue of A **if and only if** the above equation has a nontrivial solution. The **set** of solutions to the above equation is $N(A - \lambda I)$, which is a subspace of R^n. Hence, if λ is an eigenvalue of A, then $N(A - \lambda I) \neq \{\mathbf{0}\}$, and **any nonzero** vector in $N(A - \lambda I)$ is **an eigenvector** belonging to λ. The **subspace** $N(A - \lambda I)$ is called the ***eigenspace*** **corresponding** to the eigenvalue λ.

Equation $(A - \lambda I)\mathbf{x} = \mathbf{0}$ will have a **nontrivial** solution **if and only if** $A - \lambda I$ is **singular**, or, **equivalently**,

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

If the determinant in the above equation is expanded, we obtain an nth-degree polynomial in the variable λ:

$$p(\lambda) = \det(A - \lambda I)$$

This polynomial is called the ***characteristic polynomial***, and the equation above is called the *characteristic equation*, for the matrix A. The roots of the characteristic polynomial are the eigenvalues of A. If roots are counted according to multiplicity, then the characteristic polynomial will have exactly n roots. Thus, A will have n eigenvalues, some of which may be repeated and some of which may be complex numbers. To take care of the latter case, it will be necessary to expand our field of scalars to the complex numbers and to allow complex entries for our vectors and matrices.

We have now established a number of equivalent conditions for λ to be an eigenvalue of A.

Let A be an $n \times n$ matrix and λ be a scalar. The following statements are equivalent:

(a) λ is an eigenvalue of A.
(b) $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has a **nontrivial** solution.
(c) $N(A - \lambda I) \neq \{\mathbf{0}\}$
(d) $A - \lambda I$ is singular.
(e) $\det(A - \lambda I) = 0$

EXAMPLE 4: Let

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

Find the eigenvalues and the corresponding eigenspaces.

Solution

$$\begin{vmatrix} 2-\lambda & -3 & 1 \\ 1 & -2-\lambda & 1 \\ 1 & -3 & 2-\lambda \end{vmatrix} = -\lambda(\lambda - 1)^2$$

Thus, the characteristic polynomial has roots $\lambda_1 = 0$, $\lambda_2 = \lambda_3 = 1$. The eigenspace corresponding to $\lambda_1 = 0$ is $N(A)$, which we determine in the usual manner:

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

Setting $x_3 = \alpha$, we find that $x_1 = x_2 = x_3 = \alpha$. Consequently, the **eigenspace** corresponding to $\lambda_1 = 0$ consists of all vectors of the form $\alpha(1,1,1)^T$. To find the eigenspace corresponding to $\lambda = 1$, we solve the system $(A - I)\mathbf{x} = \mathbf{0}$:

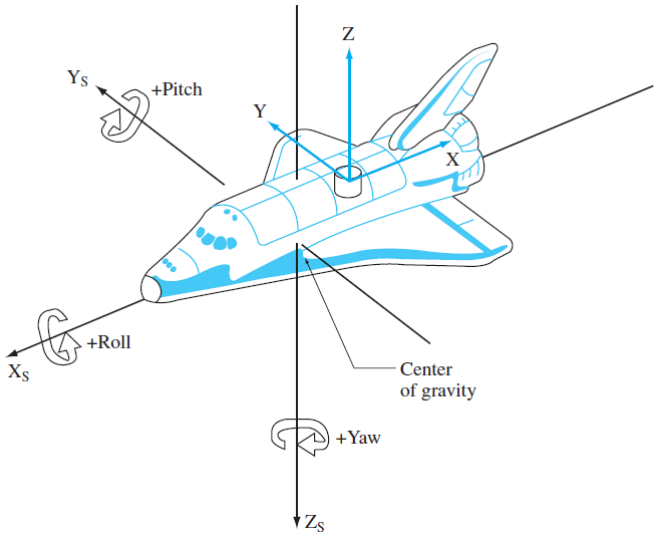
$$\left[\begin{array}{ccc|c} 1 & -3 & 1 & 0 \\ 1 & -3 & 1 & 0 \\ 1 & -3 & 1 & 0 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & -3 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Setting $x_2 = \alpha$ and $x_3 = \beta$, we get $x_1 = 3\alpha - \beta$. Thus, the **eigenspace** corresponding to $\lambda = 1$ consists of all vectors of the form

$$\begin{bmatrix} 3\alpha - \beta \\ \alpha \\ \beta \end{bmatrix} = \alpha \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

APPLICATION 1: Structures—Buckling of a Beam.

APPLICATION 2: Aerospace: The Orientation of a Space Shuttle (Page 294)



Complex Eigenvalues

If A is an $n \times n$ matrix with real entries, then the characteristic polynomial of A will have real coefficients, and hence all its complex roots must occur in conjugate pairs. Thus, if $\lambda = a + bi$ $(b \neq 0)$ is an eigenvalue of A, then $\bar{\lambda} = a - bi$ must also be an eigenvalue of A. Here the symbol $\bar{\lambda}$ (read *lambda bar*) is used to denote the complex conjugate of λ. A similar notation can be used for matrices. If $A = (a_{ij})$ is a matrix with complex entries, then $\bar{A} = (\overline{a_{ij}})$ is the matrix formed from A by conjugating each of its entries. We define a *real matrix* to be a matrix with the property that $\bar{A} = A$. In general, if A and B are matrices with complex entries and the multiplication AB is possible, then $\overline{AB} = \bar{A}\bar{B}$.

Not only do the complex eigenvalues of a real matrix occur in conjugate pairs, but so do the eigenvectors. Indeed, if λ is a complex eigenvalue of a real $n \times n$ matrix A and $\mathbf{z}$ is an eigenvector belonging to λ, then

$$A\bar{\mathbf{z}} = \bar{A}\bar{\mathbf{z}} = \overline{A\mathbf{z}} = \overline{\lambda\mathbf{z}} = \bar{\lambda}\bar{\mathbf{z}}$$

Thus, $\bar{\mathbf{z}}$ is an eigenvector of A belonging to $\bar{\lambda}$.

The Product and Sum of the Eigenvalues

It is easy to determine the product and sum of the eigenvalues of an $n \times n$ matrix A. If $p(\lambda)$ is the characteristic polynomial of A, then

$$p(\lambda) = \det(A - \lambda I) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & & a_{nn} - \lambda \end{vmatrix}$$

If $\lambda_1, \cdots, \lambda_n$ are the eigenvalues of A,

$$\lambda_1 \cdot \lambda_2 \cdots \lambda_n = p(0) = \det(A)$$

$$\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} a_{ii}$$

The sum of the diagonal elements of A is called the *trace* of A and is denoted by $\mathrm{tr}(A)$.

If the eigenvalues of A have been computed by some numerical method, one way to check their accuracy is to compare their sum with the trace of A.

Similar Matrices

We close this section with an important result about the eigenvalues of similar matrices. Recall that a matrix B is said to be *similar* to a matrix A if there exists a nonsingular matrix S such that $B = S^{-1}AS$.

Theorem 6.1.1

Let A and B be $n \times n$ matrices. If B is similar to A, then the two matrices have the same characteristic polynomial and, consequently, the same eigenvalues.

EXERCISES

6. Let λ be an eigenvalue of A and let $\mathbf{x}$ be an eigenvector belonging to λ. Use mathematical induction to show that, for $m \geq 1$, λ^m is an eigenvalue of A^m and $\mathbf{x}$ is an eigenvector of A^m belonging to λ^m.

12. Show that A and A^T have the same eigenvalues. Do they necessarily have the same eigenvectors? (No)

21. Let Q be an orthogonal matrix.

(a) Show that if λ is an eigenvalue of Q, then $|\lambda| = 1$. $(\langle \mathbf{x}, \mathbf{x} \rangle = \langle Q\mathbf{x}, Q\mathbf{x} \rangle = \langle \lambda\mathbf{x}, \lambda\mathbf{x} \rangle = \lambda^2 \langle \mathbf{x}, \mathbf{x} \rangle)$

(b) Show that $|\det(Q)| = 1$. $(\det(Q) = \lambda_1 \lambda_2 \cdots \lambda_n)$

6.2 Systems of Linear Differential Equations

Eigenvalues play an important role in the solution of systems of linear differential equations. In this section, we see how they are used in the solution of systems of linear differential equations with constant coefficients. We begin by considering systems of first-order equations of the form

$$\begin{aligned} y_1' &= a_{11}y_1 + a_{12}y_2 + \cdots + a_{1n}y_n \\ y_2' &= a_{21}y_1 + a_{22}y_2 + \cdots + a_{2n}y_n \\ &\vdots \end{aligned}$$

$$y_n' = a_{n1}y_1 + a_{n2}y_2 + \cdots + a_{nn}y_n$$

where $y_i = f_i(t)$ is a function in $C^1[a,b]$ for each i. If we let

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \text{and} \quad \mathbf{Y}' = \begin{bmatrix} y_1' \\ y_2' \\ \vdots \\ y_n' \end{bmatrix}$$

then the system can be written in the form

$$\mathbf{Y}' = A\mathbf{Y}$$

$\mathbf{Y}$ and $\mathbf{Y}'$ are both vector functions of t. Let us consider the simplest case first. When $n = 1$, the system is simply

$$y' = ay$$

Clearly, any function of the form

$$y(t) = ce^{at} \quad (c \text{ an arbitrary constant})$$

satisfies equation $y' = ay$. A natural generalization of this solution for the case $n > 1$ is to take

$$\mathbf{Y} = \begin{bmatrix} x_1 e^{\lambda t} \\ x_2 e^{\lambda t} \\ \vdots \\ x_n e^{\lambda t} \end{bmatrix} = e^{\lambda t}\mathbf{x}$$

where $\mathbf{x} = (x_1, x_2, \cdots, x_n)^T$. To verify that a vector function of this form does work, we compute the derivative

$$\mathbf{Y}' = \lambda e^{\lambda t}\mathbf{x} = \lambda\mathbf{Y}$$

Now, if we choose λ to be an eigenvalue of A and $\mathbf{x}$ to be an eigenvector belonging to λ, then

$$A\mathbf{Y} = e^{\lambda t}A\mathbf{x} = \lambda e^{\lambda t}\mathbf{x} = \lambda\mathbf{Y} = \mathbf{Y}'$$

Hence, $\mathbf{Y}$ is a solution of the system. Thus, if λ is an eigenvalue of A and $\mathbf{x}$ is an eigenvector belonging to λ, then $e^{\lambda t}\mathbf{x}$ is a solution of the system $\mathbf{Y}' = A\mathbf{Y}$. This will be true whether λ is real or complex. Note that if $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are both solutions of $\mathbf{Y}' = A\mathbf{Y}$, then $\alpha\mathbf{Y}_1 + \beta\mathbf{Y}_2$ is also a solution, since

$$\begin{aligned}(\alpha\mathbf{Y}_1 + \beta\mathbf{Y}_2)' &= \alpha\mathbf{Y}_1' + \beta\mathbf{Y}_2' \\ &= \alpha A\mathbf{Y}_1 + \beta A\mathbf{Y}_2 \\ &= A(\alpha\mathbf{Y}_1 + \beta\mathbf{Y}_2)\end{aligned}$$

It follows by induction that if $\mathbf{Y}_1, \cdots, \mathbf{Y}_n$ are solutions of $\mathbf{Y}' = A\mathbf{Y}$, then any linear combination $c_1\mathbf{Y}_1 + \cdots + c_n\mathbf{Y}_n$ will also be a solution.

In general, the solutions of an $n \times n$ first-order system of the form

$$\mathbf{Y}' = A\mathbf{Y}$$

will form an n-dimensional subspace of the vector space of all continuous vector-valued functions. If, in addition, we require that $\mathbf{Y}(t)$ take on a prescribed value $\mathbf{Y}_0$ when $t = 0$ then a standard theorem from differential equations guarantees that the problem will have a unique solution. A problem of the form

$$\mathbf{Y}' = A\mathbf{Y}, \quad \mathbf{Y}(0) = \mathbf{Y}_0$$

is called an *initial value problem*.

APPLICATION 1: Mixtures (Page 303)

Complex Eigenvalues

Let A be a real $n \times n$ matrix with a complex eigenvalue $\lambda = a + bi$, and let $\mathbf{x}$ be an eigenvector belonging to λ. The vector $\mathbf{x}$ can be split up into its real and imaginary parts.

$$\mathbf{x} = \begin{bmatrix} \operatorname{Re} x_1 + i \operatorname{Im} x_1 \\ \operatorname{Re} x_2 + i \operatorname{Im} x_2 \\ \vdots \\ \operatorname{Re} x_n + i \operatorname{Im} x_n \end{bmatrix} = \begin{bmatrix} \operatorname{Re} x_1 \\ \operatorname{Re} x_2 \\ \vdots \\ \operatorname{Re} x_n \end{bmatrix} + i \begin{bmatrix} \operatorname{Im} x_1 \\ \operatorname{Im} x_2 \\ \vdots \\ \operatorname{Im} x_n \end{bmatrix} = \operatorname{Re}\mathbf{x} + i \operatorname{Im}\mathbf{x}$$

Since the entries of A are all real, it follows that $\bar{\lambda} = a - bi$ is also an eigenvalue of A with eigenvector

$$\bar{\mathbf{x}} = \begin{bmatrix} \operatorname{Re} x_1 - i \operatorname{Im} x_1 \\ \operatorname{Re} x_2 - i \operatorname{Im} x_2 \\ \vdots \\ \operatorname{Re} x_n - i \operatorname{Im} x_n \end{bmatrix} = \operatorname{Re}\mathbf{x} - i \operatorname{Im}\mathbf{x}$$

and hence $e^{\lambda t}\mathbf{x}$ and $e^{\bar{\lambda} t}\bar{\mathbf{x}}$ are both solutions of the first-order system $\mathbf{Y}' = A\mathbf{Y}$. Any linear combination of

these two solutions will also be a solution. Thus, if we set

$$\mathbf{Y}_1 = \frac{1}{2}\left(e^{\lambda t}\mathbf{x} + e^{\bar{\lambda} t}\bar{\mathbf{x}}\right) = \mathrm{Re}\left(e^{\lambda t}\mathbf{x}\right)$$

and

$$\mathbf{Y}_2 = \frac{1}{2i}\left(e^{\lambda t}\mathbf{x} - e^{\bar{\lambda} t}\bar{\mathbf{x}}\right) = \mathrm{Im}\left(e^{\lambda t}\mathbf{x}\right)$$

then the vector functions $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are real-valued solutions of $\mathbf{Y}' = A\mathbf{Y}$. Taking the real and imaginary parts of

$$e^{\lambda t}\mathbf{x} = e^{(a+bi)t}\mathbf{x}$$
$$= e^{at}(\cos bt + i \sin bt)(\mathrm{Re}\,\mathbf{x} + i\,\mathrm{Im}\,\mathbf{x})$$

we see that

$$\mathbf{Y}_1 = e^{at}[(\cos bt)\,\mathrm{Re}\,\mathbf{x} - (\sin bt)\,\mathrm{Im}\,\mathbf{x}]$$
$$\mathbf{Y}_2 = e^{at}[(\cos bt)\,\mathrm{Im}\,\mathbf{x} + (\sin bt)\,\mathrm{Re}\,\mathbf{x}]$$

EXAMPLE 2: (Page 305)

If the $n \times n$ coefficient matrix A of the system $\mathbf{Y}' = A\mathbf{Y}$ has n linearly independent eigenvectors, the general solution can be obtained by the methods that have been presented. The case when A has fewer than n linearly independent eigenvectors is more complicated; consequently we will defer discussion of this case to Section 6.3.

Higher Order Systems

Given a second-order system of the form

$$\mathbf{Y}'' = A_1\mathbf{Y} + A_2\mathbf{Y}'$$

we may translate it into a first-order system by setting

$$y_{n+1}(t) = y_1'(t)$$
$$y_{n+2}(t) = y_2'(t)$$
$$\vdots$$
$$y_{2n}(t) = y_n'(t)$$

If we let

$$\mathbf{Y}_1 = \mathbf{Y} = (y_1, y_2, \cdots, y_n)^T$$

and

$$\mathbf{Y}_2 = \mathbf{Y}' = (y_{n+1}, \cdots, y_{2n})^T$$

then

$$\mathbf{Y}_1' = O\mathbf{Y}_1 + I\mathbf{Y}_2$$

and

$$\mathbf{Y}_2' = A_1\mathbf{Y}_1 + A_2\mathbf{Y}_2$$

The equations can be combined to give the $2n \times 2n$ first-order system

$$\begin{bmatrix}\mathbf{Y}_1' \\ \mathbf{Y}_2'\end{bmatrix} = \begin{bmatrix}O & I \\ A_1 & A_2\end{bmatrix}\begin{bmatrix}\mathbf{Y}_1 \\ \mathbf{Y}_2\end{bmatrix}$$

If the values of $\mathbf{Y}_1 = \mathbf{Y}$ and $\mathbf{Y}_2 = \mathbf{Y}'$ are specified when $t = 0$, then the initial value problem will have a unique solution.

EXAMPLE 3: (Page 307)

In general, if we have an mth-order system of the form

$$\mathbf{Y}^{(m)} = A_1\mathbf{Y} + A_2\mathbf{Y}' + \cdots + A_m\mathbf{Y}^{(m-1)}$$

where each A_i is an $n \times n$ matrix, we can transform it into a first-order system by setting

$$\mathbf{Y}_1 = \mathbf{Y}, \mathbf{Y}_2 = \mathbf{Y}_1', \cdots, \mathbf{Y}_m = \mathbf{Y}_{m-1}'$$

We will end up with a system of the form

$$\begin{bmatrix}\mathbf{Y}_1' \\ \mathbf{Y}_2' \\ \vdots \\ \mathbf{Y}_{m-1}' \\ \mathbf{Y}_m'\end{bmatrix} = \begin{bmatrix}O & I & O & \cdots & O \\ O & O & I & \cdots & O \\ \vdots & & & & \\ O & O & O & \cdots & I \\ A_1 & A_2 & A_3 & \cdots & A_m\end{bmatrix}\begin{bmatrix}\mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_{m-1} \\ \mathbf{Y}_m\end{bmatrix}$$

If, in addition, we require that $\mathbf{Y}, \mathbf{Y}', \cdots, \mathbf{Y}^{(m-1)}$ take on specific values when $t = 0$, there will be exactly one

solution to the problem.

If the system is simply of the form $\mathbf{Y}^{(m)} = A\mathbf{Y}$, it is usually not necessary to introduce new variables. In this case, we need only calculate the mth roots of the eigenvalues of A. If λ is an eigenvalue of A, $\mathbf{x}$ is an eigenvector belonging to λ, σ is an mth root of λ, and $\mathbf{Y} = e^{\sigma t}\mathbf{x}$, then

$$\mathbf{Y}^{(m)} = \sigma^m e^{\sigma t}\mathbf{x} = \lambda\mathbf{Y}$$

and

$$A\mathbf{Y} = e^{\sigma t}A\mathbf{x} = \lambda e^{\sigma t}\mathbf{x} = \lambda\mathbf{Y}$$

Therefore, $\mathbf{Y} = e^{\sigma t}\mathbf{x}$ is a solution to the system.

APPLICATION 2: Harmonic Motion (Page 308)

6.3 Diagonalization

In this section, we consider the problem of factoring an $n \times n$ matrix A into a product of the form XDX^{-1}, where D is diagonal.

Theorem 6.3.1

If $\lambda_1, \lambda_2, \cdots, \lambda_k$ are **distinct** eigenvalues of an $n \times n$ matrix A with corresponding eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k$, then $\mathbf{x}_1, \cdots, \mathbf{x}_k$ are linearly independent.

An $n \times n$ matrix A is said to be **diagonalizable** if there exists a nonsingular matrix X and a diagonal matrix D such that

$$X^{-1}AX = D$$

We say that X **diagonalizes** A.

Theorem 6.3.2

An $n \times n$ matrix A is diagonalizable if and only if A has n linearly independent eigenvectors.

$$\begin{aligned} AX &= (A\mathbf{x}_1, A\mathbf{x}_2, \cdots, A\mathbf{x}_n) \\ &= (\lambda_1\mathbf{x}_1, \lambda_2\mathbf{x}_2, \cdots, \lambda_n\mathbf{x}_n) \\ &= (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \\ &= XD \end{aligned}$$

$$A\mathbf{x}_j = \lambda_j\mathbf{x}_j \quad (\lambda_j = d_{jj})$$

for each j, λ_j is an eigenvalue of A and $\mathbf{x}_j$ is an eigenvector belonging to λ_j.

Remarks

1. If A is diagonalizable, then the column vectors of the diagonalizing matrix X are eigenvectors of A and the diagonal elements of D are the corresponding eigenvalues of A.
2. The diagonalizing matrix X is **not** unique. Reordering the columns of a given diagonalizing matrix X or multiplying them by nonzero scalars will produce a new diagonalizing matrix.
3. If A is $n \times n$ and A has n distinct eigenvalues, then A is diagonalizable. If the eigenvalues are not distinct, then A may or may not be diagonalizable depending on whether A has n linearly independent eigenvectors.
4. If A is diagonalizable, then A can be factored into a product XDX^{-1}.

It follows from remark 4 that

$$A^2 = (XDX^{-1})(XDX^{-1}) = XD^2X^{-1}$$

and, in general,

$$\begin{aligned} A^k &= XD^kX^{-1} \\ &= X\begin{bmatrix} (\lambda_1)^k & & & \\ & (\lambda_2)^k & & \\ & & \ddots & \\ & & & (\lambda_n)^k \end{bmatrix}X^{-1} \end{aligned}$$

Once we have a factorization $A = XDX^{-1}$, it is easy to compute powers of A.

If an $n \times n$ matrix A has fewer than n linearly independent eigenvectors, we say that A is *defective*. It follows from Theorem 6.3.2 that a defective matrix is not diagonalizable.

EXERCISES

14. Let A be a diagonalizable matrix and let X be the diagonalizing matrix. Show that the column vectors of X that correspond to nonzero eigenvalues of A form a basis for $R(A)$.

18. Let A be a diagonalizable $n \times n$ matrix. Prove that if B is any matrix that is similar to A, then B is diagonalizable.

6.4 Hermitian Matrices

Let C^n denote the vector space of all n-tuples of complex numbers. The set C of all complex numbers will be taken as our field of scalars.

Complex Inner Products

If $\alpha = a + bi$ is a complex scalar, the length of α is given by

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

The length of a vector $\mathbf{z} = (z_1, z_2, \cdots, z_n)^T$ in C^n is given by

$$\begin{aligned} \|\mathbf{z}\| &= (|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2)^{1/2} \\ &= (\bar{z}_1 z_1 + \bar{z}_2 z_2 + \cdots + \bar{z}_n z_n)^{1/2} \\ &= (\bar{\mathbf{z}}^T \mathbf{z})^{1/2} \end{aligned}$$

As a notational convenience, we write $\mathbf{z}^H$ for the transpose of $\bar{\mathbf{z}}$, Thus

$$\bar{\mathbf{z}}^T = \mathbf{z}^H \quad \text{and} \quad \|\mathbf{z}\| = (\mathbf{z}^H \mathbf{z})^{1/2}$$

Let V be a vector space over the complex numbers. An inner product on V is an operation that assigns to each pair of vectors $\mathbf{z}$ and $\mathbf{w}$ in V a complex number $\langle \mathbf{z},\mathbf{w} \rangle$ satisfying the following conditions.

I. $\langle \mathbf{z},\mathbf{w} \rangle \geq 0$, with equality if and only if $\mathbf{z} = 0$.

II. $\langle \mathbf{z},\mathbf{w} \rangle = \overline{\langle \mathbf{w},\mathbf{z} \rangle}$ for all $\mathbf{z}$ and $\mathbf{w}$ in V.

III. $\langle \alpha\mathbf{z} + \beta\mathbf{w}, \mathbf{u} \rangle = \alpha\langle \mathbf{z},\mathbf{u} \rangle + \beta\langle \mathbf{w},\mathbf{u} \rangle$.

Note that for a complex inner product space, $\langle \mathbf{z},\mathbf{w} \rangle = \overline{\langle \mathbf{w},\mathbf{z} \rangle}$, rather than $\langle \mathbf{w},\mathbf{z} \rangle$. If we make the proper modifications to allow for this difference, the theorems on real inner product spaces in Chapter 5, Section 5.5, will all be valid for complex inner product spaces. In particular, let us recall Theorem 5.5.2: If $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ is an orthonormal basis for a real inner product space V and

$$\mathbf{x} = \sum_{i=1}^{n} c_i \mathbf{u}_i$$

then

$$c_i = \langle \mathbf{u}_i,\mathbf{x} \rangle = \langle \mathbf{x}, \mathbf{u}_i \rangle \quad \text{and} \quad \|\mathbf{x}\|^2 = \sum_{i=1}^{n} c_i^2$$

In the case of a complex inner product space, if $\{\mathbf{w}_1, \cdots, \mathbf{w}_n\}$ is an orthonormal basis and

$$\mathbf{z} = \sum_{i=1}^{n} c_i \mathbf{w}_i$$

Then

$$c_i = \langle \mathbf{z}, \mathbf{w}_i \rangle, \bar{c}_i = \langle \mathbf{w}_i,\mathbf{z} \rangle \quad \text{and} \quad \|\mathbf{z}\|^2 = \sum_{i=1}^{n} c_i \bar{c}_i$$

We can define an inner product on C^n by

$$\langle \mathbf{z},\mathbf{w} \rangle = \mathbf{w}^H \mathbf{z}$$

for all $\mathbf{z}$ and $\mathbf{w}$ in C^n. The complex inner product space C^n is similar to the real inner product space R^n. The main difference is that in the complex case it is necessary to conjugate before transposing when taking an inner product.

R^n	C^n
$\langle \mathbf{x},\mathbf{y} \rangle = \mathbf{y}^T\mathbf{x}$	$\langle \mathbf{z},\mathbf{w} \rangle = \mathbf{w}^H\mathbf{z}$
$\mathbf{x}^T\mathbf{y} = \mathbf{y}^T\mathbf{x}$	$\mathbf{z}^H\mathbf{w} = \overline{\mathbf{w}^H\mathbf{z}}$
$\|\mathbf{x}\|^2 = \mathbf{x}^T\mathbf{x}$	$\|\mathbf{z}\|^2 = \mathbf{z}^H\mathbf{z}$

Hermitian Matrices

Let $M = (m_{ij})$ be an $m \times n$ matrix with $m_{ij} = a_{ij} + ib_{ij}$ for each i and j. We may write M in the form

$$M = A + iB$$

where $A = (a_{ij})$ and $B = (b_{ij})$ have real entries. We define the conjugate of M by

$$\bar{M} = A - iB$$

Thus, $\bar{M}$ is the matrix formed by conjugating each of the entries of M. The transpose of $\bar{M}$ will be denoted by M^H. The vector space of all $m \times n$ matrices with complex entries is denoted by $\mathrm{C}^{m\times n}$. If A and B are elements of $\mathrm{C}^{m\times n}$ and $C \in \mathrm{C}^{n\times r}$, then the following rules are easily verified:

I. $(A^H)^H = A$
II. $(\alpha A + \beta B)^H = \bar{\alpha}A^H + \bar{\beta}B^H$
III. $(AC)^H = C^H A^H$

A matrix M is said to be **Hermitian** if $M = M^H$.

EXAMPLE 2: The matrix

$$M = \begin{bmatrix} 3 & 2-i \\ 2+i & 4 \end{bmatrix}$$

is Hermitian, since

$$M^H = \begin{bmatrix} \bar{3} & \overline{2-i} \\ \overline{2+i} & \bar{4} \end{bmatrix}^T = \begin{bmatrix} 3 & 2-i \\ 2+i & 4 \end{bmatrix} = M$$

If M is a matrix with **real** entries, then $M^H = M^T$. In particular, if M is a **real symmetric** matrix, then M is Hermitian. Thus we may view Hermitian matrices as the complex analogue of real symmetric matrices.

Theorem 6.4.1

The eigenvalues of a Hermitian matrix are **all real**. Furthermore, eigenvectors belonging to distinct eigenvalues are orthogonal.

Proof

Let A be a Hermitian matrix. Let λ be an eigenvalue of A and let $\mathbf{x}$ be an eigenvector belonging to λ. If $\alpha = \mathbf{x}^H A\mathbf{x}$, (a number) then

$$\bar{\alpha} = \alpha^H = (\mathbf{x}^H A\mathbf{x})^H = \mathbf{x}^H A\mathbf{x} = \alpha$$

Thus, α is real. It follows that

$$\alpha = \mathbf{x}^H A\mathbf{x} = \mathbf{x}^H \lambda\mathbf{x} = \lambda\|\mathbf{x}\|^2$$

and hence

$$\lambda = \frac{\alpha}{\|\mathbf{x}\|^2}$$

is real. If $\mathbf{x}_1$ and $\mathbf{x}_2$ are eigenvectors belonging to distinct eigenvalues λ_1 and λ_2, respectively, then

$$(A\mathbf{x}_1)^H\mathbf{x}_2 = \mathbf{x}_1^H A^H\mathbf{x}_2 = \mathbf{x}_1^H A\mathbf{x}_2 = \lambda_2\mathbf{x}_1^H\mathbf{x}_2$$

and

$$(A\mathbf{x}_1)^H\mathbf{x}_2 = (\mathbf{x}_2^H A\mathbf{x}_1)^H = (\lambda_1\mathbf{x}_2^H\mathbf{x}_1)^H = \lambda_1\mathbf{x}_1^H\mathbf{x}_2$$

Consequently,

$$\lambda_1\mathbf{x}_1^H\mathbf{x}_2 = \lambda_2\mathbf{x}_1^H\mathbf{x}_2$$

and since $\lambda_1 \neq \lambda_2$, it follows that (In real space, we can just replace $\mathbf{x}^H$ with $\mathbf{x}^T$)

$$\langle \mathbf{x}_2, \mathbf{x}_1 \rangle = \mathbf{x}_1^H\mathbf{x}_2 = 0$$

(The proof above is important)

An $n \times n$ matrix U is said to be **unitary** if its column vectors form an orthonormal set in C^n.

Thus U is unitary if and **only** if $U^H U = I$. $(U^{-1} = U^H)$

A real unitary matrix is an orthogonal matrix.

Corollary 6.4.2

If the eigenvalues of a Hermitian matrix A are distinct, then there exists a unitary matrix U that diagonalizes A.

(Let $\mathbf{x}_i$ be an eigenvector belonging to λ_i for each eigenvalue λ_i of A. Let $\mathbf{u}_i = (1/\|\mathbf{x}_i\|\mathbf{x}_i)$. Thus $\mathbf{u}_i$ is a unit eigenvector belonging to λ_i for each i. Following from Theorem 6.4.1 that $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ is an orthonormal set in C^n. Matrix $U = \{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ is unitary and U diagonalizes A.)

EXAMPLE 3:

Let

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

Find a unitary matrix U that diagonalizes A.

The eigenvalues of A are $\lambda_1 = 3$ and $\lambda_2 = 0$, with corresponding eigenvectors $\mathbf{x}_1 = (1-i, 1)^T$ and $\mathbf{x}_2 = (-1, 1+i)^T$. Let

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{x}_1\|}\mathbf{x}_1 = \frac{1}{\sqrt{3}}(1-i, 1)^T$$

and

$$\mathbf{u}_2 = \frac{1}{\|\mathbf{x}_2\|}\mathbf{x}_2 = \frac{1}{\sqrt{3}}(-1, 1+i)^T$$

Thus

$$U = \frac{1}{\sqrt{3}}\begin{bmatrix} 1-i & -1 \\ 1 & 1+i \end{bmatrix}$$

and

$$U^H A U = \frac{1}{3}\begin{bmatrix} 1+i & 1 \\ -1 & 1-i \end{bmatrix}\begin{bmatrix} 2 & 1-i \\ 1+i & 1 \end{bmatrix}\begin{bmatrix} 1-i & -1 \\ 1 & 1+i \end{bmatrix}$$
$$= \begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}$$

Actually, Corollary 6.4.2 is valid even if the eigenvalues of A are not distinct.

Theorem 6.4.3 Schur's Theorem

For each $n \times n$ matrix A, there exists a unitary matrix U such that $U^H A U$ is upper triangular.

The factorization $A = UTU^H$ is often referred to as the *Schur decomposition* of A. In the case that A is Hermitian, the matrix T will be diagonal.

Theorem 6.4.4 Spectral Theorem

If A is Hermitian, then there exists a unitary matrix U that diagonalizes A.

EXAMPLE 4: Given

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

find an orthogonal matrix U that diagonalizes A. (Page 335: Applying the Gram–Schmidt process to obtain an orthonormal basis for the eigenspace)

The Real Schur Decomposition

A subspace S of R^n is said to be **invariant** under a matrix A if, for each $\mathbf{x} \in S$, $A\mathbf{x} \in S$.

Lemma 6.4.5

Let A be a real $n \times n$ matrix with eigenvalue $\lambda_1 = a + bi$ (where a and b are real and $b \neq 0$), and let $\mathbf{z}_1 = \mathbf{x} + i\mathbf{y}$ (where $\mathbf{x}$ and $\mathbf{y}$ are vectors in R^n) be an eigenvector belonging to λ_1. If $S = \mathrm{Span}(\mathbf{x}, \mathbf{y})$, then $\dim S = 2$

and S is invariant under A.

Theorem 6.4.6 The Real Schur Decomposition

If A is an $n \times n$ matrix with real entries, then A can be factored into a product QTQ^T, where Q is an orthogonal matrix and T is in Schur form ((2) in page 337).

Corollary 6.4.7 Spectral Theorem for Real Symmetric Matrices

If A is a **real symmetric matrix**, then there is an orthogonal matrix Q that diagonalizes A; that is, $Q^T AQ = D$, where D is diagonal.

Normal Matrices

There are non-Hermitian matrices that possess complete orthonormal sets of eigenvectors. For example, skew-symmetric and skew-Hermitian matrices have this property.

> A matrix A is said to be **normal** if $AA^H = A^H A$.

We have shown that if a matrix has a complete orthonormal set of eigenvectors, then it is normal. The converse is also true. ($A = UDU^H$)

Theorem 6.4.8

A matrix A is normal if and only if A possesses a complete orthonormal set of eigenvectors.

EXERCISES

6. Show that the diagonal entries of a Hermitian matrix must be real.

6.5 The Singular Value Decomposition

(unfinished)

6.6 Quadratic Forms

In this section, we will see that matrices also play an important role in the study of quadratic equations. With each quadratic equation, we can associate a vector function $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$. Such a vector function is called a *quadratic form*. Quadratic forms arise in a wide variety of applied problems. They are particularly important in the study of optimization theory.

> A **quadratic equation** in two variables x and y is an equation of the form
> $$ax^2 + 2bxy + cy^2 + dx + ey + f = 0$$
> The above equation may be rewritten in the form
> $$[x \quad y]\begin{bmatrix} a & b \\ b & c \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + [d \quad e]\begin{bmatrix} x \\ y \end{bmatrix} + f = 0$$
> Let
> $$\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$
> The term
> $$\mathbf{x}^T A\mathbf{x} = ax^2 + 2bxy + cy^2$$
> is called the **quadratic form** associated with the quadratic equation above.

Conic Sections

The graph of an equation of the form $ax^2 + 2bxy + cy^2 + dx + ey + f = 0$ is called a *conic section*. [If there are no ordered pairs (x, y) which satisfy the form, we say that the equation represents an imaginary conic.] If the graph of the form consists of a single point, a line, or a pair of lines, we say that the form represents a degenerate conic. Of more interest are the nondegenerate conics. Graphs of nondegenerate conics turn out to be circles, ellipses, parabolas, or hyperbolas.

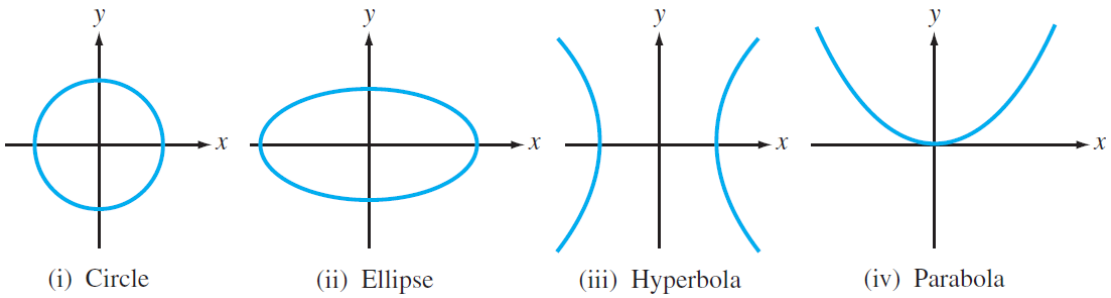


Figure 6. 1

The graph of a conic is particularly easy to sketch when its equation can be put into one of the following standard forms:

i. $x^2 + y^2 = r^2$ (circle)

ii. $\frac{x^2}{\alpha^2} + \frac{y^2}{\beta^2} = 1$ (ellipse)

iii. $\frac{x^2}{\alpha^2} - \frac{y^2}{\beta^2} = 1$ or $\frac{y^2}{\alpha^2} - \frac{x^2}{\beta^2} = 1$ (hyperbola)

iv. $x^2 = \alpha y$ or $y^2 = \alpha x$ (parabola)

Here α, β, and r are nonzero real numbers. Note that the circle is a special case of the ellipse ($\alpha = \beta = r$). A conic section is said to be in *standard position* if its equation can be put into one of these four standard forms.

What about the conics that are not in standard position? Let us consider the following cases:

Case 1. The conic section has been translated horizontally from the standard position. This occurs when the x^2 and x terms in the form of the quadratic equation both have nonzero coefficients.

Case 2. The conic section has been translated vertically from the standard position. This occurs when the y^2 and y terms in the form of the quadratic equation have nonzero coefficients (i.e., $c \neq 0$ and $e \neq 0$).

Case 3. The conic section has been rotated from the standard position by an angle θ that is not a multiple of 90°. This occurs when the coefficient of the xy term is nonzero (i.e., $b \neq 0$).

In general, we may have anyone or any combination of these three cases. To graph a conic section that is not in standard position, we usually find a new set of axes x' and y' such that the conic section is in standard position with respect to the new axes. This is not difficult if the conic has only been translated horizontally or vertically, in which case the new axes can be found by completing the squares.

EXAMPLE 1: Sketch the graph of the equation (Page 358)

$$9x^2 - 18x + 4y^2 + 16y - 11 = 0$$

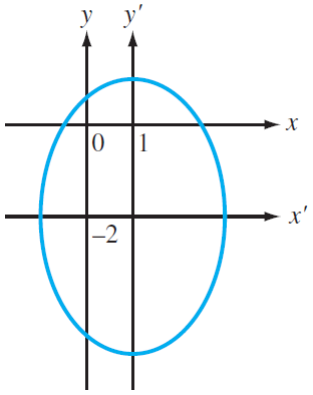


Figure 6. 2

If, however, the conic section has also been rotated from the standard position, it is necessary to change coordinates so that the equation in terms of the new coordinates x' and y' involves no $x'y'$ term. Let $\mathbf{x} = (x, y)^T$ and $\mathbf{x}' = (x', y')^T$. Since the new coordinates differ from the old coordinates by a rotation, we have

$$\mathbf{x} = Q\mathbf{x}' \quad \text{or} \quad \mathbf{x}' = Q^T\mathbf{x}$$

where

$$Q = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \quad \text{or} \quad Q^T = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

With this change of variables, $\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} a & b \\ b & c \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} d & e \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + f = 0$ becomes

$$(\mathbf{x}')^T(Q^TAQ)\mathbf{x}' + \begin{bmatrix} d' & e' \end{bmatrix}\mathbf{x}' + f = 0$$

where $\begin{bmatrix} d' & e' \end{bmatrix} = \begin{bmatrix} d & e \end{bmatrix}Q$. This equation will involve no $x'y'$ term if and only if Q^TAQ is diagonal. Since A is symmetric, it is possible to find a pair of orthonormal eigenvectors $\mathbf{q}_1 = (x_1, -y_1)^T$ and $\mathbf{q}_2 = (y_1, x_1)^T$. Thus, if we set $\cos\theta = x_1$ and $\sin\theta = y_1$, then

$$Q = [\mathbf{q}_1 \quad \mathbf{q}_2] = \begin{bmatrix} x_1 & y_1 \\ -y_1 & x_1 \end{bmatrix}$$

diagonalizes A and $(\mathbf{x}')^T(Q^TAQ)\mathbf{x}' + \begin{bmatrix} d' & e' \end{bmatrix}\mathbf{x}' + f = 0$ simplifies to

$$\lambda_1(x')^2 + \lambda_2(y')^2 + d'x' + e'y' + f = 0$$

EXAMPLE 2: Consider the conic section (Page 359)

$$3x^2 + 2xy + 3y^2 - 8 = 0$$

To summarize, a quadratic equation in the variables x and y can be written in the form

$$\mathbf{x}^TA\mathbf{x} + B\mathbf{x} + f = 0$$

where $\mathbf{x} = (x, y)^T$, A is a 2×2 symmetric matrix, B is a 1×2 matrix, and f is a scalar. If A is **nonsingular**, then, by rotating and translating the axes, it is possible to rewrite the equation in the form ($\det(A) = \lambda_1\lambda_2$)

$$\lambda_1(x')^2 + \lambda_2(y')^2 + f' = 0$$

where λ_1 and λ_2 are the eigenvalues of A. If the above equation represents a real nondegenerate conic, it will be either an ellipse or a hyperbola, depending on whether λ_1 and λ_2 agree in sign or differ in sign. If A is singular and exactly one of its eigenvalues is zero, the quadratic equation can be reduced to either

$$\lambda_1(x')^2 + e'y' + f' = 0 \quad \text{or} \quad \lambda_2(y')^2 + d'x' + f' = 0$$

These equations will represent parabolas, provided that e' and d' are nonzero.

There is no reason to limit ourselves to two variables. We could just as well have quadratic equations and quadratic forms in any number of variables. Indeed, a quadratic equation in n variables $x_1, \cdots, x_n$ is one of the form

$$\mathbf{x}^TA\mathbf{x} + B\mathbf{x} + \alpha = 0$$

where $\mathbf{x} = (x_1, \cdots, x_n)^T$, A is an $n \times n$ symmetric matrix, B is a $1 \times n$ matrix, and α is a scalar. The vector function

$$f(\mathbf{x}) = \mathbf{x}^TA\mathbf{x} = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij}x_j\right)x_i$$

is the *quadratic form* in n variables associated with the quadratic equation.

In the case of **three** unknowns, if

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad A = \begin{bmatrix} a & d & e \\ d & b & f \\ e & f & c \end{bmatrix}, \quad B = \begin{bmatrix} g \\ h \\ i \end{bmatrix}^T$$

then the quadratic equation becomes

$$ax^2 + by^2 + cz^2 + 2dxy + 2exz + 2fyz + gx + hy + iz + \alpha = 0$$

The graph of a quadratic equation in three variables is called a *quadric surface*. (Page 363)

As in the two-dimensional case, we can use translations and rotations to transform the equation into the standard form

$$\lambda_1(x')^2 + \lambda_2(y')^2 + \lambda_3(z')^2 + \alpha = 0$$

where λ_1, λ_2, λ_3 are the eigenvalues of A. For the general n-dimensional case, the quadratic form can always be translated to a simpler diagonal form.

Theorem 6.6.1 Principal Axes Theorem

If A is a real symmetric $n \times n$ matrix, then there is a change of variables $\mathbf{u} = Q^T\mathbf{x}$ such that $\mathbf{x}^TA\mathbf{x} = \mathbf{u}^TD\mathbf{u}$, where D is a diagonal matrix.

Optimization: An Application to the Calculus

Let us consider the problem of maximizing and minimizing functions of several variables. In particular, we would like to determine the nature of the critical points of a real-valued vector function $w = F(\mathbf{x})$. If the function is a quadratic form, $w = \mathbf{x}^T A\mathbf{x}$, then $\mathbf{0}$ is a critical point. Whether it is a maximum, minimum, or saddle point depends on the eigenvalues of A. More generally, if the function to be maximized or minimized is sufficiently differentiable, it behaves locally like a quadratic form. Thus, each critical point can be tested by determining the signs of the eigenvalues of the matrix of an associated quadratic form.

> Let $F(\mathbf{x})$ be a real-valued vector function on R^n. A point $\mathbf{x}_0$ in R^n is said to be a **stationary point** of F if all the first partial derivatives of F at $\mathbf{x}_0$ exist and are zero.

If $F(\mathbf{x})$ has either a local maximum or a local minimum at a point $\mathbf{x}_0$ and the first partials of F exist at $\mathbf{x}_0$, they will all be zero. Thus, if $F(\mathbf{x})$ has first partials everywhere, its local maxima and minima will occur at stationary points.

Consider the quadratic form

$$f(x, y) = ax^2 + 2bxy + cy^2$$

The first partials of f are

$$f_x = 2ax + 2by$$
$$f_y = 2bx + 2cy$$

Setting these equal to zero, we see that $(0,0)$ is a stationary point. Moreover, if the matrix

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

is **nonsingular**, this will be the only critical point. Thus, if A is nonsingular, f will have either a global minimum, a global maximum, or a saddle point at $(0,0)$. ($f_x = 0$ and $f_y = 0 \rightarrow A\mathbf{x} = \mathbf{0}$)

Let us write f in the form

$$f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} \quad \text{where} \quad \mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$$

Since $f(\mathbf{0}) = 0$, it follows that f will have a global minimum at $\mathbf{0}$ if and only if

$$\mathbf{x}^T A\mathbf{x} > 0 \quad \text{for all} \quad \mathbf{x} \neq \mathbf{0}$$

and f will have a global maximum at $\mathbf{0}$ if and only if

$$\mathbf{x}^T A\mathbf{x} < 0 \quad \text{for all} \quad \mathbf{x} \neq \mathbf{0}$$

If $\mathbf{x}^T A\mathbf{x}$ changes sign, then $\mathbf{0}$ is a saddle point.

In general, if f is a quadratic form in n variables, then, for each $\mathbf{x} \in \mathrm{R}^n$,

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

where A is a symmetric $n \times n$ matrix.

> A quadratic form $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is said to be **definite** if it takes on only one sign as $\mathbf{x}$ varies over all nonzero vectors in R^n. The form is **positive definite** if $\mathbf{x}^T A\mathbf{x} > 0$ for all nonzero $\mathbf{x}$ in R^n and **negative definite** if $\mathbf{x}^T A\mathbf{x} < 0$ for all nonzero $\mathbf{x}$ in R^n. A quadratic form is said to be **indefinite** if it takes on values that differ in sign. If $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} \geq 0$ and assumes the value 0 for some $\mathbf{x} \neq \mathbf{0}$, then $f(\mathbf{x})$ is said to be **positive semidefinite**. If $f(\mathbf{x}) \leq 0$ and assumes the value 0 for some $\mathbf{x} \neq \mathbf{0}$, then $f(\mathbf{x})$ is said to be **negative semidefinite**.

Whether the quadratic form is positive definite or negative definite depends on the matrix A. If the quadratic form is positive definite, we say simply that A is positive definite. The preceding definition can then be restated as follows.

A real symmetric matrix A is said to be

I. **positive definite** if $\mathbf{x}^T A\mathbf{x} > 0$ for all nonzero $\mathbf{x}$ in R^n.
II. **negative definite** if $\mathbf{x}^T A\mathbf{x} < 0$ for all nonzero $\mathbf{x}$ in R^n.
III. **positive semidefinite** if $\mathbf{x}^T A\mathbf{x} \geq 0$ for all nonzero $\mathbf{x}$ in R^n.
IV. **negative semidefinite** if $\mathbf{x}^T A\mathbf{x} \leq 0$ for all nonzero $\mathbf{x}$ in R^n.
V. **indefinite** if $\mathbf{x}^T A\mathbf{x}$ takes on values that differ in sign.

If A is nonsingular, then $\mathbf{0}$ will be the only stationary point of $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$. It will be a global minimum if A is positive definite and a global maximum if A is negative definite. If A is indefinite, then $\mathbf{0}$ is a saddle point. To classify the stationary point, we must then classify the matrix A.

Theorem 6.6.2

Let A be a real symmetric $n \times n$ matrix. Then A is positive definite if and only if all its eigenvalues are positive. (Page 365)

If the eigenvalues of A are all negative, A must be negative definite. If A has eigenvalues that differ in sign, then A is indefinite.

Suppose now that we have a function $F(x, y)$ with a stationary point (x_0, y_0). If F has continuous third partials in a neighborhood of (x_0, y_0), it can be expanded in a Taylor series about that point.

$$F(x_0 + h, y_0 + k) = F(x_0, y_0) + \left[hF_x(x_0, y_0) + kF_y(x_0, y_0)\right] + \frac{1}{2}\left[h^2 F_{xx}(x_0, y_0) + 2hkF_{xy}(x_0, y_0) + k^2 F_{yy}(x_0, y_0)\right] + R$$

$$= F(x_0, y_0) + \frac{1}{2}(ah^2 + 2bhk + ck^2) + R$$

where

$$a = F_{xx}(x_0, y_0), \quad b = F_{xy}(x_0, y_0), \quad c = F_{yy}(x_0, y_0)$$

and the remainder R is given by

$$R = \frac{1}{6}\left[h^3 F_{xxx}(\mathbf{z}) + 3h^2 kF_{xxy}(\mathbf{z}) + 3hk^2 F_{xyy}(\mathbf{z}) + k^3 F_{yyy}(\mathbf{z})\right]$$

$$\mathbf{z} = (x_0 + \theta h, y_0 + \theta k), \quad 0 < \theta < 1$$

If h and k are sufficiently small, $|R|$ will be less than $\frac{1}{2}|ah^2 + 2bhk + ck^2|$, and hence $[F(x_0 + h, y_0 + k) - F(x_0, y_0)]$ will have the same sign as $(ah^2 + 2bhk + ck^2)$. The expression

$$f(h, k) = ah^2 + 2bhk + ck^2$$

is a quadratic form in the variables h and k. Thus, $F(x, y)$ will have a local minimum (maximum) at (x_0, y_0) if and only if $f(h, k)$ has a minimum (maximum) at $(0,0)$. Let

$$H = \begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} F_{xx}(x_0, y_0) & F_{xy}(x_0, y_0) \\ F_{xy}(x_0, y_0) & F_{yy}(x_0, y_0) \end{bmatrix}$$

and let λ_1 and λ_2 be the eigenvalues of H. If H is **nonsingular**, then λ_1 and λ_2 are nonzero and we can classify the stationary points as follows:

i. F has a minimum at (x_0, y_0) if $\lambda_1 > 0$, $\lambda_2 > 0$.
ii. F has a maximum at (x_0, y_0) if $\lambda_1 < 0$, $\lambda_2 < 0$.
iii. F has a saddle point at (x_0, y_0) if λ_1 and λ_2 differ in sign.

EXAMPLE 5: The graph of the function (Page 367)

$$F(x, y) = \frac{1}{3}x^3 + xy^2 - 4xy + 1$$

We can now generalize our method of classifying stationary points to functions of more than two variables. Let $F(\mathbf{x}) = F(x_1, \cdots, x_n)$ be a real-valued function whose third partial derivatives are all continuous. Let $\mathbf{x}_0$ be a stationary point of F and define the matrix $H = H(\mathbf{x}_0)$ by

$$h_{ij} = F_{x_i x_j}(\mathbf{x}_0)$$

$H(\mathbf{x}_0)$ is called the ***Hessian*** of F at $\mathbf{x}_0$.

The stationary point can be classified as follows:

i. $\mathbf{x}_0$ is a local minimum of F if $H(\mathbf{x}_0)$ is positive definite.
ii. $\mathbf{x}_0$ is a local maximum of F if $H(\mathbf{x}_0)$ is negative definite.
iii. $\mathbf{x}_0$ is a saddle point of F if $H(\mathbf{x}_0)$ is indefinite.

EXAMPLE 6: Find the local minima of the function

$$F(x, y, z) = x^2 + xz - 3\cos y + z^2$$