THE ALGEBRA OF VECTORS AND MATRICES

by

THOMAS L. WADE

Florida State University

1951

ADDISON-WESLEY PRESS, INC.

CAMBRIDGE 42, MASS.

Printed in the United States of America

PREFACE

In recent years there has been an extensive development of modern algebraic theories and application of them not only to other divisions of mathematics but also to the physical sciences, the social sciences, and statistics. There have been published several excellent treatises which deal with group theory, number theory, matrix algebra, and formal abstract algebra, either compositely or separately, in an advanced way. But the elementary expository presentations of matrix algebra are distinctly limited in number.

This book is intended to give such an exposition. It is concerned largely with an elementary exposition of *the algebra of vectors and matrices*, that exposition being articulated with the basic concepts of modern algebra in the broad sense, to wit, *group*, *integral domain*, *field*, *ring*, *basis*, *dimension*, and *isomorphism*. While the student using this text will be primarily learning about vectors and matrices, he will be getting that knowledge in the setting of modern algebraic theory. This book, therefore, is suitable for use in the first course of a sequence of courses devoted to modern algebraic theories.

There are good reasons for believing that beyond the very basic concepts of groups and fields, that part of "modern higher algebra" which is of greatest interest and use to the applied scientist — the physicist, engineer, theoretical chemist, statistician, or psychometrician — is the algebra of vectors and matrices. To some extent the same is true of the student of mathematics who is mainly interested in geometry or analysis. It is believed that the applied scientist will find here in a single source, in relatively simple, readable language, the essential features of matrix algebra which for him constitutes a powerful and widely used tool.

Considerable use is made of the summation notation, the index symbolism, and transformations, all of which are common in tensor theory. So this book may well serve as a prerequisite for work in matrix and tensor calculus.

A course for applied scientists might omit the material on permutation groups, rings, and groups and matrices (Section 1–3, Section 1–6, Section 6–9, and Chapter 7). If Chapter 7 is omitted, one should take as a definition that "two matrices A and B are similar if $B = P^{-1}AP$ for some nonsingular matrix P." If the book is used by a group of students who have had some vector methods in analytic geometry, Chapter 2 may be used as a reading assignment.

The mathematical background of the student using this book is assumed to be a first course in analytic geometry and an acquaintanceship with the usual theorems on determinants as given in a standard college algebra book.

A major portion of the material in this book has been used in a trial lithoprinted edition at Florida State University during the past two years. The author takes this occasion to express his appreciation to the students who have participated in the use of that material and who have aided in eliminating a number of "bugs" from it.

Thomas L. Wade

Tallahassee, Florida
April, 1951

CONTENTS

CHAPTER 1

BASIC CONCEPTS

1–1 System of elements. Our first basic notion is that of a class, collection, aggregate, or *system* S of undefined objects or entities $a, b, c, \ldots$ which we usually call the *elements* of S. Sometimes we shall specify these elements, but they will usually be any abstract objects. Such a system may contain either a *finite* or an *infinite* number of elements.

Some finite systems are: (1) the system of non-negative integers less than a positive integer n, (2) all the combinations of a given finite set of n elements taken r at a time $(r < n)$, and (3) the n nth roots of unity, for a given positive integer n. Some infinite systems are: (1) all positive integers, (2) all prime integers, and (3) all real numbers.

1–2 Groups. Consider a system S of elements $a, b, c, \ldots$, with a rule of combination or operation designated so that element a followed by element b (which may be any element in the set including a) determines some element, say c, uniquely. While we do not usually state the nature of this operation, we abstractly call $c = ab$ the *product* of a and b, and speak of the operation as *multiplication.*

A system is called a *group* G with respect to a designated operation if and only if it satisfies the following four conditions:

(1) The system is closed under the operation, that is, if a and b are elements of G, then $c = ab$ is a unique element of G.

(2) The associative law must hold, that is, if a, b, c are elements of G, then $(ab)c = a(bc)$. The notation abc may therefore be used without ambiguity.

(3) The set G contains a single element i, called the *identity* element, such that for every element a of G we have $ai = ia = a$.

(4) Each element a of G has a unique *inverse* a^{-1} such that $a^{-1}a = aa^{-1} = i$. These four conditions may be combined compositely in the following definition.

A *group* G is a system of elements which can be combined by a single-valued operation which is associative, and with respect to which G contains an identity element and with each element an inverse.

Systems satisfying condition (1) are called *closed systems*. The relevancy of speaking of such systems as closed is evidenced by the fact that a system satisfying the condition (1) is such that the result of combining any two of its elements is itself an element of the system.

A system selected at random may or may not be closed and may or may not be a group. Consider the system S of the positive integers from 1 to 20 inclusive in connection with the operation of ordinary multiplication. Note that while $2 \cdot 3 = 6$, which lies in S, $3 \cdot 8 = 24$, which is not in S; so this set is not closed. Let S be the set of all natural numbers 1, 2, 3, . . . with respect to the operation of ordinary multiplication. Then condition (1) is satisfied, that is, S is closed; (2) is obviously satisfied; and so is (3), the identity element being 1; and if n is any integer $n \cdot 1 = 1 \cdot n = n$. However, (4) is not satisfied; of the elements in the set S only 1 has an inverse in S.

Let S be the set consisting of all the positive and negative integers including zero, and with addition as the rule of combination. Then one can readily see that S is a group, zero being the identity element, and each element having its own negative for inverse.

A group is said to be *finite* or *infinite* according as the number of elements in it is finite or infinite. If the number of elements in G is the finite number n, then n is said to be the *order* of the finite group G. The numbers $1, -1, i, -i$ ($i^2 = -1$) constitute a group of order 4, the operation being ordinary multiplication.

A group G is called a *commutative* group if its rule of combination is commutative, that is, if $ab = ba$. The numbers 1, -1, i, and $-i$ under ordinary multiplication form a commutative finite group; the negative and positive integers and zero under the operation of addition form a commutative infinite group. Noncommutative groups will appear later.

It should be clearly realized that the elements of a group are not necessarily numbers. They may be motions or acts of many sorts.

Consider the rotations of a six-spoke wheel, spokes evenly spaced, and one pair of spokes horizontal. Let the rotations be counterclockwise and each such that it leaves the wheel with one pair of spokes horizontal. Since consecutive spokes are 60 degrees apart, rotations must be multiples of 60 degrees. Let the rotations of the wheel counterclockwise be designated as follows:

$$a = \text{rotation of 60 degrees}, \quad b = \text{rotation of 120 degrees},$$
$$c = \text{rotation of 180 degrees}, \quad d = \text{rotation of 240 degrees},$$
$$e = \text{rotation of 300 degrees}, \quad i = \text{rotation of 360 degrees}.$$

Note that

$$a + b = c, \quad a + c = d, \quad a + d = e, \quad a + i = a, \text{ etc.}$$

That is, any two of these six rotations performed in succession are equivalent to some other rotation of the set. Further, one can readily verify that the rotations a, b, c, d, e, and i satisfy the other conditions necessary to form a group, ordinary addition being the group operation and i being the identity element.

Not only may the elements of a group be other than numbers, but also the group operation is not necessarily addition or subtraction or any of the usual operations in arithmetic or algebra.

Exercises

1. Verify that the following systems form groups: (a) All real numbers, the rule of combination being addition. (b) All real numbers except zero, the rule of combination being ordinary multiplication. (Here unity is the identity element, and the inverse of an element is its reciprocal.) (c) The five fifth roots of unity under the operation of ordinary multiplication of complex numbers.

2. Show that the functions x, $1/x$, $1 - x$, $1/(1 - x)$, $(x - 1)/x$, $x/(x - 1)$ form a group, the operation being the substitution of the second function for x in the preceding one.

1–3 Permutation groups. As examples of groups which have group operations different from the usual arithmetic and algebraic operations, we now consider permutations. Such groups are of interest in themselves; moreover, there are distinct advantages in working with elementary mathematical systems with operations different from the familiar ones, for later with vectors and matrices we do just that.

Let $a_1, a_2, \ldots a_n$ denote n finite distinct letters or abstract objects. Let $b_1, b_2, \ldots b_n$ be any arrangement of the same n objects which replaces the given arrangement; the operation which effects this rearrangement is called a permutation of the n objects and may be denoted by the symbol

$$\begin{pmatrix} a_1 & a_2 & \ldots & a_n \\ b_1 & b_2 & \ldots & b_n \end{pmatrix}.$$

This means that each letter in the first row is replaced by the letter directly under it. Any order of the given letters may be used provided each has under it the proper letter; thus

$$\begin{pmatrix} a & b & c \\ b & c & a \end{pmatrix} = \begin{pmatrix} b & a & c \\ c & b & a \end{pmatrix} = \begin{pmatrix} c & a & b \\ a & b & c \end{pmatrix}; \text{ etc.}$$

Hence there are six ways of symbolizing the permutation which replaces a by b, b by c, and c by a. The identity

$$i = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ a_1 & a_2 & \cdots & a_n \end{pmatrix}$$

is included among the permutations, even though it does not change the letters. The number of symbols in a permutation is called its *degree*. A permutation of degree two, different from i, is called a *transposition*. From college algebra the total number of permutations on n letters $a_1, a_2, \ldots, a_n$ is $n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$.

For two letters a, b there are two permutations,

$$i = \begin{pmatrix} a & b \\ a & b \end{pmatrix} \quad \text{and} \quad p = \begin{pmatrix} a & b \\ b & a \end{pmatrix}.$$

For three letters a, b, c there are six permutations,

$$i = \begin{pmatrix} a & b & c \\ a & b & c \end{pmatrix}, \quad p = \begin{pmatrix} a & b & c \\ b & c & a \end{pmatrix}, \quad q = \begin{pmatrix} a & b & c \\ c & a & b \end{pmatrix}, \quad r = \begin{pmatrix} a & b & c \\ a & c & b \end{pmatrix},$$

$$s = \begin{pmatrix} a & b & c \\ b & a & c \end{pmatrix}, \quad t = \begin{pmatrix} a & b & c \\ c & b & a \end{pmatrix}.$$

A permutation such as

$$\begin{pmatrix} a_1 & a_2 & \cdots & a_{n-1} & a_n \\ a_2 & a_3 & \cdots & a_n & a_1 \end{pmatrix}$$

is called a *circular permutation* or *cycle*, and for brevity is denoted by the symbol $(a_1a_2 \ldots a_n)$. This means that each letter is replaced by the one which follows it. Clearly a cycle may be represented in several ways, for $(a_1a_2 \ldots a_n) = (a_2a_3 \ldots a_na_1) = (a_3a_4 \ldots a_na_1a_2) = \ldots$ The cycle (abc) may be denoted equivalently by (bca) and (cab).

For two letters the permutations in cyclical form are i and $p = (ab)$. For three letters the permutations in cyclical form are i, $p = (abc)$, $q = (acb)$, $r = (bc)$, $s = (ab)$, $t = (ac)$. A single letter cycle is usually omitted, with the understanding that such a letter is left unchanged by the permutation. Thus t above is a shortened version of $(b)(ac)$. In the two illustrations just given i is used to represent no change in the letters under consideration.

The result obtained by applying to $a_1, a_2, \ldots, a_n$ two permutations successively can be accomplished by applying a single permutation. The latter permutation is called the *product* of the first two. Thus for the products of some permutations on three letters a, b, c, we have

$$pq = \begin{pmatrix} a & b & c \\ b & c & a \end{pmatrix}\begin{pmatrix} a & b & c \\ c & a & b \end{pmatrix} = \begin{pmatrix} a & b & c \\ b & c & a \end{pmatrix}\begin{pmatrix} b & c & a \\ a & b & c \end{pmatrix} = \begin{pmatrix} a & b & c \\ a & b & c \end{pmatrix} = i;$$

$$ps = \begin{pmatrix} a & b & c \\ b & c & a \end{pmatrix}\begin{pmatrix} a & b & c \\ b & a & c \end{pmatrix} = \begin{pmatrix} a & b & c \\ b & c & a \end{pmatrix}\begin{pmatrix} b & c & a \\ a & c & b \end{pmatrix} = \begin{pmatrix} a & b & c \\ a & c & b \end{pmatrix} = (bc) = r;$$

$$sp = \begin{pmatrix} a & b & c \\ b & a & c \end{pmatrix}\begin{pmatrix} a & b & c \\ b & c & a \end{pmatrix} = \begin{pmatrix} a & b & c \\ b & a & c \end{pmatrix}\begin{pmatrix} b & a & c \\ c & b & a \end{pmatrix} = \begin{pmatrix} a & b & c \\ c & b & a \end{pmatrix} = (ac) = t.$$

The process for forming the *product of two permutations* may be stated as follows: *rearrange the columns of the second permutation until its first row is the same as the second row of the first permutation; the product is the permutation whose first row is the first row of the first factor and whose second row is the second row of the second factor.* That is,

$$\begin{pmatrix} a_1 & a_2 & \ldots & a_n \\ b_1 & b_2 & \ldots & b_n \end{pmatrix}\begin{pmatrix} b_1 & b_2 & \ldots & b_n \\ c_1 & c_2 & \ldots & c_n \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & \ldots & a_n \\ c_1 & c_2 & \ldots & c_n \end{pmatrix}.$$

Note that the product of permutations, as ps and sp above, may not be commutative.

In group theory it is established that any permutation may be written in a unique manner as a product of circular permutations of which no two have a letter in common. Thus

$$\begin{pmatrix} a & b & c & d & e & f & g & h \\ e & h & g & f & a & c & b & d \end{pmatrix} = (ae)(bhdfcg) \quad \text{and} \quad \begin{pmatrix} a & b & c & d & e & f \\ e & f & d & a & c & b \end{pmatrix} = (aecd)(bf).$$

The order in which the factor cycles of a given permutation are written is immaterial, since no two of the cycles have a common letter. Thus

$$\begin{pmatrix} a & b & c & d & e \\ c & e & a & b & d \end{pmatrix} = (bed)(ac) = (ac)(bed).$$

Here we understand that by a cycle on less than a total of five letters, such as (ac) and (bed), we mean the permutation on all five letters in which the letters not written down are unchanged.

One may find the product of circular permutations directly. To find the product $(ab)(abc)$ one may reason as follows: in (ab) a is replaced by b, and in (abc) b is replaced by c; the result is that a is replaced by c. Further, in (ab) b is replaced by a, and in (abc) a is replaced by b; so b remains unchanged. Finally, in (ab) c is not mentioned, while in (abc) c is changed to a. All of these results are completely accounted for in (ac). So $(ab)(abc) = (ac)$. Note that a product of cycles having a letter in common, as $(ab)(abc)$, may be put in alternate and simpler form, but that a product of cycles with no letter in common, as $(ab)(cd)$, cannot be further simplified.

Consider the six permutations on three letters,

$$i, \quad (abc), \quad (acb), \quad (bc), \quad (ab), \quad (ac).$$

Observe that

$$(abc)(acb) = i, \quad (bc)(bc) = (ab)(ab) = (ac)(ac) = i,$$

so each element has an inverse. The associative law holds; so does the *identity law*, $pi = ip = p$, for any element p. Finally the set is closed, for

$$(abc)(bc) = (ac), \quad (ab)(abc) = (ac), \quad (ac)(ab) = (acb),$$
$$(ac)(bc) = (abc), \text{ etc.}$$

That the set is closed follows alternatively from the fact that it contains all possible permutations on the three letters. Hence these six permutations on three letters form a noncommutative group of order 3! or 6. It is shown in group theory that the set of all permutations on n letters, $n!$ in number, form a noncommutative group of order $n!$. This group is called the *symmetric group* on n letters, or the symmetric group of order $n!$.

We have previously mentioned that any permutation can be written as the product of cycles no two of which have a letter in common; further, any circular permutation can be written as the product of transpositions, for

$$(abcd \ldots l) = (ab)(ac)(ad)(ae) \ldots (al).$$

Therefore, any permutation can be expressed as the product of transpositions. Thus $(abcd) = (ab)(ac)(ad)$. Note that

$$(ab)(ac) = (ac)(bc) = (bc)(ab) = (abc)$$

and

$$(ac)(ab) = (cb)(ca) = (ba)(bc) = (acb).$$

These illustrations evidence that a permutation of degree greater than two can be expressed as a product of transpositions in more than one way, using only its own letters. A permutation is said to be *even* if it is expressible as a product of an even number of transpositions, and *odd* if it is expressible as the product of an odd number of transpositions.

We have noted that the total set of all permutations on n letters $a_1, a_2, \ldots, a_n$, $n!$ in number, constitutes the so-called *symmetric group* on the n letters, of order $n!$. It is established in group theory that the number of even permutations on $a_1, a_2, \ldots, a_n$ is equal to the number of odd permutations on these same letters. To illustrate, of the six permutations on the letters a, b, c the three odd ones are (bc), (ab), and (ac); the three even ones are i, (abc), and (acb). Now the inverse of an even permutation is even, and the product of two even permutations is even. Therefore, it follows that the even permutations on $a_1, a_2, \ldots, a_n$ form a group; it is called the *alternating group* on these letters, and is of order $\frac{1}{2}n!$. For $n = 2$ the alternating group consists of the identity element alone. For $n = 3$ the alternating group consists of i, (abc), and (acb).

If a group G contains a subset of elements H which forms a group with the same law of combination as G, then H is said to be a *subgroup* of G. The symmetric group of order 6 has four subgroups besides the identity, namely:

$$i, (ab); \quad i, (ac); \quad i, (bc); \quad \text{and} \quad i, (abc), (acb).$$

The latter, the alternating subgroup of the symmetric group on three letters, is a special kind of subgroup known as an *invariant subgroup.* In order to explain the meaning of an invariant subgroup we need to know what is meant by the transform of one element by another. Recurring to the symmetric group of order 6, multiply the element (ab) on the right by (abc) and on the left by (acb), the inverse of (abc). We obtain $(acb)(ab)(abc) = (bc)$; here (bc) is called the *transform* of (ab) by (abc). In general, if $t^{-1}st = r$, we say t *transforms* s into r. Alternately stated, if a given element of a group is multiplied on the right by another element and on the left by the inverse of that element, the result is called the *transform* of the given element by that other element. A subgroup H of G is said to be an *invariant subgroup* if every element of G transforms H into itself. By the latter we do not necessarily mean that each element of the subgroup H remains unchanged, but that each element of the subgroup H is transformed by any element of the group G into an element of the subgroup H; that is, the subgroup H as a whole is unchanged. With the concept of an invariant subgroup now stated, one can prove that the alternating group on n letters is an invariant subgroup of the symmetric group on these n letters.

Exercises

1. The product of every element by every element of a given group may be arranged in a multiplication table for the group. Verify in detail that the following is a multiplication table for the symmetric group on the three letters a, b, c. In each row of the table the common multiplier is used on the right of each of the six elements, and in each column the common multiplier is used on the left of each of the elements.

i	(abc)	(acb)	(ab)	(ac)	(bc)
(ab)	(bc)	(ac)	i	(acb)	(abc)
(abc)	(acb)	i	(ac)	(bc)	(ab)
(ac)	(ab)	(bc)	(abc)	i	(acb)
(acb)	i	(abc)	(bc)	(ab)	(ac)
(bc)	(ac)	(ab)	(acb)	(abc)	i

Thus the results recorded in the second column are to be interpreted as $(abc)(ab) = (bc)$, $(abc)(abc) = (acb)$, $(abc)(ac) = (ab)$, $(abc)(acb) = i$, $(abc)(bc) = (ac)$.

2. Show in detail that i, (abc), (acb) is an invariant subgroup of the symmetric group on a, b, c.

3. Write in cyclic form all the permutations of the symmetric group on the four letters a, b, c, d. Which of these elements constitute the alternating group of degree four?

4. Construct a multiplication table for the group with the permutations i, $(abcd)$, $(ac)(bd)$, and $(adcb)$ as elements and show that this group is a subgroup of the symmetric group of degree four.

1–4 Integral domains and fields. Modern algebra is concerned with the study of a variety of mathematical systems. We have considered briefly one particular mathematical system, a group, from the postulational viewpoint which states initially the basic laws or conditions that the elements of the group satisfy.

It is well to consider some familiar mathematical systems from this postulational point of view. One of the oldest mathematical systems is that of all the *integers* (positive, negative, and zero). For all such integers a, b, and c the following formal laws are satisfied:

		Addition		Multiplication
Commutative law:	$I(i)$	$a + b = b + a$	(ii)	$ab = ba$
Associative law:	$II(i)$	$a + (b + c) = (a + b) + c$	(ii)	$a(bc) = (ab)c$
Distributive law:	III	$a(b + c) = ab + ac.$		

The number zero has the property that it leaves unaltered any number to which it is added; we say that zero is an identity element for addition. The number 1 is an identity element for multiplication. These observations give us the

Identities: $IV(i)\ a + 0 = a \qquad (ii)\ a \cdot 1 = a$ for all a.

The negative $-a$ of a constitutes an

Additive inverse: $V \qquad a + (-a) = 0.$

Quite familiar is the *cancellation law for multiplication:*

$$VI \quad \text{If } ab = 0, \text{ then either } a = 0 \text{ or } b = 0.$$

The second pair of these laws renders the use of parentheses unnecessary for continued sums $a + b + c$ and for continued products abc. Laws $I(ii)$ and VI hold particular interest for us, because the mathematical system of matrices to which we shall give particular attention later will satisfy all of the above laws except these two.

These formal laws which apply to the system of all integers also apply to other systems; namely, to the set of real numbers, to the set of rational numbers, to the set of all complex numbers of the form $a + b\sqrt{-1}$ (a and b real), and to other sets.

A system of elements $a, b, c, \ldots$ which is closed under the operations of addition and multiplication and which satisfies the formal laws just stated for integers is called an *integral domain.* Notice that such a system contains the distinct elements 0 (zero) and 1 (unity) which are identity elements for addition and multiplication, respectively.

An integral domain is a group with respect to the operation of addition, for under that operation the four group properties hold: (1) the system is closed under addition; (2) by $II(i)$ the associative law for addition holds; (3) the system contains a unique element 0 with the property $a + 0 = a$, so zero is the identity element for addition; (4) under addition, with zero as the identity element, $-a$ is the unique additive inverse of a, since $a + (-a) = 0$. From the properties of the zero identity element and the additive inverse the *cancellation law for addition* follows, for if $a + b = a + c$, then $b = c$. Also, since

$$a \cdot 0 + a \cdot a = a(0 + a) = a \cdot a = 0 + a \cdot a,$$

we have $a \cdot 0 = 0$. Similarly, we can prove that $0 \cdot a = 0$.

We elected above to term the law "if $ab = 0$, then either $a = 0$ or

$b = 0$" the *cancellation law for multiplication.* Often the cancellation law for multiplication is stated: if $a \neq 0$ and $ab = ac$, then $b = c$. It should be clear that in the presence of the other postulates for an integral domain either form of the cancellation law for multiplication follows from the other. Suppose $a \neq 0$, and that $ab = ac$. From the latter, $ab - ac = a(b - c) = 0$. Then by 17 above, $b - c = 0$, or $b = c$.

It should be noted that in some integral domains division is possible, while in others it is sometimes possible, but not always. To illustrate, in the integral domain of all integers division is not always possible, that is, if a and b are integers, the equation $ax = b$ does not always have integral solutions. But in many integral domains division is always possible, as in the integral domains of rational numbers, of real numbers, and of complex numbers. Integral domains of this latter type we call *fields.*

A *field* **F** is an integral domain which contains for each element $a \neq 0$ an inverse element a^{-1} satisfying the relation $a^{-1}a = 1$.

THEOREM I. *Division except by zero is possible and unique in any field.*

If $a \neq 0$, we multiply $ax = b$ by a^{-1} and get $a^{-1}ax = a^{-1}b$ or $x = a^{-1}b$, which clearly satisfies $ax = b$. It is the only solution, for $ax = b$ and $ay = b$ together imply $ax = ay$ and, by the cancellation law, $x = y$. We commonly write $a^{-1} = 1/a$ and $a^{-1}b = b/a$.

THEOREM II. *Subtraction is possible and unique for all elements in any field.*

Clearly $x = (-a) + b$ is a solution of the equation $a + x = b$. If there were two solutions x and y, then $a + x = b$, $a + y = b$, and so the two solutions would be identical.

As a consequence of these two theorems, we may state that a set S of two or more elements (numbers) is called a *field* **F** if, when p and $q (q \neq 0)$ are any elements whatever of S, then

$$(1) \qquad p + q, \quad p - q, \quad p \cdot q, \quad p/q$$

are also elements of S.

In testing an integral domain to determine whether it is a field, we should keep in mind simply that *a field is an integral domain in which division by a nonzero element always exists.*

A *subfield* of a given field **F** is a subset of **F** which itself is a field under the same operations which characterize **F**. Thus the set of real numbers is a subfield of the field of complex numbers.

It is of interest to note that a field is a system **F** of elements such that: (*i*) under addition **F** is a commutative group with 0 as the identity element; (*ii*) the elements of **F** not zero form under multiplication another commutative group; (*iii*) addition of elements of **F** is distributive under multiplication.

Exercises

1. Let a/b and c/d be any two rational numbers, a, b, c, and d being integers. Combine them as indicated in (1), and thus verify that rational numbers form a field.

2. Let $a + b\sqrt{-1}$ and $c + d\sqrt{-1}$ be any two complex numbers, a, b, c, and d being real. Subject these two numbers to the combinations of (1), and thereby show that complex numbers constitute a field **C**.

3. The set of all real numbers of the form $a + b\sqrt{2}$, with rational coefficients a and b, is a field. Prove this by showing that any two such numbers when combined in accordance with (1) yield a number of the same form. This is a subfield of the field **R** of all real numbers, and **R** is a subfield of the field **C** of all complex numbers.

1-5 Scalars. In our subsequent consideration of vectors and matrices we shall often speak of *scalar quantities*, or simply *scalars*. By a *scalar* we shall mean an element of some field. Rational numbers, real numbers, and complex numbers are scalars.

The concept of a field has been developed because it facilitates the specification of the kind of number universe with which we are working. Subsequently, when no particular field is specified, the reader should think of some familiar field, as the field **C** of complex numbers or the field **R** of real numbers.

Any "unit number" u gives rise to its multiples, for

$$u = 1 \cdot u, \quad u + u = 2u, \quad 2u + u = 3u, \quad \ldots .$$

From this viewpoint the integers 1, 2, 3, . . . are symbols of "multipliers," rather than numbers in the reference field. There are two cases: (1) either all the multiples $nu \neq 0$, or (2) there is a least n for which $nu = 0$. In the latter case the integer n must be a prime number p. For if n factors in the manner $n = n_1n_2$ (both n_1 and n_2 different from 1), then we would have

$$nu = n_1n_2u = n_1u \cdot n_2u = 0.$$

Consequently, either n_1u or n_2u would be equal to 0, which is incompatible with n being the least value for which $nu = 0$. The characteristic p of the field is the least positive prime integer such that $pu = 0$. If nu is never zero for positive n, the field is said to be of characteristic ∞. Some writers use "characteristic 0" instead of "characteristic ∞." So a field is either of characteristic ∞ or of characteristic $p \neq \infty$. The term "modular field" is sometimes used for a field of characteristic $p \neq \infty$, and similarly the term nonmodular field is sometimes used for a field of characteristic ∞.

In this book when we speak of "an arbitrary field" we shall mean "an arbitrary field of characteristic ∞," and *when we speak of a scalar we shall mean an element of an arbitrary field of characteristic* ∞.

1–6 Rings. For the purposes of this book it is necessary that the reader be familiar with the basic concepts of groups, integral domains, and fields, but not necessarily with the concept of rings. However, since rings constitute one of the important ideas of modern algebra, it appears desirable that we state the properties of a ring, and contrast rings with integral domains and fields. Thereby when we come to the study of matrices we shall be in a position to observe that matrices of order n with elements in a given field constitute a noncommutative ring, and later be in a better position to relate the present material with modern algebraic theories.

A *ring* is a system of elements $a, b, c, \ldots$ which is closed under the two operations of *addition* for which the *sum* $a + b = c$, and *multiplication* for which the product $ab = c$, with the additional properties:

I $\quad a + b = b + a.$

II(*i*) $\quad a + (b + c) = (a + b) + c. \qquad$ (*ii*) $\quad a(bc) = (ab)c.$

III $\quad a(b + c) = ab + ac.$

IV $\quad a + 0 = a.$

V $\quad a + (-a) = 0.$

Evidently a ring is a commutative group with respect to addition, zero being the identity element and $-a$ being the additive inverse of a. If we compare the properties of a ring with those of an integral domain we see that:

For a ring in general the commutative law for multiplication $ab = ba$, does not hold. A ring for which $ab = ba$ is called a *commutative ring*.

For a ring in general there is no identity element of multiplication. If a ring contains an element i such that $ai = ia = a$ for every element a in the ring, i is called the *unity element*, and the ring is called a *ring with unit element*.

For a ring in general the cancellation law for multiplication does not hold. If there is a nonzero element b such that $ab = 0$, or a nonzero element c such that $ca = 0$, then a is said to be a *divisor of zero*. Evidently 0 is always a divisor of zero except for the trivial system with 0 as the only element. A nonzero divisor of zero is called a *proper* divisor of zero.

Thus we see that *a commutative ring with a unity element and without divisors of zero is an integral domain.*

A ring is said to be a *division ring* if for every nonzero element a and arbitrary element b of the ring the equation $ax = b$ always has a solution. It can be shown that a division ring has a unit element, has no divisors of zero, and has an inverse for every nonzero element.* Hence *a commutative division ring is a field.*

We conclude our brief consideration of integral domains, fields, and rings with some examples. It should be evident that rings are the more general systems. An integral domain is a specialized ring, and a field is a specialized integral domain. Since we have not discussed number congruences and residue classes, at this time we have no illustration of a ring with a proper divisor of zero; later we shall see that the set of all square matrices of order n with elements in a field furnishes an excellent illustration of a ring with unit element and with divisors of zero.

Examples of a Ring

(1) All integers.
(2) All even integers.
(3) All even integers divisible by the arbitrary prime p.
(4) All complex numbers.

---

* See N. H. McCoy, "Rings and Ideals," *Carus Mathematical Monograph* No. **8**, the Mathematical Association of America (1948), pp. 19–20.

CLASSIFICATION OF THE EXAMPLES OF A RING

Commutative		With unit element	With proper divisor of zero	Integral domain	Field
1	Yes	Yes	No	Yes	No
2	Yes	No	No	No	No
3	Yes	No	No	No	No
4	Yes	Yes	No	Yes	Yes

In the illustrations just given of rings, integral domains, and fields the operations of addition and multiplication were the familiar operations by those names. However, it should not be assumed that this is necessarily true; the operations of *addition* and *multiplication* for these systems may be abstract and unfamiliar operations, as with the operation of a group.

CHAPTER 2

VECTORS OF TWO AND THREE DIMENSIONS

2–1 Vectors in classical elementary physics. Several types of quantities appear in physics. Of one class are certain quantities called *scalars*, each of which is completely described by a single measure number or *magnitude* which relates the given quantity to a chosen unit. Mass, density, temperature, volume, and population are scalars in this sense. The common ground of all scalars is the field of real numbers, and the physicist's use of the word scalar generally corresponds to the mathematician's use of that word as an element of the field of all real numbers.

Other quantities in physics, such as force, velocity, and acceleration, require for their determination not only *magnitude* but also *direction*. For example, the effect of a force on an object O depends not only on the magnitude of the force but also on its direction; this force may be represented by an arrow or directed line segment $\alpha = OP$ (Fig. 2–1). The combined effect of two such forces $\alpha = OP$ and $\beta = OQ$ acting simultaneously on O, called the *resultant* or sum of α and β, is a third force γ, and we write $\gamma = \alpha + \beta$. It is found empirically that γ is represented in direction and magnitude by the diagonal $\gamma = OR$ of the completed parallelogram with sides $\alpha = OP$ and $\beta = OQ$. This rule for finding $\alpha + \beta$ is called the *parallelogram law*.

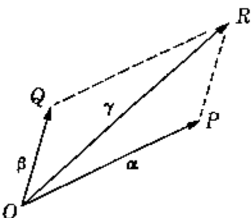

FIG. 2–1.

Physical quantities such as force, which require both direction and magnitude for their specification, are commonly referred to as *line vectors* or simply *vectors*. A line vector may require information in addition to magnitude and direction for its complete specification, such as its point or line of application.

It is noteworthy that the method of obtaining the sum of the line vectors α and β by the parallelogram law does not have a postulational mathematical basis, but rests upon experimentation in physics. We shall see that a line vector in a plane is determined by an ordered pair of scalars (henceforth used in the sense of Section 1–5). Clearly we could define the addition of an ordered pair of scalars in a variety of ways; we shall see that the definition we do give of such addition is reconcilable with the addition of (plane) line vectors. This indicates the method used throughout this book; namely, to reflect how abstractions of lower order may be related to experience, and how experience may suggest these lower order abstractions; then in turn to reflect how these lower order abstractions or simple mathematical systems may suggest more complex mathematical systems; and lastly to indicate how these rationally conceived and rather complex mathematical systems may be used to explore modern mathematical physics, multiple factor analysis of psychology, and multivariate statistical analysis. It is through relatively complex mathematical systems that the human mind explores quantitatively "conceived worlds of reality," even though such worlds are not amenable to human perception.

EXERCISES

1. Two forces of 50 pounds and 80 pounds respectively act on an object at an angle of 30 degrees. With the aid of the law of cosines of trigonometry, find the magnitude and direction of the resultant force.

2. The wind drives a steamer east with a force which would carry it along at a speed of 12 miles per hour, and its propeller is driving it southeast with a force which would carry it along at a speed of 15 miles per hour. With the aid of trigonometry, find the distance it will actually travel in an hour, and the direction of its path.

2–2 Two-dimensional vectors. We shall now develop the theory of vectors along purely algebraic lines; that is, as ordered sets of scalars with algebraic rules for the manipulation of such sets. We shall see that such a vector, when of a simple kind, is susceptible of interpretation as a line vector, but it should be kept in mind that this is an incidental rather than a fundamental feature of the vector idea.

To facilitate scientific investigations of quantitative phenomena we often find it convenient to introduce schemes known as coordinate systems. These reference systems may be of different kinds, but in this chapter we limit ourselves to (right-handed) rectangular cartesian coordinates. In a significant sense these coordinate systems are extraneous to certain concepts which we consider, for we shall see that the most important properties of vectors are those which are essentially independent of the reference system used. In this chapter, when we speak of the coordinates of a point we mean *nonhomogeneous* rectangular cartesian coordinates, whereby the position of a point in two-dimensional space is determined by a pair of scalars and the position of a point in three-dimensional space is determined by a triplet of scalars.

Ordered pairs of scalars such as $\alpha = (a_1, a_2)$, $\beta = (b_1, b_2)$, $\gamma = (c_1, c_2)$, . . . which obey certain rules of combination are called *vectors* (*of two dimensions*). The basic operations of vector algebra are the *multiplication of a vector* α *by a scalar* k:

(1) $$k\alpha = (ka_1, ka_2);$$

and the *addition of two vectors* α *and* β:

(2) $$\alpha + \beta = (a_1 + b_1, a_2 + b_2).$$

Clearly one may associate with a vector $\alpha = (a_1, a_2)$, defined as an ordered set of scalars, a point having the coordinates (a_1, a_2) with reference to the X_1, X_2 cartesian coordinate system; also one may associate with this vector α a direction, by regarding (a_1, a_2) as the direction numbers of a line. The procedure for doing the latter we discuss in the next section.

We ordinarily use small Greek letters $\alpha, \beta, \gamma, \delta, \ldots$ to denote vectors, and small Latin letters $a, b, c, d, \ldots$ to represent scalars. An exception will be the use of θ_1 and θ_2 for the direction angles of a line, and the use of $O = (0, 0)$ for the *null* or zero vector.

2–3 Direction numbers in two-dimensional space. To make clear the notion of direction numbers, consider any line L, with a definite direction indicated, lying in the coordinate plane. Let OP be a line drawn through the origin in the same direction as L. The smallest possible angles, θ_1 and θ_2 respectively, which the directed line OP makes with the positive directions of the X_1, and X_2 axes are called the *direction angles* of L; these angles as drawn in Fig. 2–2 are

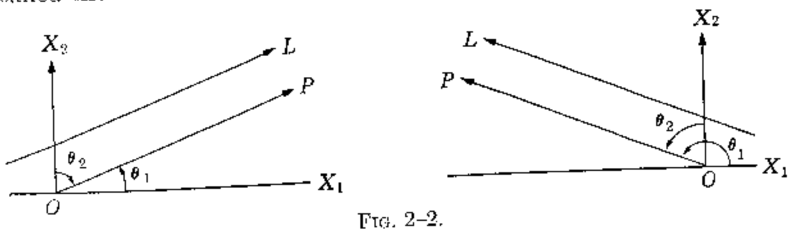

FIG. 2–2.

considered as positive, whether clockwise or countercloc cosines of these angles, $\cos\theta_1$ and $\cos\theta_2$, are called the *dire* of L. If the direction of L is reversed, the new directi and θ_2' are respective plements of θ_1 and θ_2.

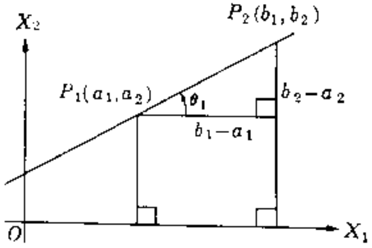

FIG. 2-3.

Let $P_1(a_1, a_2)$ an be two points on L. 2-3 we see that we ha

(3) $\cos\theta_1 = \frac{b_1 -}{d}$

and

$\cos\theta_2 = \frac{b_2 -}{d}$

where d is the distance from P_1 to P_2. Squaring and latter relations we get

(4) $$\cos^2\theta_1 + \cos^2\theta_2 = 1.$$

That is, the sum of the squares of the direction cosines of a line is equal to 1. In particular consider the directed li origin $O(0, 0)$ to $P_1(a_1, a_2)$. Then

$$\cos\theta_1 = \frac{a_1}{d}, \quad \cos\theta_2 = \frac{a_2}{d}.$$

If P_1 is a point on the unit circle (a circle with center a and radius 1), then $d = 1$, and the coordinates of P_1 a with the direction cosines of the line OP_1. We have thus a one-to-one correspondence between the points on the and all sets of direction cosines, and have proved that pair of real numbers is a set of direction cosines if and only of their squares is equal to 1.

If a line has direction cosines $\cos\theta_1$, $\cos\theta_2$, then an numbers n_1, n_2 proportional to $\cos\theta_1$, $\cos\theta_2$ are called *dir bers* of L. This relationship may be expressed in the forn

(5) $$\frac{\cos\theta_1}{n_1} = \frac{\cos\theta_2}{n_2} = c,$$

whence $\cos\theta_1 = n_1c$, $\cos\theta_2 = n_2c$. Squaring, adding, and we get

$$c = 1 \div \sqrt{n_1^2 + n_2^2}.$$

Therefore, we have

$$(6)\qquad \cos\theta_1 = \frac{a_1}{\pm\sqrt{a_1^2 + a_2^2}}, \quad \cos\theta_2 = \frac{a_2}{\pm\sqrt{a_1^2 + a_2^2}}.$$

These relations enable us to calculate the direction cosines of a line when a set of direction numbers for it are known.

From the definition of direction numbers of a line it follows that if a_1, a_2 are direction numbers of a line L, so are ka_1, ka_2, where k is any real number; that is, the direction numbers of a line L are not unique, although the direction cosines of a directed line are unique. Accordingly, we commonly speak of *a pair of direction numbers* of a directed line L; but we speak of *the direction cosines* of L. However, we speak of c_1, c_2 as being *the direction numbers of the directed line segment* P_1P_2; by this we mean that c_1, c_2 constitute a pair of direction numbers of the line of infinite extent determined by P_1P_2, and further that if P_1 has the coordinates (a_1, a_2) and P_2 has the coordinates (b_1, b_2), then $c_1 = b_1 - a_1$ and $c_2 = b_2 - a_2$. Therefore, for a specified pair of direction numbers there is a line segment of determinable length and direction. From the formulas (3) and the definition of direction numbers, we have

THEOREM I. *If $P_1(a_1, a_2)$ and $P_2(b_1, b_2)$ are two distinct points, then $b_1 - a_1$, $b_2 - a_2$ are the pair of direction numbers of the directed line segment P_1P_2.*

We designate the *directed line segment* determined by the two points $P_1(a_1, a_2)$ and $P_2(b_1, b_2)$ by either P_1P_2 or $(a_1, a_2) \to (b_1, b_2)$. When we want to refer to the line of infinite extent through P_1 and P_2 we speak of the *(directed) line* P_1P_2. As a special case of Theorem I we have

THEOREM II. *The direction numbers of OP, where P is the point (a_1, a_2), are a_1 and a_2. That is, the coordinates of any point P are the direction numbers of the line segment OP.*

Thus with $\alpha = (a_1, a_2)$ we may associate a point P with the coordinates a_1, a_2; further, these coordinates a_1, a_2 may be considered as the direction numbers of OP; for they determine a direction with direction cosines a_1/d, a_2/d, where d is the distance from O to P, and $d^2 = a_1^2 + a_2^2$.

If P_1P_2 and P_3P_4 are two directed line segments in the plane of reference with the same direction numbers, then these segments have the same length, and the lines on which the segments lie have the considered as positive, whether clockwise or counterclockwise. The cosines of these angles, $\cos \theta_1$ and $\cos \theta_2$, are called the *direction cosines* of L. If the direction of L is reversed, the new direction angles θ_1' and θ_2' are respectively the supplements of θ_1 and θ_2.

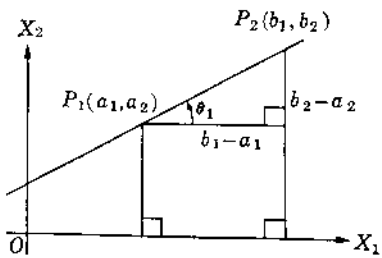

FIG. 2–3.

Let $P_1(a_1, a_2)$ and $P_2(b_1, b_2)$ be two points on L. From Fig. 2–3 we see that we have

$$(3) \qquad \cos \theta_1 = \frac{b_1 - a_1}{d}$$

and

$$\cos \theta_2 = \frac{b_2 - a_2}{d},$$

where d is the distance from P_1 to P_2. Squaring and adding the latter relations we get

$$(4) \qquad \cos^2 \theta_1 + \cos^2 \theta_2 = 1.$$

That is, the sum of the squares of the direction cosines of any directed line is equal to 1. In particular, consider the directed line from the origin $O(0, 0)$ to $P_1(a_1, a_2)$. Then

$$\cos \theta_1 = \frac{a_1}{d}, \quad \cos \theta_2 = \frac{a_2}{d}.$$

If P_1 is a point on the unit circle (a circle with center at the origin and radius 1), then $d = 1$, and the coordinates of P_1 are identical with the direction cosines of the line OP_1. We have thus established a one-to-one correspondence between the points on the unit circle and all sets of direction cosines, and have proved that an ordered pair of real numbers is a set of direction cosines if and only if the sum of their squares is equal to 1.

If a line has direction cosines $\cos \theta_1$, $\cos \theta_2$, then any two real numbers n_1, n_2 proportional to $\cos \theta_1$, $\cos \theta_2$ are called *direction numbers* of L. This relationship may be expressed in the form

$$(5) \qquad \frac{\cos \theta_1}{n_1} = \frac{\cos \theta_2}{n_2} = c,$$

whence $\cos \theta_1 = n_1 c$, $\cos \theta_2 = n_2 c$. Squaring, adding, and using (4), we get

$$c = 1/\pm\sqrt{n_1^2 + n_2^2}.$$

Therefore, we have

$$\text{(6)} \qquad \cos\theta_1 = \frac{n_1}{\pm\sqrt{n_1^2 + n_2^2}}, \quad \cos\theta_2 = \frac{n_2}{\pm\sqrt{n_1^2 + n_2^2}}.$$

These relations enable us to calculate the direction cosines of a line when a set of direction numbers for it are known.

From the definition of direction numbers of a line it follows that if n_1, n_2 are direction numbers of a line L, so are kn_1, kn_2, where k is any real number; that is, the direction numbers of a line L are not unique, although the direction cosines of a directed line are unique. Accordingly, we commonly speak of *a pair of direction numbers* of a directed line L; but we speak of *the direction cosines* of L. However, we speak of c_1, c_2 as being *the direction numbers of the directed line segment* P_1P_2; by this we mean that c_1, c_2 constitute a pair of direction numbers of the line of infinite extent determined by P_1P_2, and further that if P_1 has the coordinates (a_1, a_2) and P_2 has the coordinates (b_1, b_2), then $c_1 = b_1 - a_1$ and $c_2 = b_2 - a_2$. Therefore, for a specified pair of direction numbers there is a line segment of determinable length and direction. From the formulas (3) and the definition of direction numbers, we have

THEOREM I. *If $P_1(a_1, a_2)$ and $P_2(b_1, b_2)$ are two distinct points, then $b_1 - a_1$, $b_2 - a_2$ are the pair of direction numbers of the directed line segment P_1P_2.*

We designate the *directed line segment* determined by the two points $P_1(a_1, a_2)$ and $P_2(b_1, b_2)$ by either P_1P_2 or $(a_1, a_2) \to (b_1, b_2)$. When we want to refer to the line of infinite extent through P_1 and P_2 we speak of the *(directed) line* P_1P_2. As a special case of Theorem I we have

THEOREM II. *The direction numbers of OP, where P is the point (a_1, a_2), are a_1 and a_2. That is, the coordinates of any point P are the direction numbers of the line segment OP.*

Thus with $\alpha = (a_1, a_2)$ we may associate a point P with the coordinates a_1, a_2; further, these coordinates a_1, a_2 may be considered as the direction numbers of OP; for they determine a direction with direction cosines a_1/d, a_2/d, where d is the distance from O to P, and $d^2 = a_1^2 + a_2^2$.

If P_1P_2 and P_3P_4 are two directed line segments in the plane of reference with the same direction numbers, then these segments have the same length, and the lines on which the segments lie have the same direction. If we slide the segment P_1P_2 in the plane, with its length and direction unchanged, until P_1 falls on P_3, then P_2 will fall on P_4. All segments into which P_1P_2 may be slid in this way are said to be equivalent to P_1P_2. Consequently *two directed segments are equivalent when they have the same direction numbers.* Clearly, with any vector $\alpha = (a_1, a_2)$ we may associate an infinite number of directed line segments, all with the same length and direction numbers. It is usual to speak of the entire collection of all such equivalent directed line segments as the *line vector* $\alpha = (a_1, a_2)$, and regard any one of them as a particular representation of the vector α. Thus *two directed line segments represent the same vector when and only when they are equivalent, that is, when and only when they have the same direction numbers.*

The viewpoint of regarding two directed line segments P_1P_2 and P_3P_4 with the same direction numbers as equivalent should be weighed carefully. In adhering to this viewpoint we are emphasizing that the things of particular concern about the directed segment P_1P_2 are its length and its direction.

It is convenient to speak of the scalar components a_1, a_2 of the vector $\alpha = (a_1, a_2)$ as the *coordinates* of the vector. Thus the coordinates of the line vector determined by the points $P_1(a_1, a_2)$ and $P_2(b_1, b_2)$ are $b_1 - a_1, b_2 - a_2$. The coordinates of the line vector $(0, 0) \to (a_1, a_2)$ are a_1, a_2, which are the coordinates of the point $P_1(a_1, a_2)$.

The distance P_1P_2 is given by (refer to Fig. 2–3)

$$(P_1P_2)^2 = (b_1 - a_1)^2 + (b_2 - a_2)^2. \tag{7}$$

Let $(c_1, c_2) = (b_1 - a_1, b_2 - a_2)$. Then (c_1, c_2) are *the pair of direction numbers for the directed line segment* P_1P_2, and (c_1, c_2) *are a pair of direction numbers* for the line P_1P_2. However, any pair of ordered scalars proportional to (c_1, c_2), as (kc_1, kc_2), is also a pair of direction numbers for the line P_1P_2. By assigning to k any positive value we please, we obtain a particular pair of direction numbers of the line P_1P_2; any such particular pair is representative of the entire aggregate of pairs of direction numbers of this line. If we assign to k a negative value we obtain a pair of direction numbers of the oppositely directed line P_2P_1.

Exercises

1. Determine the direction numbers of the following directed line segments:

$(3, 4) \to (8, 7)$; $(-3, 8) \to (7, 9)$; $(-4, -5) \to (-1, -2)$.

2. Show that the direction numbers of P_1P_2 are the negative of the direction numbers of P_2P_1.

3. Show that when the first direction number of P_1P_2 is zero, the line through P_1 and P_2 is parallel to the X_2 axis. Show that when the second direction number of P_1P_2 is zero, the line through P_1 and P_2 is parallel to the X_1 axis.

4. Determine the coordinates of the terminal point of the line segment whose initial point is $(2, 3)$ and whose direction numbers are $(1, 4)$.

5. Show that $(0, 0) \to (5, 2)$ and $(-2, 2) \to (3, 4)$ represent the same vector.

6. Show that if P_1P_2 and P_3P_4 represent the same vector, then P_1P_3 and P_2P_4 represent the same vector.

7. For $P_1(2, 3)$ and $P_2(4, 7)$, find: (*a*) the direction numbers of P_1P_2; (*b*) two pairs of direction numbers for the line P_1P_2. (*c*) Does it matter which of these pairs of direction numbers we use in determining the direction cosines of the line P_1P_2?

8. Show that the direction cosines of the line with direction numbers (n_1, n_2) are the same as the direction cosines of the line with direction numbers (kn_1, kn_2).

2–4 Line vectors and position vectors. We have seen that a two-dimensional vector has two distinct aspects, one of which may be called its algebraic aspect and the other its geometric aspect. In the sense that an algebraic vector determines a point, we may relevantly speak of it as a *position vector*. In turn, since a *line vector* is determined by two points, we may say that a line vector is determined by two position vectors. The line vector determined by the position vectors $\alpha = (a_1, a_2)$ and $\beta = (b_1, b_2)$ is denoted by $\beta + (-1)\alpha = \beta - \alpha$, or sometimes by $\boldsymbol{\alpha\beta}$, the initial point being α and the terminal point being β; we shall refer to the line of infinite extent determined by these position vectors as the line $\boldsymbol{\alpha\beta}$. Thus the direction numbers of this line vector are the components of the vector

$$\delta = \beta - \alpha = (b_1 - a_1, b_2 - a_2).$$

The length of the line segment from α to β is called the *magnitude* (*mag.*) of the line vector $\boldsymbol{\alpha\beta}$. So

(8) $$\text{mag}^2(\beta - \alpha) = (b_1 - a_1)^2 + (b_2 - a_2)^2.$$

In general the magnitude of a vector $\alpha = (a_1, a_2)$ is the positive square root of the sum of the squares of the coordinates of the vector; that is,

$$\text{mag } \alpha = \sqrt{a_1^2 + a_2^2}.$$

A line vector takes on its simplest form when one end point is the origin $O(0, 0)$. For if $\alpha = (a_1, a_2)$ is a position vector, and $O(0, 0)$ is the position vector determining the fundamental point of reference, the line vector

(9) $$\mathbf{O}\alpha = \alpha - O = (a_1, a_2) - (0, 0) = (a_1, a_2) = \alpha.$$

Thus $\mathbf{O}\alpha = \alpha$; that is, *the position vector α and the line vector $\mathbf{O}\alpha$ have the same algebraic form.* To avoid ambiguity we make the convention that *when we speak of the vector α in a geometric setting we have in mind the position vector α, except when α is to be regarded as a line vector, and then we shall speak of it as the vector $\mathbf{O}\alpha$.*

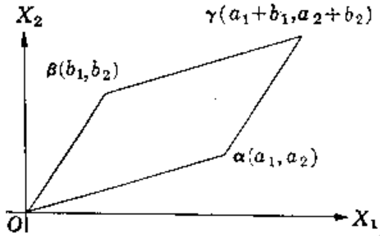

FIG. 2–4.

Consider the figure (Fig. 2–4) determined by $O = (0, 0)$, $\alpha = (a_1, a_2)$, $\beta = (b_1, b_2)$, and let $\gamma = (a_1 + b_1, a_2 + b_2) = \alpha + \beta$.

The projection of $\mathbf{O}\alpha$ on the X_1 axis is a_1 and on the X_2 axis is a_2. Similarly the projection of $\mathbf{O}\beta$ on the X_1 and X_2 axes are b_1 and b_2 respectively. If $\boldsymbol{\alpha\gamma}$ is drawn with the same direction and length as $\mathbf{O}\beta$, then its projections on the reference axes are equal to those of $\mathbf{O}\beta$, namely b_1 and b_2. Consequently the projections of $\mathbf{O}\gamma$, considered as the line vector obtained by adding $\mathbf{O}\alpha$ and $\mathbf{O}\beta$, are $a_1 + b_1$ *and* $a_2 + b_2$. But these are precisely the components of the vector obtained by the algebraic addition of $\alpha = (a_1, a_2)$ and $\beta = (b_1, b_2)$. In other words, *the algebraic addition of vectors defined by* (2) *is reconciled with the geometric addition of vectors by the parallelogram law.*

EXERCISES

1. Determine the magnitude of $\gamma = \alpha + \beta$, where $\alpha = (2, 3)$ and $\beta = (5, 7)$.

2. Find two representative segments of the vector $\alpha = (3, 4)$. What is the magnitude of this vector?

3. Find the line vector $\boldsymbol{\alpha\beta}$ where $\alpha = (1, 1)$ and $\beta = (3, 2)$. How does this compare with the vector $\mathbf{O}\gamma$ where $\gamma = (2, 1)$?

4. For $\alpha = (3, 4)$, $\beta = (-5, 6)$, $\gamma = (5, -2)$ find algebraically:

(a) $\alpha + \beta$. (b) $\alpha + 2\beta$. (c) $\alpha - \beta$. (d) $\frac{1}{3}(\alpha + \beta + \gamma)$.

2–5 Three-dimensional vectors. We assume familiarity with a rectangular cartesian coordinate system whereby the position of a point in ordinary space is determined by its perpendicular distances from three mutually perpendicular planes. These three perpendicular planes intersect in three mutually perpendicular lines, the X_1 axis, X_2 axis, and X_3 axis.

Ordered triads of scalars as $\alpha = (a_1, a_2, a_3)$, $\beta = (b_1, b_2, b_3)$, $\gamma = (c_1, c_2, c_3)$, . . . which obey certain rules of combination are called *vectors (of three dimensions)*. The basic operations of the algebra of such vectors are the *multiplication of a vector* α *by a scalar* k:

$$k\alpha = (ka_1, ka_2, ka_3); \tag{10}$$

and the *addition of two vectors* α *and* β:

$$\alpha + \beta = (a_1 + b_1, a_2 + b_2, a_3 + b_3). \tag{11}$$

There exists a unique vector $O = (0, 0, 0)$ such that for any vector α, $\alpha + O = \alpha$. If α is any non-null vector, we shall denote by $-\alpha$ the vector $(-1)\alpha$. Two three-dimensional vectors $\alpha = (a_1, a_2, a_3)$ and $\beta = (b_1, b_2, b_3)$ are *equal* if corresponding scalar elements are equal. The equality of two such vectors implies the satisfying of three scalar equations: $a_1 = b_1$, $a_2 = b_2$, $a_3 = b_3$.

Analogous to the procedure with two-dimensional vectors, we may regard α as a *position vector* determining a point, or as a *line vector* of which $\mathbf{O}\alpha$ *is a* representation.

Let $\epsilon_1 = (1, 0, 0)$, $\epsilon_2 = (0, 1, 0)$, $\epsilon_3 = (0, 0, 1)$ be three unit vectors (each with magnitude 1); then $\mathbf{O}\epsilon_1$, $\mathbf{O}\epsilon_2$, $\mathbf{O}\epsilon_3$ are unit line vectors along the X_1, X_2, and X_3 axes. Note that

$$\alpha = (a_1, a_2, a_3) = (a_1, 0, 0) + (0, a_2, 0) + (0, 0, a_3),$$

or

$$\alpha = a_1\epsilon_1 + a_2\epsilon_2 + a_3\epsilon_3. \tag{12}$$

The vectors $a_1\epsilon_1$, $a_2\epsilon_2$, $a_3\epsilon_3$ which appear in the linear combination (12) are called *the (vector) components* of α with respect to the coordinate system just described, and the scalar coefficients a_1, a_2, a_3 in this linear combination are termed *the (scalar) coordinates* of α with reference to this coordinate system.

Consider the directed line L. Suppose L does not pass through the origin; draw L' parallel to L through the origin, as indicated in Fig. 2–5. The directed line L' makes definite angles θ_1, θ_2, θ_3 with the positive direction of each of the X_1, X_2, X_3 axes. These angles are called the *direction angles* of L' and of L. Each of these direction angles is between 0 and 180 degrees inclusive. If the direction of L is reversed, the new directed line has direction angles θ_1', θ_2', θ_3', where the latter are the supplements of θ_1, θ_2, θ_3 respectively. The cosines of the direction angles, $\cos\theta_1$, $\cos\theta_2$, $\cos\theta_3$ are called the *direction cosines* of the line L.

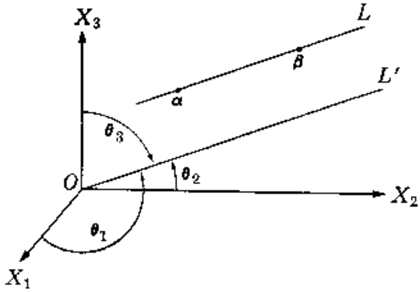

FIG. 2–5.

Let $\alpha = (a_1, a_2, a_3)$ and $\beta = (b_1, b_2, b_3)$ be two points on the line L. Comparable to Section 2–3, we have

$$(13)\qquad \cos\theta_1 = \frac{b_1 - a_1}{d}, \quad \cos\theta_2 = \frac{b_2 - a_2}{d}, \quad \cos\theta_3 = \frac{b_3 - a_3}{d},$$

where d is the distance from α to β; and

$$d^2 = (b_1 - a_1)^2 + (b_2 - a_2)^2 + (b_3 - a_3)^2.$$

Squaring both sides of each of the relations (13) and adding the results, we get

$$(14)\qquad \cos^2\theta_1 + \cos^2\theta_2 + \cos^2\theta_3 = 1.$$

Suppose $\alpha = (a_1, a_2, a_3)$ is a point on L', a line through the origin. Then $\cos\theta_1 = a_1/d$, $\cos\theta_2 = a_2/d$, $\cos\theta_3 = a_3/d$, where d is the distance from O to α; that is, $d^2 = a_1^2 + a_2^2 + a_3^2$. If $\alpha = (a_1, a_2, a_3)$ is a point of the unit sphere (a sphere with center at the origin and radius 1), then $d = 1$, and the coordinates of α are identical with the direction cosines of the line $\mathbf{O}\alpha$.

If a line has direction cosines $\cos\theta_1$, $\cos\theta_2$, $\cos\theta_3$, any three real numbers n_1, n_2, n_3 proportional to $\cos\theta_1$, $\cos\theta_2$, $\cos\theta_3$ are called *direction numbers* of the line L. This relation may be expressed in the form

$$(15)\qquad \frac{\cos\theta_1}{n_1} = \frac{\cos\theta_2}{n_2} = \frac{\cos\theta_3}{n_3} = c,$$

whence $\cos\theta_1 = n_1c$, $\cos\theta_2 = n_2c$, $\cos\theta_3 = n_3c$. Squaring, adding, and using (14), we get $c = 1/\pm\sqrt{n_1^2 + n_2^2 + n_3^2}$. Therefore, we have

$$(16)\qquad (\cos\theta_1, \cos\theta_2, \cos\theta_3) = \frac{1}{\pm\sqrt{n_1^2 + n_2^2 + n_3^2}}(n_1, n_2, n_3).$$

If (n_1, n_2, n_3) is a triad of direction numbers for a line, so also is (kn_1, kn_2, kn_3); we appropriately speak of *a triad of direction numbers* for a line $\boldsymbol{\alpha\beta}$. However, as in two dimensions, we speak of $(c_1, c_2, c_3) = (b_1 - a_1, b_2 - a_2, b_3 - a_3)$ as being *the triad of direction numbers* for the directed line segment $\boldsymbol{\alpha\beta}$, where $\alpha = (a_1, a_2, a_3)$ and $\beta = (b_1, b_2, b_3)$. From relations (13) and the definition of direction numbers, we have

THEOREM III. *If $\alpha = (a_1, a_2, a_3)$ and $\beta = (b_1, b_2, b_3)$ are two distinct points, then*

$$\gamma = \beta - \alpha = (b_1 - a_1, b_2 - a_2, b_3 - a_3)$$

is the triad of direction numbers for $\boldsymbol{\alpha\beta}$.

Oftentimes we shall speak simply of the *direction of a line* and by this we shall mean the difference in the position vectors associated with any two points on the line. Otherwise stated, the direction of a line is the line vector corresponding to any segment on that line; the *direction of the line* $\boldsymbol{\alpha\beta}$ is $\gamma = \beta - \alpha$. As a particular case of Theorem III, we have

THEOREM IV. *The direction numbers of* $\mathbf{O\alpha}$, *where* $\alpha = (a_1, a_2, a_3)$ *and* $O = (0, 0, 0)$, *are* (a_1, a_2, a_3). *That is, the coordinates of any point* α *are the direction numbers of* $\mathbf{O}\alpha$.

EXERCISES

1. Find the direction numbers of the following directed line segments:

$$(3, 4, 5) \to (8, 7, 6); \quad (-4, -5, -6) \to (-1, -2, 3).$$

2. Find the terminal point of the directed line segment whose initial point is $\alpha = (3, 4, -5)$ and whose direction numbers are $(2, 4, 7)$.

3. For $\alpha = (2, 3, 4)$ and $\beta = (4, 7, 6)$ find: the direction numbers for the segment $\boldsymbol{\alpha\beta}$; two triads of direction numbers for the line $\boldsymbol{\alpha\beta}$. Does it matter which of these triads of direction numbers we use in determining the direction cosines of the line $\boldsymbol{\alpha\beta}$?

4. Using the relation (16), show that the direction cosines of the line with direction numbers $\alpha = (n_1, n_2, n_3)$ are the same as the direction cosines of the line with direction numbers $k\alpha = (kn_1, kn_2, kn_3)$.

5. For $\alpha = (3, 4, 1)$, $\beta = (-5, 6, 2)$, $\gamma = (5, -2, 1)$ find algebraically:

$$\alpha + \beta; \quad \alpha + 2\beta; \quad \beta - \gamma; \quad \tfrac{1}{3}(\alpha + \beta + \gamma).$$

6. Given that $\alpha = a_1\epsilon_1 + a_2\epsilon_2 + a_3\epsilon_3$ and $\beta = b_1\epsilon_1 + b_2\epsilon_2 + b_3\epsilon_3$, find $\alpha + \beta$ and $\alpha - \beta$ in terms of the unit vectors ϵ_1, ϵ_2, ϵ_3 and the coordinates of α and β.

2–6 Laws of vectors. As we said in Section 1–5, in general when we speak of a scalar we mean an element of an arbitrary field (of characteristic ∞); however, the reader will not miss anything essential if he thinks of a scalar as a real number, provided he keeps an open mind toward the possibility of the theory under consideration holding for other fields. We use the symbol $V_n(\mathbf{F})$ to designate a system of n-dimensional vectors with coordinates in the field $\mathbf{F}$; we often have occasion to speak of $V_n(\mathbf{C})$, the vectors over the field $\mathbf{C}$ of complex numbers, and $V_n(\mathbf{R})$, the vectors over the field $\mathbf{R}$ of real numbers. Vectors of complex numbers, $V_n(\mathbf{C})$, play an important role in both pure and applied mathematics; their great disadvantage, particularly in an introductory presentation, lies in the difficulty of conceiving a geometric picture of them. The graphical picture of $V_1(\mathbf{C})$ is indistinguishable from that of $V_2(\mathbf{R})$, and a graphical picture of $V_n(\mathbf{C})$ for $n > 1$ seems to be inconceivable by the human mind. For this reason it is understood that when we relate the vector concepts of this chapter to a coordinate system we have in mind the vectors with elements or coordinates in the field of real numbers.

We now study in considerable detail the algebra of three-dimensional vectors. It should be clear that just as a point in the X_1, X_2 coordinate plane may be considered as a point $\alpha = (a_1, a_2, 0)$ with reference to the X_1, X_2, X_3 space coordinate system, so may two-dimensional vectors be considered as three-dimensional vectors in which the third components are zero. Unless stated specifically to the contrary, the concepts and theorems which we now consider for three-dimensional vectors hold also for two-dimensional vectors.

In these theorems on vector algebra we need to bear in mind the fundamental laws of scalar algebra, which we considered in Section 1–4 as the characterizing relations of an integral domain. Those laws of scalars which we have immediate use for we list here, and we designate them by certain symbols for convenience in reference.

Commutative Laws

$$a + b = b + a \quad (SIa) \qquad ab = ba \quad (SIb)$$

Associative Laws

$$a + (b + c) = (a + b) + c \quad (SIIa) \qquad a(bc) = (ab)c \quad (SIIb)$$

Distributive Law

$$(a + b)c = ac + bc \quad (SIII)$$

The properties of a field as reflected in these laws, in conjunction with the definition of a vector in Section 2–5, imply certain relations satisfied by vectors. To some readers these relations may appear as obvious consequences of the definition of a vector; others may find the detailed exhibition of Theorem V instructive.

THEOREM V. *The addition of two vectors is commutative:*

$$\alpha + \beta = \beta + \alpha.$$

For

$$\begin{aligned} \alpha + \beta &= (a_1, a_2, a_3) + (b_1, b_2, b_3) \\ &= (a_1 + b_1, a_2 + b_2, a_3 + b_3) && \text{By (11)} \\ &= (b_1 + a_1, b_2 + a_2, b_3 + a_3) && \text{By } (SIa) \\ &= \beta + \alpha. && \text{By (11)} \end{aligned}$$

In like manner the following theorems may be established.

THEOREM VI. *The addition of vectors is associative:*

$$(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma).$$

THEOREM VII. *For any scalars r and s and any vectors α and β:*

(*i*) $(r + s)\alpha = r\alpha + s\alpha$. (*ii*) $r(s\alpha) = (rs)\alpha$.
(*iii*) $r(\alpha + \beta) = (r\alpha) + (r\beta)$. (*iv*) $0\alpha = O$. (*v*) $1\alpha = \alpha$.

2–7 The inner product of two vectors. While two scalars may be multiplied in only one way, it is not so with vectors. There are three conventional ways of combining two vectors α and β by processes called multiplication. One of these results in a scalar, the second in a vector, and the third in an entity more complex than a vector; the third will not be considered until we study matrices in a later chapter.

In the preceding sections we have made no mention of lengths of line vectors or of angles between line vectors. These cannot be defined in terms of the vector operations considered so far, rather they are expressed in terms of *inner products* of vectors, which we define below. We shall see that an expression for a length and an angle entails the square root of the sum of squares. While the operations of vector addition, of multiplication of a vector by a scalar, and of the inner product of two vectors hold for an arbitrary field, when we take the square root of the sum of squares trouble may arise. Therefore, when we deal with the topics of lengths, of angles, and of normalizing factors of vectors, we assume that we are limiting ourselves to vectors whose scalar coordinates are in the real field.

By the *inner product* (also known as the "dot product," the "scalar product," and the "direct product") $\alpha \cdot \beta$ of the two vectors $\alpha = (a_1, a_2, a_3)$ and $\beta = (b_1, b_2, b_3)$ with real coordinates, we mean the scalar quantity

$$\alpha \cdot \beta = a_1b_1 + a_2b_2 + a_3b_3. \tag{17}$$

Inner products have four important properties as expressed in the following theorems; these are consequences of the definition (17) and the laws of scalar algebra.

THEOREM VIII. *The inner product of two vectors is commutative:*

$$\alpha \cdot \beta = \beta \cdot \alpha.$$

For

$$\begin{aligned} \alpha \cdot \beta &= a_1b_1 + a_2b_2 + a_3b_3 && \text{By (17)} \\ &= b_1a_1 + b_2a_2 + b_3a_3 && \text{By } (SIb) \\ &= \beta \cdot \alpha. && \text{By (17)} \end{aligned}$$

In like manner, we may establish

THEOREM IX. *The product of a scalar and the inner product of two vectors is associative:* $k(\alpha \cdot \beta) = (k\alpha) \cdot \beta$.

THEOREM X. *The inner product of two vectors is distributive with respect to addition:* $(\alpha + \beta) \cdot \gamma = \alpha \cdot \gamma + \beta \cdot \gamma$.

For if $\alpha = (a_1, a_2, a_3)$ and $\beta = (b_1, b_2, b_3)$, then $\alpha + \beta = (a_1 + b_1, a_2 + b_2, a_3 + b_3)$. Consequently,

$$\begin{aligned} (\alpha + \beta) \cdot \gamma &= (a_1 + b_1)c_1 + (a_2 + b_2)c_2 + (a_3 + b_3)c_3 && \text{By (17)} \\ &= a_1c_1 + b_1c_1 + a_2c_2 + b_2c_2 + a_3c_3 + b_3c_3 && \text{By } (SIII) \\ &= a_1c_1 + a_2c_2 + a_3c_3 + b_1c_1 + b_2c_2 + b_3c_3 && \text{By } (SIa) \\ &= \alpha \cdot \gamma + \beta \cdot \gamma. && \text{By (17)} \end{aligned}$$

THEOREM XI. *The inner product of a non-null (real) vector α and itself is positive:* $\alpha \cdot \alpha > 0$.

Clearly $\alpha \cdot \alpha > 0$ if the coordinates are real, for $\alpha \cdot \alpha = a_1^2 + a_2^2 + a_3^2$.

The *magnitude* $|\alpha|$ of a vector α is the expression

$$|\alpha| = \sqrt{\alpha \cdot \alpha} = \sqrt{a_1^2 + a_2^2 + a_3^2}. \tag{18}$$

The magnitude of α is readily recognized as the length of $\mathbf{O}\alpha$. More generally the distance d between the points $\alpha = (a_1, a_2, a_3)$ and $\beta = (b_1, b_2, b_3)$ is given by

$$d^2 = |\beta - \alpha|^2 = (b_1 - a_1)^2 + (b_2 - a_2)^2 + (b_3 - a_3)^2. \tag{19}$$

THEOREM XII. *The cosine of the angle between two line vectors is equal to the inner product of the two vectors, divided by the product of their magnitudes.*

If α and β are any two vectors, then for the triangle with sides $\mathbf{O}\alpha$, $\mathbf{O}\beta$, and $\boldsymbol{\alpha\beta} = \beta - \alpha$, the trigonometric law of cosines gives

$$(20) \qquad |\beta - \alpha|^2 = |\alpha|^2 + |\beta|^2 - 2\,|\alpha| \cdot |\beta| \cos\theta.$$

However by (18) and Theorem X,

$$|\beta - \alpha|^2 = (\beta - \alpha) \cdot (\beta - \alpha) = \beta \cdot \beta - 2\beta \cdot \alpha + \alpha \cdot \alpha,$$

or

$$(21) \qquad |\beta - \alpha|^2 = |\beta|^2 - 2\beta \cdot \alpha + |\alpha|^2.$$

Combining (20) and (21) we get

$$|\alpha| \cdot |\beta| \cos\theta = \alpha \cdot \beta.$$

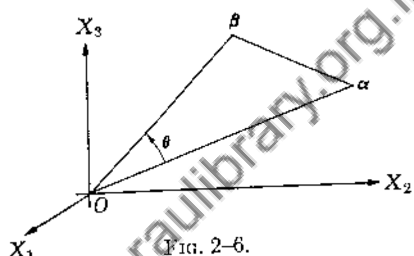

FIG. 2–6.

Consequently,

$$(22) \qquad \cos\theta = \frac{\alpha \cdot \beta}{|\alpha| \cdot |\beta|}.$$

In words, *the cosine of the angle between two line vectors* $\mathbf{O}\alpha$ *and* $\mathbf{O}\beta$ *is equal to the inner product of* α *and* β *divided by the product of the magnitudes of* α *and* β.

Let α and β be the directions of two lines, and recall that $\cos 90° = 0$. Then from (22), we have

THEOREM XIII. *The necessary and sufficient condition for two lines to be perpendicular is that the inner product of their directions shall be zero.*

EXERCISES

1. Prove that $(\alpha + \beta) \cdot (\gamma + \delta) = \alpha \cdot \gamma + \alpha \cdot \delta + \beta \cdot \gamma + \beta \cdot \delta$.
2. Verify that $\epsilon_1 \cdot \epsilon_2 = 0$, $\epsilon_1 \cdot \epsilon_3 = 0$, $\epsilon_2 \cdot \epsilon_3 = 0$.
3. Verify that $\epsilon_1 \cdot \epsilon_1 = 1$, $\epsilon_2 \cdot \epsilon_2 = 1$, $\epsilon_3 \cdot \epsilon_3 = 1$.
4. Using the results of Exs. 2 and 3, take the inner product of both sides of the vector equation $\alpha = a_1\epsilon_1 + a_2\epsilon_2 + a_3\epsilon_3$ with $\epsilon_1, \epsilon_2, \epsilon_3$ in turn, and show that $a_1 = \alpha \cdot \epsilon_1$, $a_2 = \alpha \cdot \epsilon_2$, $a_3 = \alpha \cdot \epsilon_3$. Thereby show that we may write any vector α in the form
$$\alpha = (\alpha \cdot \epsilon_1)\epsilon_1 + (\alpha \cdot \epsilon_2)\epsilon_2 + (\alpha \cdot \epsilon_3)\epsilon_3.$$
5. Find the angle between the lines $\mathbf{O}\alpha$ and $\mathbf{O}\beta$, where $\alpha = (6, 3, -2)$ and $\beta = (3, 5, -8)$.
6. Show that the lines with directions $\alpha = (-6, -3, -2)$ and $\beta = (-2, 6, -3)$ are perpendicular.

7. Show that the triangle with vertices $\alpha = (7, 3, 4)$, $\beta = (1, 0, 6)$, and $\gamma = (4, 5, -2)$ is a right isosceles triangle.

8. A unit vector is a vector whose magnitude is one. Prove that if α and β are unit vectors, then $\alpha \cdot \beta = \cos\theta$, where θ is the angle between $\mathbf{O}\alpha$ and $\mathbf{O}\beta$.

9. Using the results of Exs. 4 and 8, show that any unit vector α may be written in the form $\alpha = \epsilon_1 \cos\theta_1 + \epsilon_2 \cos\theta_2 + \epsilon_3 \cos\theta_3$, where θ_1, θ_2, θ_3 are the direction angles of $\mathbf{O}\alpha$. Alternately, $\alpha = (\cos\theta_1, \cos\theta_2, \cos\theta_3)$.

10. Using Ex. 8, derive the formula for $\cos(a - b)$. Let α and β be unit vectors in the X_1X_2 plane and making angles a and b with the positive X_1 axis. Then $\alpha = (\cos a, \sin a)$ and $\beta = (\cos b, \sin b)$. Using the formula $\cos\theta = \alpha \cdot \beta$, we get $\cos(a - b) = \cos a \cos b + \sin a \sin b$. Compare the ease of generality of this formula by vector methods with the procedure of establishing the generality of it by the conventional methods in trigonometry.

11. What is the magnitude of $\gamma = \alpha + \beta$, where $\alpha = (4, 5, 6)$ and $\beta = (2, 3, 7)$? Find the direction cosines of $\mathbf{O}\gamma$.

2–8 Rotation of axes. Invariants. Let $\mathbf{O}\epsilon_1$, $\mathbf{O}\epsilon_2$, $\mathbf{O}\epsilon_3$ and $\mathbf{O}\epsilon_1'$, $\mathbf{O}\epsilon_2'$, $\mathbf{O}\epsilon_3'$ be two rectangular cartesian coordinate systems, S and S', with a common origin, ϵ_1, ϵ_2, ϵ_3 and ϵ_1', ϵ_2', ϵ_3' being unit vectors. Denote the angle between $\mathbf{O}\epsilon_i$ and $\mathbf{O}\epsilon_j'$ by θ_{ij}. Then

$$
(23)\qquad
\begin{aligned}
\epsilon_1 \cdot \epsilon_1' &= \cos\theta_{11}, & \epsilon_2 \cdot \epsilon_1' &= \cos\theta_{21}, & \epsilon_3 \cdot \epsilon_1' &= \cos\theta_{31},\\
\epsilon_1 \cdot \epsilon_2' &= \cos\theta_{12}, & \epsilon_2 \cdot \epsilon_2' &= \cos\theta_{22}, & \epsilon_3 \cdot \epsilon_2' &= \cos\theta_{32},\\
\epsilon_1 \cdot \epsilon_3' &= \cos\theta_{13}, & \epsilon_2 \cdot \epsilon_3' &= \cos\theta_{23}, & \epsilon_3 \cdot \epsilon_3' &= \cos\theta_{33}.
\end{aligned}
$$

Note that θ_{11}, θ_{21}, θ_{31} are the direction angles of $\mathbf{O}\epsilon_1'$ relative to the S coordinate system, etc.; and θ_{11}, θ_{12}, θ_{13} are the direction angles of $\mathbf{O}\epsilon_1$ relative to the S' coordinate system, etc. Since there are six lines involved, there are twelve relations on their direction cosines, six relations from the sum of the squares of the direction cosines being equal to one, and six from the cosine of the angle between two perpendicular lines being zero.

Let the coordinates of α relative to the old coördinate system S be a_1, a_2, a_3 and the coordinates of this vector relative to the new coördinate system S' be a_1', a_2', a_3'. Clearly the sum of the projections of $\mathbf{O}\alpha$ on either set of coordinate axes must be the same; therefore

$$
(24)\qquad \alpha = a_1\epsilon_1 + a_2\epsilon_2 + a_3\epsilon_3
$$

and

$$
(25)\qquad \alpha = a_1'\epsilon_1' + a_2'\epsilon_2' + a_3'\epsilon_3'.
$$

If we take the inner product of both sides of the vector equation (25) by ϵ_1, ϵ_2, ϵ_3 respectively, we obtain (as in Ex. 4 of Section 2-7)

$$\begin{aligned} \epsilon_1 \cdot \alpha &= a_1'\epsilon_1 \cdot \epsilon_1' + a_2'\epsilon_1 \cdot \epsilon_2' + a_3'\epsilon_1 \cdot \epsilon_3', \\ \epsilon_2 \cdot \alpha &= a_1'\epsilon_2 \cdot \epsilon_1' + a_2'\epsilon_2 \cdot \epsilon_2' + a_3'\epsilon_2 \cdot \epsilon_3', \\ \epsilon_3 \cdot \alpha &= a_1'\epsilon_3 \cdot \epsilon_1' + a_2'\epsilon_3 \cdot \epsilon_2' + a_3'\epsilon_3 \cdot \epsilon_3'; \end{aligned} \tag{26}$$

r

$$\begin{aligned} a_1 &= a_1' \cos\theta_{11} + a_2' \cos\theta_{12} + a_3' \cos\theta_{13}, \\ a_2 &= a_1' \cos\theta_{21} + a_2' \cos\theta_{22} + a_3' \cos\theta_{23}, \\ a_3 &= a_1' \cos\theta_{31} + a_2' \cos\theta_{32} + a_3' \cos\theta_{33}. \end{aligned} \tag{27}$$

t is of interest to note that the relations (26), and equivalently (27), re obtained by direct use of vector algebra, without resort to geo-netric figures. Only in case one desires a geometric picture or in-erpretation of the scalar product of two vectors such as $\epsilon_1 \cdot \epsilon_2'$ as the osine of the angle between two unit vectors does one need to in-roduce geometry.

The vector α has magnitude $|\alpha|$ given by $|\alpha|^2 = a_1^2 + a_2^2 + a_3^2$ in he S coordinate system and the vector α' has the magnitude $|\alpha'|$ given by $|\alpha'|^2 = (a_1')^2 + (a_2')^2 + (a_3')^2$ in the S' coordinate system. By using the relations on the direction cosines of the coordinate axes and (27) it can be shown that

$$a_1^2 + a_2^2 + a_3^2 = (a_1')^2 + (a_2')^2 + (a_3')^2. \tag{28}$$

Therefore *the magnitude of a vector is a scalar invariant under the rotation of rectangular cartesian coordinate axes.*

More generally, let

$$\alpha = a_1\epsilon_1 + a_2\epsilon_2 + a_3\epsilon_3 = a_1'\epsilon_1' + a_2'\epsilon_2' + a_3'\epsilon_3'$$

and

$$\beta = b_1\epsilon_1 + b_2\epsilon_2 + b_3\epsilon_3 = b_1'\epsilon_1' + b_2'\epsilon_2' + b_3'\epsilon_3'$$

be two vectors expressed in terms of their coordinates in the S and S' reference systems. Corresponding to the relations (27) for the transformation of the a's, we have equations as

$$b_1 = b_1' \cos\theta_{11} + b_2' \cos\theta_{12} + b_3' \cos\theta_{13}$$

for the transformation of the b's. Proceeding for $\alpha \cdot \beta$ as we did above for $\alpha \cdot \alpha$, we can show that

$$a_1b_1 + a_2b_2 + a_3b_3 = a_1'b_1' + a_2'b_2' + a_3'b_3'. \tag{29}$$

That is, *the inner product of two vectors is an invariant under the rotation of rectangular cartesian coordinate axes.*

Since distances and angles are expressible in terms of inner products and magnitudes, it follows that these concepts are likewise invariant under rotation of axes.

7. Show that the triangle with vertices $\alpha = (7, 3, 4)$, $\beta = (1, 0, 6)$, and $\gamma = (4, 5, -2)$ is a right isosceles triangle.

8. A unit vector is a vector whose magnitude is one. Prove that if α and β are unit vectors, then $\alpha \cdot \beta = \cos\theta$, where θ is the angle between $\mathbf{O}\alpha$ and $\mathbf{O}\beta$.

9. Using the results of Exs. 4 and 8, show that any unit vector α may be written in the form $\alpha = \epsilon_1 \cos\theta_1 + \epsilon_2 \cos\theta_2 + \epsilon_3 \cos\theta_3$, where θ_1, θ_2, θ_3 are the direction angles of $\mathbf{O}\alpha$. Alternately, $\alpha = (\cos\theta_1, \cos\theta_2, \cos\theta_3)$.

10. Using Ex. 8, derive the formula for $\cos(a - b)$. Let α and β be unit vectors in the X_1X_2 plane and making angles a and b with the positive X_1 axis. Then $\alpha = (\cos a, \sin a)$ and $\beta = (\cos b, \sin b)$. Using the formula $\cos\theta = \alpha \cdot \beta$, we get $\cos(a - b) = \cos a \cos b + \sin a \sin b$. Compare the ease of generality of this formula by vector methods with the procedure of establishing the generality of it by the conventional methods in trigonometry.

11. What is the magnitude of $\gamma = \alpha + \beta$, where $\alpha = (4, 5, 6)$ and $\beta = (2, 3, 7)$? Find the direction cosines of $\mathbf{O}\gamma$.

2–8 Rotation of axes. Invariants. Let $\mathbf{O}\epsilon_1$, $\mathbf{O}\epsilon_2$, $\mathbf{O}\epsilon_3$ and $\mathbf{O}\epsilon_1'$, $\mathbf{O}\epsilon_2'$, $\mathbf{O}\epsilon_3'$ be two rectangular cartesian coordinate systems, S and S', with a common origin, $\epsilon_1, \epsilon_2, \epsilon_3$ and $\epsilon_1', \epsilon_2', \epsilon_3'$ being unit vectors. Denote the angle between $\mathbf{O}\epsilon_i$ and $\mathbf{O}\epsilon_j'$ by θ_{ij}. Then

$$
(23)\qquad
\begin{aligned}
\epsilon_1 \cdot \epsilon_1' &= \cos\theta_{11}, & \epsilon_2 \cdot \epsilon_1' &= \cos\theta_{21}, & \epsilon_3 \cdot \epsilon_1' &= \cos\theta_{31},\\
\epsilon_1 \cdot \epsilon_2' &= \cos\theta_{12}, & \epsilon_2 \cdot \epsilon_2' &= \cos\theta_{22}, & \epsilon_3 \cdot \epsilon_2' &= \cos\theta_{32},\\
\epsilon_1 \cdot \epsilon_3' &= \cos\theta_{13}, & \epsilon_2 \cdot \epsilon_3' &= \cos\theta_{23}, & \epsilon_3 \cdot \epsilon_3' &= \cos\theta_{33}.
\end{aligned}
$$

Note that $\theta_{11}, \theta_{21}, \theta_{31}$ are the direction angles of $\mathbf{O}\epsilon_1'$ relative to the S coordinate system, etc.; and $\theta_{11}, \theta_{12}, \theta_{13}$ are the direction angles of $\mathbf{O}\epsilon_1$ relative to the S' coordinate system, etc. Since there are six lines involved, there are twelve relations on their direction cosines, six relations from the sum of the squares of the direction cosines being equal to one, and six from the cosine of the angle between two perpendicular lines being zero.

Let the coordinates of α relative to the old coordinate system S be a_1, a_2, a_3 and the coordinates of this vector relative to the new coordinate system S' be a_1', a_2', a_3'. Clearly the sum of the projections of $\mathbf{O}\alpha$ on either set of coordinate axes must be the same; therefore

$$(24)\qquad \alpha = a_1\epsilon_1 + a_2\epsilon_2 + a_3\epsilon_3$$

and

$$(25)\qquad \alpha = a_1'\epsilon_1' + a_2'\epsilon_2' + a_3'\epsilon_3'.$$

If we take the inner product of both sides of the vector equation (25) by $\epsilon_1, \epsilon_2, \epsilon_3$ respectively, we obtain (as in Ex. 4 of Section 2–7)

$$
\begin{aligned}
\epsilon_1 \cdot \alpha &= a_1'\epsilon_1 \cdot \epsilon_1' + a_2'\epsilon_1 \cdot \epsilon_2' + a_3'\epsilon_1 \cdot \epsilon_3', \\
\epsilon_2 \cdot \alpha &= a_1'\epsilon_2 \cdot \epsilon_1' + a_2'\epsilon_2 \cdot \epsilon_2' + a_3'\epsilon_2 \cdot \epsilon_3', \qquad (26)\\
\epsilon_3 \cdot \alpha &= a_1'\epsilon_3 \cdot \epsilon_1' + a_2'\epsilon_3 \cdot \epsilon_2' + a_3'\epsilon_3 \cdot \epsilon_3';
\end{aligned}
$$

or

$$
\begin{aligned}
a_1 &= a_1' \cos\theta_{11} + a_2' \cos\theta_{12} + a_3' \cos\theta_{13}, \\
a_2 &= a_1' \cos\theta_{21} + a_2' \cos\theta_{22} + a_3' \cos\theta_{23}, \qquad (27)\\
a_3 &= a_1' \cos\theta_{31} + a_2' \cos\theta_{32} + a_3' \cos\theta_{33}.
\end{aligned}
$$

It is of interest to note that the relations (26), and equivalently (27), are obtained by direct use of vector algebra, without resort to geometric figures. Only in case one desires a geometric picture or interpretation of the scalar product of two vectors such as $\epsilon_1 \cdot \epsilon_2'$ as the cosine of the angle between two unit vectors does one need to introduce geometry.

The vector α has magnitude $|\alpha|$ given by $|\alpha|^2 = a_1^2 + a_2^2 + a_3^2$ in the S coordinate system and the vector α' has the magnitude $|\alpha'|$ given by $|\alpha'|^2 = (a_1')^2 + (a_2')^2 + (a_3')^2$ in the S' coordinate system. By using the relations on the direction cosines of the coordinate axes and (27) it can be shown that

$$a_1^2 + a_2^2 + a_3^2 = (a_1')^2 + (a_2')^2 + (a_3')^2. \qquad (28)$$

Therefore *the magnitude of a vector is a scalar invariant under the rotation of rectangular cartesian coordinate axes.*

More generally, let

$$\alpha = a_1\epsilon_1 + a_2\epsilon_2 + a_3\epsilon_3 = a_1'\epsilon_1' + a_2'\epsilon_2' + a_3'\epsilon_3'$$

and

$$\beta = b_1\epsilon_1 + b_2\epsilon_2 + b_3\epsilon_3 = b_1'\epsilon_1' + b_2'\epsilon_2' + b_3'\epsilon_3'$$

be two vectors expressed in terms of their coordinates in the S and S' reference systems. Corresponding to the relations (27) for the transformation of the a's, we have equations as

$$b_1 = b_1' \cos\theta_{11} + b_2' \cos\theta_{12} + b_3' \cos\theta_{13}$$

for the transformation of the b's. Proceeding for $\alpha \cdot \beta$ as we did above for $\alpha \cdot \alpha$, we can show that

$$a_1b_1 + a_2b_2 + a_3b_3 = a_1'b_1' + a_2'b_2' + a_3'b_3'. \qquad (29)$$

That is, *the inner product of two vectors is an invariant under the rotation of rectangular cartesian coordinate axes.*

Since distances and angles are expressible in terms of inner products and magnitudes, it follows that these concepts are likewise invariant under rotation of axes.

EXERCISES

1. Proceeding as in Section 2–8, obtain the formulas

$$a_1 = a_1' \cos\theta_{11} + a_2' \cos\theta_{12} \qquad \text{or} \qquad a_1 = a_1' \cos\theta - a_2' \sin\theta$$
$$a_2 = a_1' \cos\theta_{21} + a_2' \cos\theta_{22} \qquad \qquad a_2 = a_1' \sin\theta + a_2' \cos\theta$$

for the rotation of axes in two-dimensional analytics.

2. Using the result of Ex. 1, show that if $\alpha = (a_1, a_2)$ and $\beta = (b_1, b_2)$, then $\alpha \cdot \beta = a_1b_1 + a_2b_2$ is an invariant under the rotation of axes. Similarly, show that $a_1b_2 - a_2b_1$ is an invariant under such rotation.

3. Obtain the inverse of equations (27) by multiplying scalarly equation (24) by ϵ_1', ϵ_2', ϵ_3' (by this we mean to form the indicated inner products and thereby obtain scalar relations). Interpret the resulting coefficients of a_1, a_2, a_3 in terms of direction cosines.

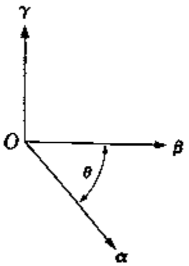

FIG. 2–7.

2–9 The vector product of two vectors. Let $\alpha = (a_1, a_2, a_3)$ and $\beta = (b_1, b_2, b_3)$ be two vectors such that $c\alpha \neq \beta$, that is, if α and β are direction vectors they represent different directions. Through O draw $\mathbf{O}\alpha$ and $\mathbf{O}\beta$; these lines determine a plane. Draw $\mathbf{O}\gamma \perp$ to this plane, where $\gamma = (c_1, c_2, c_3)$; here $\mathbf{O}\alpha$, $\mathbf{O}\beta$, and $\mathbf{O}\gamma$ are oriented so as to form a right-handed triple as indicated in Fig. 2–7. Since $\mathbf{O}\gamma$ is perpendicular to $\mathbf{O}\alpha$ and to $\mathbf{O}\beta$, we have, respectively,

$$\alpha \cdot \gamma = a_1c_1 + a_2c_2 + a_3c_3 = 0,$$

and

$$\beta \cdot \gamma = b_1c_1 + b_2c_2 + b_3c_3 = 0.$$

Solving these two homogeneous equations for c_1, c_2, c_3, we get *

(30) $$(c_1, c_2, c_3) = k\left(\begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}, \; -\begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}, \; \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}\right)$$

or

(31) $$\gamma = k(a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1).$$

The initial condition of the problem prohibits all of the coefficients of k from being zero. Also

$$|\gamma|^2 = c_1^2 + c_2^2 + c_3^2 = k^2[(a_1^2 + a_2^2 + a_3^2)(b_1^2 + b_2^2 + b_3^2) - (a_1b_1 + a_2b_2 + a_3b_3)^2]$$
$$= k^2[|\alpha|^2 |\beta|^2 - |\alpha|^2 |\beta|^2 \cos^2\theta],$$

* Here we are using the familiar second order determinant expression $\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$

or

(32) $$|\gamma|^2 = k^2 |\alpha|^2 |\beta|^2 \sin^2 \theta,$$

where θ is the smaller angle from $\mathbf{O}\alpha$ to $\mathbf{O}\beta$.

Since the magnitudes of α, β, γ and the angle θ are invariants under rotation of axes, it follows that k is also invariant. To specify uniquely the vector γ, we take $k = 1$ in (31); the vector so obtained is called the *vector product* of α and β. The vector product of α and β is denoted by $\alpha \times \beta$, and so

(33) $$\alpha \times \beta = \left(\begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}, \; -\begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix}, \; \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}\right)$$

or

(34) $$\alpha \times \beta = (a_2b_3 - a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1)$$

or

(35) $$\alpha \times \beta = (a_2b_3 - a_3b_2)\epsilon_1 + (a_3b_1 - a_1b_3)\epsilon_2 + (a_1b_2 - a_2b_1)\epsilon_3.$$

From (32) we see that the magnitude of $\gamma = \alpha \times \beta$ is given by

(36) $$|\alpha \times \beta| = |\alpha| \cdot |\beta| \sin \theta.$$

This magnitude of the vector product of α and β may be interpreted as the area of the parallelogram with $\mathbf{O}\alpha$ and $\mathbf{O}\beta$ for adjacent sides and θ the smaller angle between them.

Recalling that a determinant of the second order changes its sign if its rows are interchanged, we have directly from (33)

THEOREM XIV. *The vector product of two vectors is* not *commutative; for* $\alpha \times \beta = -\beta \times \alpha$.

Also from elementary determinant theory and (33), or otherwise by using (34), we can establish

THEOREM XV. *For any scalar k and any vectors* α, β, *and* γ,

$$(i) \quad (k\alpha) \times \beta = k(\alpha \times \beta) = \alpha \times (k\beta).$$
$$(ii) \quad \alpha \times (\beta + \gamma) = \alpha \times \beta + \alpha \times \gamma.$$
$$(iii) \quad (k\alpha) \times \alpha = O = (0, 0, 0).$$

THEOREM XVI. *If* $|\alpha\beta\gamma|$ *denotes the determinant of three three-dimensional vectors, as*

$$|\alpha\beta\gamma| = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}, \quad \textit{then} \quad |\alpha\beta\gamma| = \alpha \cdot (\beta \times \gamma) = \beta \cdot (\gamma \times \alpha) = \gamma \cdot (\alpha \times \beta).$$

If we solve equation (36) for $\sin \theta$, we get

(37) $$\sin \theta = \frac{|\alpha \times \beta|}{|\alpha| \cdot |\beta|}.$$

For emphasis and reference, we record this result as

THEOREM XVII. *The sine of the angle between two three-dimensional line vectors is equal to the magnitude of the vector product of these vectors, divided by the product of their magnitudes.*

From the definition of the magnitude of a vector, Theorem XV (*iii*) and Theorem XVII, there follows

THEOREM XVIII. *The necessary and sufficient condition for two line vectors* $\mathbf{O}\alpha$ *and* $\mathbf{O}\beta$ *to be parallel is for* α *and* β *to be proportional, that is,* $\alpha = k\beta$.

While the relation of this theorem to Theorem XVII is of interest, clearly the content of Theorem XVIII is not new, for we have in fact considered it in connection with direction numbers.

EXERCISES

1. Using the definition of the vector product of two vectors as given by formula (33), prove that $\epsilon_1 \times \epsilon_2 = \epsilon_3$, $\epsilon_2 \times \epsilon_3 = \epsilon_1$, $\epsilon_3 \times \epsilon_1 = \epsilon_2$; $\epsilon_1 \times \epsilon_1 = O$, $\epsilon_2 \times \epsilon_2 = O$, $\epsilon_3 \times \epsilon_3 = O$.

2. Using the results in Ex. 1, expand the product $\alpha \times \beta = (a_1\epsilon_1 + a_2\epsilon_2 + a_3\epsilon_3) \times (b_1\epsilon_1 + b_2\epsilon_2 + b_3\epsilon_3)$ and obtain (35).

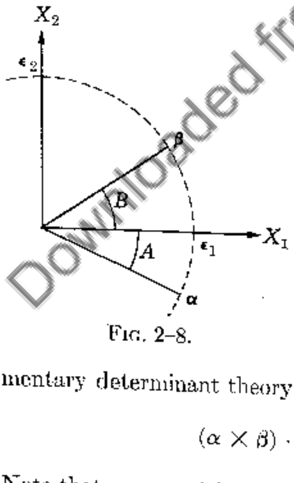

FIG. 2–8.

3. Show that $\alpha \times \beta = \begin{vmatrix} \epsilon_1 & \epsilon_2 & \epsilon_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$.

4. Let $\alpha = (a_1, a_2, 0)$ and $\beta = (b_1, b_2, 0)$ be two unit vectors in the X_1, X_2 plane and let A and B be the angles they make with the X_1 axis, as shown in Fig. 2–8. Then $\alpha = \epsilon_1 \cos A - \epsilon_2 \sin A$ and $\beta = \epsilon_1 \cos B + \epsilon_2 \sin B$. Use $\sin (A + B)\epsilon_3 = \alpha \times \beta$ to derive the formula for $\sin (A + B)$.

5. Use the definition of the vector product of two vectors as given by (33) and elementary determinant theory to prove that

$$(\alpha \times \beta) \cdot (\gamma \times \delta) = \begin{vmatrix} \alpha \cdot \gamma & \alpha \cdot \delta \\ \beta \cdot \gamma & \beta \cdot \delta \end{vmatrix}.$$

Note that as a special case of this result

$$(\alpha \times \beta) \cdot (\alpha \times \beta) = \begin{vmatrix} \alpha \cdot \alpha & \alpha \cdot \beta \\ \alpha \cdot \beta & \beta \cdot \beta \end{vmatrix}.$$

More particularly, for two-dimensional vectors $\alpha = (a_1, a_2)$ and $\beta = (b_1, b_2)$, the latter becomes

$$|\alpha\beta| \cdot |\alpha\beta| = \begin{vmatrix} \alpha \cdot \alpha & \alpha \cdot \beta \\ \alpha \cdot \beta & \beta \cdot \beta \end{vmatrix}, \qquad \text{where } |\alpha\beta| = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}.$$

6. Observe that for two two-dimensional vectors the formula (37) becomes

$$\sin\theta = \frac{|\alpha\beta|}{|\alpha| \cdot |\beta|}.$$

Show as a consequence of this relation that the condition for parallelism, as expressed in Theorem XVIII, holds for two-dimensional vectors as well as for three-dimensional vectors.

7. Show that if $\alpha = (a_1, a_2)$ and $\beta = (b_1, b_2)$, then

$$|s_1\alpha + s_2\beta, \quad t_1\alpha + t_2\beta| = \begin{vmatrix} s_1a_1 + s_2b_1, & s_1a_2 + s_2b_2 \\ t_1a_1 + t_2b_1, & t_1a_2 + t_2b_2 \end{vmatrix} = \begin{vmatrix} s_1 & s_2 \\ t_1 & t_2 \end{vmatrix} \cdot \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$$
$$= (s_1t_2 - s_2t_1)|\alpha\beta|.$$

8. If α, β, γ, δ are three-dimensional vectors, prove that

$$(s_1\alpha + s_2\beta) \times (t_1\gamma + t_2\delta) = s_1t_1(\alpha \times \gamma) + s_1t_2(\alpha \times \delta) + s_2t_1(\beta \times \gamma) + s_2t_2(\beta \times \delta).$$

(Use Theorem XV(*ii*).)

9. As a special case of Ex. 8, show that

$$(s_1\alpha + s_2\beta) \times (t_1\alpha + t_2\beta) = (s_1t_2 - s_2t_1)(\alpha \times \beta).$$

10. Prove that if α, β, γ are three-dimensional vectors, then

$$(s_1\alpha + s_2\beta + s_3\gamma) \times (t_1\alpha + t_2\beta + t_3\gamma) = (s_2t_3 - s_3t_2)(\beta \times \gamma)$$
$$+ (s_3t_1 - s_1t_3)(\gamma \times \alpha) + (s_1t_2 - s_2t_1)(\alpha \times \beta).$$

CHAPTER 3

VECTOR METHODS IN GEOMETRY; LINEAR DEPENDENCE OF VECTORS

3–1 Vector relations independent of the origin. Suppose $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_m$ are position vectors with respect to some coordinate system with origin $O = (0, 0, 0)$, and consider the linear relation

(1) $$k_1\alpha_1 + k_2\alpha_2 + k_3\alpha_3 + \cdots + k_m\alpha_m = O,$$

where the k's are scalar coefficients. Let $\alpha_1', \alpha_2', \alpha_3', \ldots, \alpha_m'$ be corresponding position vectors with respect to a new origin O' whose position vector relative to O is ξ. Then

(2) $$\alpha_1 = \alpha_1' + \xi, \quad \alpha_2 = \alpha_2' + \xi, \quad \ldots, \quad \alpha_m = \alpha_m' + \xi.$$

Substituting from (2) in (1) we get

(3) $$k_1\alpha_1' + k_2\alpha_2' + k_3\alpha_3' + \cdots + k_m\alpha_m' + (k_1 + k_2 + k_3 + \cdots + k_m)\xi = O.$$

If the relation (1) is to be true for vectors independent of the origin, that is, if we are to have

(4) $$k_1\alpha_1' + k_2\alpha_2' + k_3\alpha_3' + \cdots + k_m\alpha_m' = O,$$

then from (3) the condition

(5) $$k_1 + k_2 + k_3 + \cdots + k_m = 0$$

must be satisfied. The condition (5) is therefore necessary; it should be clear that it is also sufficient. So we have established

THEOREM I. *The necessary and sufficient condition that a linear relation connecting any number of position vectors shall be independent of the origin is that the algebraic sum of the scalar coefficients of the vectors be zero.*

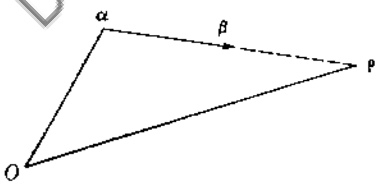

FIG. 3–1.

3–2 Vector equation of a line. Consider the problem of finding the equation of the line (Fig. 3–1) through the given point $\alpha = (a_1, a_2, a_3)$ and having the direction of the given line vector $\beta = (b_1, b_2, b_3)$. Let the *variable vector* $\rho = (x_1, x_2, x_3)$ be any point on the line. Then the line vector $\alpha - \rho$ is parallel to β, and so by Theorem XVIII of Chapter 2, for some scalar variable s which takes on all values between plus and minus infinity, we have

$$\rho - \alpha = s\beta \qquad \text{or} \qquad \rho = \alpha + s\beta. \tag{6}$$

As a special case of (6) we have for the equation of the line through the origin $O = (0, 0, 0)$ parallel to the line vector β,

$$\rho = s\beta. \tag{7}$$

To find the vector equation of the line passing through the fixed points α and β we observe that the line has for its direction $\beta - \alpha$. So either by Theorem XVIII of Chapter 2, or as an application of (6), we have for the desired equation

$$\rho - \alpha = s(\beta - \alpha) \qquad \text{or} \qquad \rho = (1 - s)\alpha + s\beta. \tag{8}$$

This equation may be put in the form

$$\rho = (1 - s)\alpha - s\beta = O, \tag{9}$$

from which it is seen that the sum of the coefficients of the vectors is zero, as it should be. For (9) expresses the condition that the position vectors ρ, α, and β be collinear, a property independent of the origin. This leads us to

THEOREM II. *The necessary and sufficient condition for three points in three-dimensional space to be collinear is that there exist a linear relation on their position vectors in which the algebraic sum of the scalar coefficients is equal to zero.*

We have just proved that it is a necessary condition. The condition is also sufficient; for assuming that α_1, α_2, α_3 satisfy the relation

$$k_1\alpha_1 + k_2\alpha_2 + k_3\alpha_3 = O \quad \text{such that} \quad k_1 + k_2 + k_3 = 0, \tag{10}$$

we may make the coefficient of α_1 unity by dividing by k_1 $(k_1 \neq 0)$, getting

$$\alpha_1 + \frac{k_2}{k_1}\alpha_2 + \frac{k_3}{k_1}\alpha_3 = O.$$

Under the appropriate transformation of the coefficients this may be put in the form (9), showing that α_1 is a point on the line $\alpha_2\alpha_3$.

3–3 The point ρ which divides the segment $\alpha\beta$ in a given ratio. Let the ratio of the length of the segment from α to ρ and that of the segment from ρ to β be equal to r/s. Then

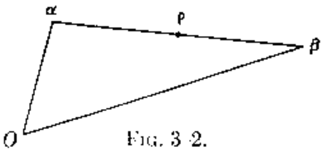

Fig. 3-2.

$$s(\rho - \alpha) = r(\beta - \rho),$$

whence

$$(11) \qquad \rho = \frac{s\alpha + r\beta}{s + r}.$$

It is noteworthy that equations (6), (7), (8), (9), and (11) hold for two- and three-dimensional vectors.

Exercises

1. Use Theorem I to show that the property expressed by (11) is independent of the origin.

2. Obtain from (8) the corresponding scalar equations for $n = 2$ and for $n = 3$. *Hint:* Equate corresponding scalar coordinates.

3. Obtain from (6), (7), and (11) the corresponding scalar equations for $n = 2$ and $n = 3$.

4. Find the position vector dividing the segment joining $\alpha = (3, 6, 7)$ and $\beta = (8, 9, 13)$ in the ratio 2:3.

3–4 Vector equation of a plane. The equation of the plane through the fixed point α and parallel to the nonequivalent line-vectors $\mathbf{O}\beta$ and $\mathbf{O}\gamma$ is

$$(12) \qquad \rho = \alpha + s\beta + t\gamma,$$

where s and t are scalar variables assuming all values between plus and minus infinity. For since ρ is any variable point on the plane, $\rho - \alpha$ is in the plane determined by β and γ; therefore $\rho - \alpha$ can be resolved into components parallel to β and γ; from this it follows that $\rho - \alpha = s\beta + t\gamma$, and (12) is a modification of this relation. As a special case of (12) we have that the equation of the plane through the origin and parallel to β and γ is

$$(13) \qquad \rho = s\beta + t\gamma.$$

To find the vector equation of the plane passing through the fixed points α, β, and γ we note that the plane is parallel to $\beta - \alpha$ and $\gamma - \alpha$. So from (12), we have

$$(14) \qquad \rho = \alpha + s(\beta - \alpha) + t(\gamma - \alpha)$$

or

$$(15) \qquad (1 - s - t)\alpha + s\beta + t\gamma - \rho = O.$$

In this equation the sum of the scalar coefficients of the vectors α, β, γ, and ρ is zero. Analogous to Theorem II, we have

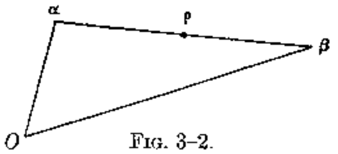

FIG. 3–2.

$$s(\rho - \alpha) = r(\beta - \rho),$$

whence

(11) $$\rho = \frac{s\alpha + r\beta}{s + r}.$$

It is noteworthy that equations (6), (7), (8), (9), and (11) hold for two- and three-dimensional vectors.

EXERCISES

1. Use Theorem I to show that the property expressed by (11) is independent of the origin.

2. Obtain from (8) the corresponding scalar equations for $n = 2$ and for $n = 3$. *Hint:* Equate corresponding scalar coordinates.

3. Obtain from (6), (7), and (11) the corresponding scalar equations for $n = 2$ and $n = 3$.

4. Find the position vector dividing the segment joining $\alpha = (3, 6, 7)$ and $\beta = (8, 9, 13)$ in the ratio 2:3.

3–4 Vector equation of a plane. The equation of the plane through the fixed point α and parallel to the inequivalent line vectors $\mathbf{O}\beta$ and $\mathbf{O}\gamma$ is

(12) $$\rho = \alpha + s\beta + t\gamma,$$

where s and t are scalar variables assuming all values between plus and minus infinity. For since ρ is any variable point on the plane, $\rho - \alpha$ is in the plane determined by β and γ; therefore $\rho - \alpha$ can be resolved into components parallel to β and γ; from this it follows that $\rho - \alpha = s\beta + t\gamma$, and (12) is a modification of this relation. As a special case of (12) we have that the equation of the plane through the origin and parallel to β and γ is

(13) $$\rho = s\beta + t\gamma.$$

To find the vector equation of the plane passing through the fixed points α, β, and γ we note that the plane is parallel to $\beta - \alpha$ and $\gamma - \alpha$. So from (12), we have

(14) $$\rho = \alpha + s(\beta - \alpha) + t(\gamma - \alpha)$$

or

(15) $$(1 - s - t)\alpha + s\beta + t\gamma - \rho = O.$$

In this equation the sum of the scalar coefficients of the vectors α, β, γ, and ρ is zero. Analogous to Theorem II, we have

THEOREM III. *The necessary and sufficient condition for four points in three-space to be coplanar is that there exist a linear relation on their position vectors in which the algebraic sum of the scalar coefficients is equal to zero.*

A symmetrical form of equation (15) may be obtained by assuming a linear relation on the vectors α, β, γ, ρ of the form

$$k_1\alpha + k_2\beta + k_3\gamma + k_4\rho = O \quad \text{where} \quad k_1 + k_2 + k_3 + k_4 = 0. \tag{16}$$

Eliminating k_4, we get

$$k_1(\rho - \alpha) + k_2(\rho - \beta) + k_3(\rho - \gamma) = O. \tag{17}$$

To derive the equation of the plane through α and perpendicular to $O\beta$, we recognize that for any variable point ρ in the plane, $\rho - \alpha$ is perpendicular to the direction β. So by Theorem XIII of Chapter 2 the desired equation is

$$(\rho - \alpha) \cdot \beta = 0. \tag{18}$$

EXERCISES

1. Write the scalar form of each of the equations (12), (13), (14), and (18), knowing that for the plane $n = 3$.

2. For $n = 2$ equation (18) is the equation of a straight line. Obtain the scalar form of this, and reconcile the latter with the normal form of the equation of a straight line as conventionally given in texts on plane analytics.

3. Find the equation ($n = 3$) of the sphere with center at α and radius equal to k.

4. Give both the vector and the scalar form of the equation of Ex. 3 for $n = 2$.

5. Give a vector proof of the theorem that the diagonals of a parallelogram bisect each other. *Hint:* Let α, β, γ, δ be the vertices and let ρ be the point of intersection of the diagonals; use Theorem II to write four linear expressions for ρ.

3-5 Linear dependence of vectors. We have previously spoken of two vectors α and β as being proportional if $\alpha = k\beta$, for some scalar k. Sometimes it is more convenient to say equivalently that α and β are proportional if for some scalars s_1 and s_2 we have

$$s_1\alpha + s_2\beta = O. \tag{19}$$

This is merely a symmetrical form of $\alpha = k\beta$.

We now consider linear dependence, which may be regarded as a generalization of the concept of proportionality. First, a word about linear dependence of scalars. The n scalars $k_1, k_2, \ldots, k_n$ are said to be *linearly dependent* with respect to a field **F** if n scalar members of the field **F** $s_1, s_2, \ldots, s_n$, not all zero, exist such that

$$s_1k_1 + s_2k_2 + \cdots + s_nk_n = 0. \tag{20}$$

If no scalars in the field **F** exist such that the condition (20) is satisfied, then the scalars $k_1, k_2, \ldots, k_n$ are said to be *linearly independent* with respect to that field. To say whether given scalars are linearly independent or linearly dependent with respect to a field **F**, it is necessary to specify the field. To illustrate, consider the numbers $k_1 = 1$, $k_2 = \sqrt{2}$; they are linearly independent relative to the rational field but they are linearly dependent relative to the real field, because $s_1 = 2$, $s_2 = -\sqrt{2}$ and the given k's will satisfy (20). Similarly, 1 and i $(i = \sqrt{-1})$ are linearly independent relative to the rational field or the real field, but are linearly dependent relative to the complex field.

A vector α is said to belong to a field **F** if, and only if, all its scalar coordinates or elements belong to the field **F**. In particular, if all of the scalar coordinates of α are ordinary complex numbers, we say α is a *complex vector;* if all of the scalar coordinates of α are real numbers, we call α a *real vector.* The set of all n-dimensional vectors with elements in a given field **F** is called a vector space $V_n(\mathbf{F})$. In ordinary solid analytic geometry we give emphasis to the study of $V_3(\mathbf{R})$.

The vectors $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_m$ in a field **F** are said to be *linearly dependent* (in **F**) if m scalars $s_1, s_2, s_3, \ldots, s_m$ in the field **F**, not all zero, exist such that

$$s_1\alpha_1 + s_2\alpha_2 + s_3\alpha_3 + \cdots + s_m\alpha_m = O. \tag{21}$$

If no such scalars exist, the vectors are said to be *linearly independent* (in **F**). Linear dependence and linear independence are basically properties of sets of vectors; however it is usual to apply these descriptive terms to vectors themselves, as well as to sets of vectors; we may speak of "a set of linearly independent vectors" or of "a linearly independent set of vectors."

In (21) at least one of the s's is different from zero. Suppose it is s_1; then we can solve this relation for α_1, and write

$$\alpha_1 = k_2\alpha_2 + k_3\alpha_3 + \cdots + k_m\alpha_m. \tag{22}$$

Then we say α_1 is a *linear combination* of $\alpha_2, \alpha_3, \ldots, \alpha_m$.

THEOREM IV. *Two two-dimensional (position) vectors are linearly dependent if and only if they are collinear with the origin.*

Theorem IV is simply a restatement of Theorem XVIII of Chapter 2, with applicability of the latter to two-dimensional vectors. As previously stated (see Ex. 6, Section 2-9), that theorem applies alike to two- and three-dimensional vectors. Theorem IV may be stated perhaps more emphatically by saying that $O = (0, 0)$ lies on the line through the points $\alpha = (a_1, a_2)$ and $\beta = (b_1, b_2)$ when two scalar multipliers r and s can be found so that

$$r\alpha + s\beta = O. \tag{23}$$

When no linear combination of the vectors α and β is the zero vector, they are linearly independent.

THEOREM V. *Any three two-dimensional vectors are linearly dependent.*

To prove this we note first that if all three of the points are collinear with the origin, they are linearly dependent, by Theorem IV. On the other hand, let the vectors be $\alpha = (a_1, a_2)$, $\beta = (b_1, b_2)$ and $\gamma = (c_1, c_2)$, and assume that some two of them, say α and β, are not collinear with the origin. Then $|\alpha\beta| \neq 0$, and it is sufficient to show that we can find r and s such that

$$\gamma = r\alpha + s\beta, \tag{24}$$

or

$$\begin{cases} c_1 = ra_1 + sb_1 \\ c_2 = ra_2 + sb_2. \end{cases} \tag{25}$$

Solving (25) for r and s, we get

$$r = \frac{|\gamma\beta|}{|\alpha\beta|}, \quad s = \frac{|\alpha\gamma|}{|\alpha\beta|}. \tag{26}$$

Consequently, we have

$$|\alpha\beta|\,\gamma = |\gamma\beta|\,\alpha + |\alpha\gamma|\,\beta, \tag{27}$$

and the theorem is proved.

It is sometimes preferable to state Theorem V equivalently in the form of

THEOREM VI. *Any plane vector is a linear combination of two linearly independent plane vectors, that is*

$$\gamma = r\alpha + s\beta, \tag{28}$$

where α and β are two given linearly independent vectors and r and s are scalar multipliers.

Since every plane vector is a linear combination of any two linearly independent plane vectors, it is customary to say that any two linearly independent plane vectors form a *basis* for all vectors in the plane. While any two linearly independent two-dimensional vectors form such a basis, it is usually particularly convenient to take as such basis the unit vectors

(29) $$\epsilon_1 = (1, 0) \quad \text{and} \quad \epsilon_2 = (0, 1),$$

for as we have noted before, any vector $\alpha = (a_1, a_2)$ is given as a linear combination of these unit vectors by

(30) $$\alpha = a_1\epsilon_1 + a_2\epsilon_2.$$

As we observed in Chapter 2 and also in connection with (7) above, there does not exist a single pair of linearly independent vectors on a line through the origin, for every vector on the line is the product of any nonzero vector on the line by a scalar s. That is, the maximum number of linearly independent vectors on a line is one, and as we have just seen, the maximum number of linearly independent vectors in a plane is two. As we shall see later, this maximum number of linearly independent vectors which characterizes a given space is called the *dimension* of that space.

THEOREM VII. *Two three-dimensional vectors are linearly dependent when and only when they are collinear with the origin.*

Theorem VII is merely a restatement of Theorem XVIII of Chapter 2 in the terminology of linear dependence. This theorem says that $O = (0, 0, 0)$ lies on the line through the points $\alpha = (a_1, a_2, a_3)$ and $\beta = (b_1, b_2, b_3)$ when two scalar multipliers r and s can be found so that

(31) $$r\alpha + s\beta = O.$$

When no linear combination of the vectors α and β is the zero vector, they are linearly independent.

THEOREM VIII. *Three three-dimensional vectors are linearly dependent when and only when they are coplanar with the origin.*

For we see from equation (13) that γ is a linear combination of α and β of the form $\gamma = r\alpha + s\beta$ when and only when γ is in the plane determined by the origin, α, and β.

From Theorem VIII and elementary determinant theory there follows immediately

THEOREM IX. *A necessary condition for three points α, β, and γ to be coplanar with the origin is that $|\alpha\beta\gamma| = 0$.*

THEOREM X. *Any four three-dimensional vectors are linearly dependent.*

The proof of Theorem X is analogous to that for Theorem V. Clearly, if all four of the points are coplanar with the origin, they are linearly dependent. Considering the general case, let the vectors be $\alpha = (a_1, a_2, a_3)$, $\beta = (b_1, b_2, b_3)$, $\gamma = (c_1, c_2, c_3)$, and $\delta = (d_1, d_2, d_3)$. Assume that some three of these (position) vectors, say α, β, and γ, are not coplanar with the origin; then $|\alpha\beta\gamma| \neq 0$. It is sufficient to show that we can find scalars s_1, s_2, s_3 such that

(32) $$\delta = s_1\alpha + s_2\beta + s_3\gamma$$

or

(33) $$\begin{cases} d_1 = s_1a_1 + s_2b_1 + s_3c_1 \\ d_2 = s_1a_2 + s_2b_2 + s_3c_2 \\ d_3 = s_1a_3 + s_2b_3 + s_3c_3. \end{cases}$$

Solving (33) for s_1, s_2, s_3, we get

(34) $$s_1 = \frac{|\delta\beta\gamma|}{|\alpha\beta\gamma|}, \quad s_2 = \frac{|\alpha\delta\gamma|}{|\alpha\beta\gamma|}, \quad s_3 = \frac{|\alpha\beta\delta|}{|\alpha\beta\gamma|},$$

and consequently the linear combination (32) may be written

(35) $$|\alpha\beta\gamma|\,\delta = |\delta\beta\gamma|\,\alpha + |\alpha\delta\gamma|\,\beta + |\alpha\beta\delta|\,\gamma,$$

and the theorem is proved.

It is sometimes desirable to emphasize the last theorem in the form of

THEOREM XI. *Any space vector is a linear combination of any three linearly independent space vectors.*

Since every ordinary three-space vector is a linear combination of any three linearly independent space vectors, we say that any three linearly independent space vectors form a basis for all vectors in that space. It is particularly convenient to take as a basis for the space vectors the unit vectors

(36) $$\epsilon_1 = (1, 0, 0), \quad \epsilon_2 = (0, 1, 0), \quad \epsilon_3 = (0, 0, 1),$$

for we saw in Section 2–5 that any vector $\alpha = (a_1, a_2, a_3)$ is given as a linear combination of these unit vectors by

(37) $$\alpha = a_1\epsilon_1 + a_2\epsilon_2 + a_3\epsilon_3.$$

Theorem XI emphasizes that the maximum number of linearly independent vectors in ordinary three-space is three; that is precisely why we term it "three-space."

Exercises

1. Determine which of the following pairs of vectors are linearly dependent; in case a pair is linearly dependent, give values of the scalar multipliers r and s which relate the dependent vectors in the manner of relation (25).

(i) $\alpha = (1, -2)$, $\beta = (-3, 6)$. (ii) $\alpha = (1, -2)$, $\beta = (3, 6)$.
(iii) $\alpha = (1, -2, 3)$, $\beta = (4, -8, 12)$. (iv) $\alpha = (1, -2, 3)$, $\beta = (-5, 10, 7)$.

2. For what values of k will the following vectors be linearly dependent?

(i) $\alpha = (2, k)$, $\beta = (3, 2k + 1)$. (ii) $\alpha = (2, k, 6)$, $\beta = (3, k + 1, 8)$.

3. Determine which of the following triads of vectors are linearly dependent; in case a triad is linearly dependent, give values of the scalar multipliers r and s which relate the dependent vectors in the manner of equation (24).

(i) $\alpha = (3, 4)$, $\beta = (-5, 3)$, $\gamma = (9, -17)$. (ii) $\alpha = (3, 0)$, $\beta = (-5, 0)$, $\gamma = (9, 0)$.
(iii) $\alpha = (2, 2, 0)$, $\beta = (0, 2, -1)$, $\gamma = (4, 2, 1)$. (iv) $\alpha = (0, 0, 3)$, $\beta = (0, 4, 0)$, $\gamma = (8, 0, 0)$.

4. Find scalar multipliers r, s, and t which enable us to write the vector $\delta = (1, -1, 6)$ as a linear combination of $\alpha = (3, 2, 1)$, $\beta = (-4, -3, 1)$, and $\gamma = (2, 1, 1)$ in the form $\delta = r\alpha + s\beta + t\gamma$.

CHAPTER 4

VECTORS OF n DIMENSIONS

4-1 Fundamental definitions and axioms. Ordered n-tuples of scalars in a field **F**, as $\alpha = (a_1, a_2, \ldots, a_n)$, $\beta = (b_1, b_2, \ldots, b_n)$, $\ldots$, $\gamma = (c_1, c_2, \ldots, c_n)$, which obey certain rules of combination are called *vectors of n dimensions.* The basic operations of the algebra of such vectors are the *multiplication of a vector α by a scalar k:*

$$(1) \qquad k\alpha = (ka_1, ka_2, \ldots, ka_n)$$

and the *addition of two vectors α and β:*

$$(2) \qquad \alpha + \beta = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n).$$

For any positive integer n and any field **F**, the set of all the n-tuples of scalars of **F** which obey the laws of combination (1) and (2) is called the *vector space* $V_n(\mathbf{F})$ over the field **F**. Some writers refer to such a vector space as a "linear vector space," others as a "linear space."

From (2) it follows that there is a unique *null vector* $O = (0, 0, \ldots, 0)$ with the property that for any vector α,

$$(3) \qquad \alpha + O = O + \alpha = \alpha$$

for all α. Also

$$(4) \qquad 0 \cdot \alpha = O \text{ for all } \alpha, \qquad \text{and} \qquad k \cdot O = O \text{ for all } k.$$

Here, as previously for two- and three-dimensional vectors, care should be taken to avoid confusing the null vector O and the zero scalar 0. Two vectors $\alpha = (a_1, a_2, \ldots, a_n)$ and $\beta = (b_1, b_2, \ldots, b_n)$ are *equal* if corresponding scalar elements are equal. The equality of two such vectors implies the satisfying of n scalar equations.

If α is any non-null vector, we shall denote by $-\alpha$ the vector $(-1)\alpha$. From this convention and the definition of O, it follows that

$$(5) \qquad \alpha + (-1)\alpha = O.$$

Relative to the operation of addition of vectors as defined by (2), *the set of vectors $V_n(\mathbf{F})$ constitute a group* with the null vector O as the identity element of the group and $(-1)\alpha$ as the group inverse of any vector α.

4–2 Algebraic properties of vectors. The above definition of n-dimensional vectors, with the properties of a field as stated in Section 2–6, implies certain relations satisfied by vectors.

THEOREM I. *The addition of vectors is commutative:* $\alpha + \beta = \beta + \alpha$.

For we have

$$\alpha + \beta = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n) \quad \text{By (2).}$$
$$= (b_1 + a_1, b_2 + a_2, \ldots, b_n + a_n) \quad \text{By } (SIa).$$

So

$$\alpha + \beta = \beta + \alpha.$$

Similarly, we may establish the following two theorems.

THEOREM II. *The addition of vectors is associative:*

$$(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma).$$

THEOREM III. *For any scalars r and s and any vectors* α *and* β:

(*i*) $(r + s)\alpha = r\alpha + s\alpha$. (*ii*) $r(s\alpha) = (rs)\alpha$.
(*iii*) $r(\alpha + \beta) = (r\alpha) + (r\beta)$. (*iv*) $0 \cdot \alpha = O$. (*v*) $1 \cdot \alpha = \alpha$.

4–3 Vector subspaces. Union and intersection of vector spaces. A subset of the vectors of $V_n(\mathbf{F})$ which is closed with respect to the two basic operations of vector algebra (the multiplication of a vector by a scalar and the addition of two vectors) is called a *vector subspace* of $V_n(\mathbf{F})$, and is designated by ${}_1V_n(\mathbf{F})$. Otherwise stated, ${}_1V$ is a vector subspace of V if α and β are vectors of ${}_1V$ and if every linear combination $r\alpha + s\beta$ is also a vector of ${}_1V$, r and s being scalars of the underlying field $\mathbf{F}$. In general, we write ${}_1V \subseteq V$, meaning that the vector subspace ${}_1V$ may or may not contain all the vectors of V. If not every vector of V is in ${}_1V$, then ${}_1V$ is a *proper vector subspace* of V, and we write ${}_1V \subset V$. The analogy with previously considered concepts of subgroups and subfields should be apparent. Some writers refer to a vector subspace as a linear manifold.

A basic observation is that if a vector subspace contains a vector α, it also contains $\alpha - \alpha = O$. Therefore, if vector subspaces are interpreted as lines, planes, etc., we must consider only those lines, planes, etc., which pass through the origin; as a degenerate case the null vector alone is a subspace of any vector space. Vectors of the form $(a_1, a_2, 0)$ constitute a subspace, the X_1X_2 plane of ordinary three-space; those of the form $(a_1, a_2, 0, a_4)$ constitute a subspace of $V_4(\mathbf{F})$.

If ${}_1V$ and ${}_2V$ are vector subspaces of V, we define their *union*

${}_1V \smile {}_2V$ to be the set of all vectors $\alpha + \beta$ where α is in ${}_1V$ and β is in ${}_2V$.

THEOREM IV. *The union ${}_1V \smile {}_2V$ of any two subspaces of a vector space V is itself a vector subspace of V.*

For if γ_1 and γ_2 are any two vectors of ${}_1V \smile {}_2V$, then $\gamma_1 = \alpha_1 + \beta_1$, $\gamma_2 = \alpha_2 + \beta_2$, where α_1 and α_2 are in ${}_1V$, and β_1 and β_2 are in ${}_2V$. Then for any two scalars r and s of the underlying field, we have $r\gamma_1 + s\gamma_2 = (r\alpha_1 + s\alpha_2) + (r\beta_1 + s\beta_2)$. Since ${}_1V$ is a vector space, $r\alpha_1 + s\alpha_2$ is in ${}_1V$, and since ${}_2V$ is a vector space, $r\beta_1 + s\beta_2$ is in ${}_2V$. Therefore, $r\gamma_1 + s\gamma_2$ is in ${}_1V \smile {}_2V$.

If ${}_1V$ and ${}_2V$ are vector subspaces of V, we define their *intersection* to be the set ${}_1V \frown {}_2V$ of those vectors which are in both ${}_1V$ and ${}_2V$. Reasoning as for Theorem IV, we may prove

THEOREM V. *The intersection of any two subspaces of a vector space V is itself a subspace of V.*

Let us illustrate *union* and *intersection* of vector subspaces in ordinary three-space. Recall the familiar unit vectors $\epsilon_1 = (1, 0, 0)$, $\epsilon_2 = (0, 1, 0)$, $\epsilon_3 = (0, 0, 1)$, and let ${}_iV$ $(i = 1, 2, 3)$ be the one-space $k_i\epsilon_i$. That is,

$$ {}_1V = (k_1, 0, 0), \quad {}_2V = (0, k_2, 0), \quad {}_3V = (0, 0, k_3); $$

these one-spaces are respectively the X_1, X_2, X_3 axes. Then ${}_1V \smile {}_2V$ is the X_1X_2 plane; ${}_2V \smile {}_3V$ is the X_2X_3 plane; ${}_3V \smile {}_1V$ is the X_3X_1 plane; ${}_1V \smile {}_2V \smile {}_3V$ is the entire space; ${}_1V \frown {}_2V$ is $O = (0, 0, 0)$, the origin. Any two-space (plane through the origin) either coincides with ${}_1V \smile {}_2V$ or intersects it in a one-space (line through the origin).

4-4 Space of n dimensions. We are accustomed to associate a point in two-dimensional space with a two-dimensional vector, and a point in three-dimensional space with a three-dimensional vector. We ordinarily think of a space of two dimensions as a geometric space, that is, as a set of points. But in a broader sense by a space of two dimensions is meant any set of objects which may be put in a one-to-one correspondence with the totality of vectors of two dimensions (pairs of scalars); and a space of three dimensions is any set of objects which may be put in a one-to-one correspondence with the totality of vectors of three dimensions (triads of scalars).

By a space of n dimensions we mean any set of objects which may be *put in a one-to-one correspondence with the vectors of* $V_n(\mathbf{F})$. We define a *point* in this n-space to be a vector $\alpha = (a_1, a_2, \ldots, a_n)$. With two such points $\alpha = (a_1, a_2, \ldots, a_n)$ and $\beta = (b_1, b_2, \ldots, b_n)$, we may associate the *direction numbers of the line segment*, $\beta - \alpha = (b_1 - a_1, b_2 - a_2, \ldots, b_n - a_n)$.

For three-space the "points" may be the points of ordinary perceptual space, but they are not necessarily so. For four-space the points may be the events of space-time as depicted in the theory of relativity. For any value of n the points may be members of a statistical set. It should be clearly realized that the mathematician makes no attempt to visualize a world of four or greater dimensions; he does not claim to see worlds which ordinary men cannot see. It is only that the mathematician finds that certain notions and concepts are discussed most readily with an algebraic setting and in terms adapted from geometry.

One should draw the distinction between "perceptual" or visualizable spaces and "conceptual" or logically conceived spaces. Ordinary two-space is a perceptual space, in that we can draw a figure to represent an entity of that mathematical world. Ordinary three-space is usually thought of as a perceptual space, although a configuration in it of moderate complexity may not be altogether perceptual. A space of n dimensions when $n > 3$, sometimes called a hyperspace, is not, as a whole, perceptual, though one may take the viewpoint that a certain section or portion of it is perceptual; this is in the same sense that we may visualize an entire plane section of a surface, although we may not simultaneously visualize the entire surface.

The ideas mentioned above are not new by any means, for Dr. C. J. Keyser in his *Mathematical Philosophy* (1922), in Lecture XVI on Hyperspaces, says: "The concept of hyperspace, though it is a modern notion, is not strictly new; it goes back three or four generations and is now, among enlightened mathematicians, as classic and orthodox as the ordinary multiplication. Though only a short while ago it was regarded by mathematicians of the conservative and reactionary type with a good deal of suspicion as being, if not crazy, at least a bit queer, over romantic, and unsound, it is now constantly employed as a great convenience by mathematicians everywhere and even by physicists (say in the kinetic theory of gases) quite without apology." Keyser, among other things, analyzes the concept of hyperspace into three related meanings, and discusses these three in some detail. However, the attitude which we have taken here that a hyperspace is a *point*-space of n dimensions, with the understanding that the points may be objects of different kinds, may be interpreted so as to include all three. This is in full accordance with Cauchy's idea as expressed in his *Mémoire sur les lieux analytiques* (1847), in which he says "We shall call a set of n variables an analytic point . . . ."

4–5 Linear dependence of vectors. The vectors $\alpha_1, \alpha_2, \ldots, \alpha_m$ in a field **F** are said to be *linearly dependent* in that field if m scalars $s_1, s_2, s_3, \ldots, s_m$ in **F**, not all zero, exist such that

$$s_1\alpha_1 + s_2\alpha_2 + s_3\alpha_3 + \cdots + s_m\alpha_m = O. \tag{6}$$

On the other hand, if no such scalars exist, the vectors are said to be *linearly independent* (in **F**). For n-dimensional vectors the equation (6) is equivalent to the n scalar equations

$$\begin{aligned} s_1a_{11} + s_2a_{21} + s_3a_{31} + \cdots + s_ma_{m1} &= 0,\\ s_1a_{12} + s_2a_{22} + s_3a_{32} + \cdots + s_ma_{m2} &= 0,\\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad &\cdot\\ s_1a_{1n} + s_2a_{2n} + s_3a_{3n} + \cdots + s_ma_{mn} &= 0, \end{aligned}$$

where $\alpha_i = (a_{i1}, a_{i2}, a_{i3}, \ldots, a_{in})$. That is, if a set of vectors are linearly dependent, their corresponding scalar coordinates are linearly dependent.

If $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_m$ are vectors of a vector space $V_n(\mathbf{F})$, we call

$$k_1\alpha_1 + k_2\alpha_2 + k_3\alpha_3 + \cdots + k_m\alpha_m \tag{7}$$

a *linear combination* of these vectors, the k's being in **F**.

THEOREM VI. *The set of all linear combinations of any set of vectors in a vector space V is a vector subspace of V.*

For from the definitions of the two basic operations (1) and (2) there follow the relations:

$$(k_1\alpha_1 + k_2\alpha_2 + \cdots + k_m\alpha_m) + (k'_1\alpha_1 + k'_2\alpha_2 + \cdots + k'_m\alpha_m) = (k_1 + k'_1)\alpha_1 + (k_2 + k'_2)\alpha_2 + \cdots + (k_m + k'_m)\alpha_m, \tag{8}$$

and

$$k'(k_1\alpha_1 + k_2\alpha_2 + \cdots + k_m\alpha_m) = (k'k_1)\alpha_1 + (k'k_2)\alpha_2 + \cdots + (k'k_m)\alpha_m. \tag{9}$$

Of fundamental importance is

THEOREM VII. *The set of nonzero vectors $\alpha_1, \alpha_2, \ldots, \alpha_m$ in a vector space V are linearly dependent if and only if some one of these vectors α_i is a linear combination of the preceding ones.*

Suppose α_i is a linear combination of the preceding vectors and we have

$$\alpha_i = s_1\alpha_1 + s_2\alpha_2 + \cdots + s_{i-1}\alpha_{i-1}.$$

This is equivalent to

$$s_1\alpha_1 + s_2\alpha_2 + \cdots s_{i-1}\alpha_{i-1} + (-1)\alpha_i = O,$$

which has at least one nonzero coefficient. Hence the vectors are dependent. On the other hand, suppose the vectors are dependent, then we have

$$(10) \qquad s_1\alpha_1 + s_2\alpha_2 + \cdots + s_m\alpha_m = O.$$

In (10) there is at least one scalar coefficient different from zero; let i be the last subscript for which $s_i \neq 0$. We may then solve (10) for α_i, getting α_i as a linear combination of the preceding vectors, except when $i = 1$. However, this exception would require that $s_1\alpha_1 = O$, with $s_1 \neq 0$, and such a condition would necessitate $\alpha_1 = O$, but the latter is contrary to the hypothesis that the given vectors are nonzero.

A consequence of Theorem VII is that a set of vectors is linearly dependent if and only if it contains a smaller subset of vectors which generate the same vector subspace. To make that clear let us consider more generally the concept of bases.

4–6 Bases. A *basis* of a vector space $V_n(\mathbf{F})$ is a set of linearly independent vectors such that every vector in $V_n(\mathbf{F})$ is a linear combination of the vectors in the basis. A linear combination like (7) is said to be *spanned* by the vectors $\alpha_1, \alpha_2, \ldots, \alpha_m$, and the vectors $\alpha_1, \alpha_2, \ldots, \alpha_m$ are said to *span* or *generate* the vector subspace formed by the linear combination (7). As some illustrations in ordinary three-space, the space spanned by a single vector $\alpha = (a_1, a_2, a_3)$ is the set of all scalar multiples $k\alpha$; geometrically this is the line through the origin and the point α. Similarly, the subspace spanned by two noncollinear vectors $\alpha = (a_1, a_2, a_3)$ and $\beta = (b_1, b_2, b_3)$ is a linear combination $\rho = r\alpha + s\beta$, which geometrically is the plane determined by the origin and the points α and β. The same vector space may be spanned by many different sets of vectors, and even by sets which contain different numbers of vectors. Thus the vectors

$$\epsilon_1 = (1, 0) \quad \text{and} \quad \epsilon_2 = (0, 1)$$

span the same space as do

$$\alpha = (5, 2), \quad \beta = (-3, 2) \quad \text{and} \quad \gamma = (7, 8).$$

Suppose α_i is a linear combination of the preceding vectors, and we have

$$\alpha_i = s_1\alpha_1 + s_2\alpha_2 + \cdots + s_{i-1}\alpha_{i-1}.$$

This is equivalent to

$$s_1\alpha_1 + s_2\alpha_2 + \cdots s_{i-1}\alpha_{i-1} + (-1)\alpha_i = O,$$

which has at least one nonzero coefficient. Hence the vectors are dependent. On the other hand, suppose the vectors are dependent; then we have

$$(10) \qquad s_1\alpha_1 + s_2\alpha_2 + \cdots + s_m\alpha_m = O.$$

In (10) there is at least one scalar coefficient different from zero; let i be the last subscript for which $s_i \neq 0$. We may then solve (10) for α_i, getting α_i as a linear combination of the preceding vectors, except when $i = 1$. However, this exception would require that $s_1\alpha_1 = O$, with $s_1 \neq 0$, and such a condition would necessitate $\alpha_1 = O$, but the latter is contrary to the hypothesis that the given vectors are nonzero.

A consequence of Theorem VII is that a set of vectors is linearly dependent if and only if it contains a smaller subset of vectors which generate the same vector subspace. To make that clear let us consider more generally the concept of bases.

4–6 Bases. A *basis* of a vector space $V_n(\mathbf{F})$ is a set of linearly independent vectors such that every vector in $V_n(\mathbf{F})$ is a linear combination of the vectors in the basis. A linear combination like (7) is said to be *spanned* by the vectors $\alpha_1, \alpha_2, \ldots, \alpha_m$, and the vectors $\alpha_1, \alpha_2, \ldots, \alpha_m$ are said to *span* or *generate* the vector subspace formed by the linear combination (7). As some illustrations in ordinary three-space, the space spanned by a single vector $\alpha = (a_1, a_2, a_3)$ is the set of all scalar multiples $k\alpha$; geometrically this is the line through the origin and the point α. Similarly, the subspace spanned by two noncollinear vectors $\alpha = (a_1, a_2, a_3)$ and $\beta = (b_1, b_2, b_3)$ is a linear combination $\rho = r\alpha + s\beta$, which geometrically is the plane determined by the origin and the points α and β. The same vector space may be spanned by many different sets of vectors, and even by sets which contain different numbers of vectors. Thus the vectors

$$\epsilon_1 = (1, 0) \qquad \text{and} \qquad \epsilon_2 = (0, 1)$$

span the same space as do

$$\alpha = (5, 2), \quad \beta = (-3, 2) \qquad \text{and} \qquad \gamma = (7, 8).$$

However, clearly ϵ_1 and ϵ_2 are linearly independent, while α, β, and γ are linearly dependent. Two sets of vectors which span the same space are said to be *linearly equivalent.*

It may be instructive to give the alternate but equivalent definition of a basis: a set of vectors $\alpha_1, \alpha_2, \ldots, \alpha_m$ which span the vector space V_n and which are linearly independent constitute a *basis* for V_n.

It should now be clear that we may delete from a set of vectors any one vector which is a linear combination of the preceding ones, and the remaining vectors will span the same subspace. To illustrate, the vectors $\alpha = (3, 2, 1, 0)$, $\beta = (-4, -3, 1, 0)$, $\gamma = (2, 1, 1, 0)$, and $\delta = (1, -1, 6, 0)$ do not span the whole of $V_4(\mathbf{R})$ because they all lie in a $V_3(\mathbf{R})$, which is a subspace of $V_4(\mathbf{R})$. The given vectors are connected (see Ex. 4, Section 3–5) by the relation $\delta = \alpha + 2\beta + 3\gamma$. So α, β, γ span the same subspace of V_4 as do $\alpha, \beta, \gamma, \delta$.

Clearly the unit vectors $\epsilon_1 = (1, 0, \ldots, 0)$, $\epsilon_2 = (0, 1, \ldots, 0)$, $\ldots, \epsilon_n = (0, 0, \ldots, 1)$ are linearly independent, and also any vector $\alpha = (a_1, a_2, \ldots, a_n)$ of $V_n(\mathbf{F})$ is expressible as a linear combination of them in the manner $\alpha = a_1\epsilon_1 + a_2\epsilon_2 + \cdots + a_n\epsilon_n$. Therefore these unit vectors form a basis for $V_n(\mathbf{F})$. The number of vectors in a basis of V_n is equal to the *dimension* n of V_n.

If ${}_1V$ and ${}_2V$ denote subspaces of V_n, then the dimension of ${}_1V \cup {}_2V$ is at least as great as the greater of the dimensions of ${}_1V$ and ${}_2V$, and at most as great as the sum of their dimensions. Let $d({}_iV)$ denote the dimension of the vector subspace ${}_iV$ of V. Then it is a fact that

$$d({}_1V) + d({}_2V) = d({}_1V \cap {}_2V) + d({}_1V \cup {}_2V).$$

Exercises

1. Determine which of the following pairs of vectors are linearly dependent; in case a pair is linearly dependent, give values of the scalar multipliers r and s which relate the vectors in the manner $r\alpha + s\beta = O$.

(*i*) $\alpha = (3, 0, 2, -1)$, $\beta = (-1, 2, 5, 3)$.
(*ii*) $\alpha = (-1, 2, 5, 3)$, $\beta = (4, -8, -20, -12)$.

2. Determine which of the following triads of vectors are linearly dependent; in case a triad is linearly dependent, give values of the scalar coefficients r and s which relate the dependent vectors in the manner $\gamma = r\alpha + s\beta$.

(*i*) $\alpha = (2, 1, -2, -1)$, $\beta = (1, -1, -1, 3)$, $\gamma = (7, 2, -7, 4)$.
(*ii*) $\alpha = (1, 2, -1, 3)$, $\beta = (0, -2, 1, -1)$, $\gamma = (2, 2, -1, 5)$.

3. Prove that any five four-dimensional vectors are linearly dependent. Obtain general results, analogous to those in the proof of Theorem X of Chapter 3.

4. Find scalars s_1, s_2, s_3, s_4 which enable us to write the vector $\alpha_5 = (2, 13, 15, 4)$ as a linear combination of $\alpha_1 = (4, 2, 8, 6)$, $\alpha_2 = (3, -1,$ [illegible]$, 2)$, $\alpha_3 = (5, -4, -6, -3)$, and $\alpha_4 = (-1, 3, 1, -3)$ in the form

$$\alpha_5 = s_1\alpha_1 + s_2\alpha_2 + s_3\alpha_3 + s_4\alpha_4.$$

4–7 The inner product of two vectors. The reader is reminded of the statement of Section 2–7 that in our consideration of distance, of angle, and of normalizing factors, we restrict ourselves to real vectors. By the *inner product* $\alpha \cdot \beta$ of the two vectors $\alpha = (a_1, a_2, \ldots, a_n)$ and $\beta = (b_1, b_2, \ldots, b_n)$, we mean the scalar quantity

$$\alpha \cdot \beta = a_1b_1 + a_2b_2 + \cdots + a_nb_n. \tag{11}$$

As consequences of this definition and properties of the underlying field, we readily establish four important characteristics of inner products as expressed in the following four theorems.

THEOREM VIII. *The inner product of two vectors is commutative: $\alpha \cdot \beta = \beta \cdot \alpha$.*

For

$$\begin{aligned} \alpha \cdot \beta &= a_1b_1 + a_2b_2 + \cdots + a_nb_n && \text{By (11)} \\ &= b_1a_1 + b_2a_2 + \cdots + b_na_n && \text{By } (SIb) \\ &= \beta \cdot \alpha. && \text{By (11)} \end{aligned}$$

Similarly, one may prove the following:

THEOREM IX. *The product of a scalar and the inner product of two vectors is associative: $k(\alpha \cdot \beta) = (k\alpha) \cdot \beta$.*

THEOREM X. *The inner product of two vectors is distributive with respect to addition: $(\alpha + \beta) \cdot \gamma = \alpha \cdot \gamma + \beta \cdot \gamma$.*

THEOREM XI. *The inner product of a nonzero (real) vector α and itself is positive: $\alpha \cdot \alpha > 0$.*

The *magnitude* $|\alpha|$ of a vector α is the expression

$$|\alpha| = \sqrt{\alpha \cdot \alpha} = \sqrt{(a_1)^2 + (a_2)^2 + \cdots + (a_n)^2}. \tag{12}$$

The *distance* d between the points $\alpha = (a_1, a_2, \ldots, a_n)$ and $\beta = (b_1, b_2, \ldots, b_n)$ is given by

$$d^2 = |\beta - \alpha|^2 = (b_1 - a_1)^2 + (b_2 - a_2)^2 + \cdots + (b_n - a_n)^2.$$

We define the cosine of the *angle* between the directions **O**α and **O**β to be given by

$$\cos\theta = \frac{\alpha \cdot \beta}{|\alpha| \cdot |\beta|}. \tag{13}$$

It can be proved that $\cos\theta$ so defined satisfies the relation $-1 \leqq \cos\theta \leqq 1$. Two vectors α and β are defined to be *orthogonal* (perpendicular) if $\alpha \cdot \beta = 0$. In a real field (to which we are limiting ourselves in the present consideration) the only vector which is orthogonal to itself is the null vector. In the complex field this is not true; to illustrate, the complex vector $(1, i)$ is orthogonal to itself. The condition $\alpha \cdot \beta = 0$ for the orthogonality of two vectors is symmetric: if α is orthogonal to β, β is orthogonal to α. This follows as an immediate consequence of the commutativity of the inner product.

A vector $\alpha = (a_1, a_2, \ldots, a_n)$ is said to be *normalized* if its magnitude, $|\alpha|$, is 1; that is, if $(a_1)^2 + (a_2)^2 + \cdots + (a_n)^2 = 1$. Every vector $\alpha \neq 0$ possesses a normalizing factor k such that $k\alpha$ is normalized, namely $k = 1/+\sqrt{(a_1)^2 + (a_2)^2 + \cdots + (a_n)^2}$. The unit vectors $\epsilon_1 = (1, 0, \ldots, 0)$, $\epsilon_2 = (0, 1, \ldots, 0)$, $\ldots$, $\epsilon_n = (0, 0, \ldots, 1)$ are normalized, mutually perpendicular, and linearly independent. They constitute a *normal orthogonal basis* for $V_n(\mathbf{R})$. In general, nonzero orthogonal vectors are linearly independent.

4–8 Abstract vector spaces. We have presented the algebra of vectors as the algebra of n-tuples $(a_1, a_2, \ldots, a_n)$ of n scalars in some field. It seems desirable that we discuss briefly some abstract aspects of vectors.

One might approach vectors from a broader axiomatic viewpoint. We could define a vector space V over a field $\mathbf{F}$ to be a set of elements, called *vectors*, satisfying the following axioms.

I. To every pair of vectors α and β in V there is associated a vector γ, called the sum of α and β, $\gamma = \alpha + \beta$, such that:

(*i*) addition of vectors is commutative, $\alpha + \beta = \beta + \alpha$;
(*ii*) addition of vectors is associative, $\alpha + (\beta + \gamma) = (\alpha + \beta) + \gamma$;
(*iii*) there exists a unique vector O such that $O + \alpha = O$ for any α in V;
(*iv*) to every vector α in V there corresponds a unique vector $-\alpha$ with the property $\alpha + (-\alpha) = O$.

To say that the vectors of V satisfy these axioms is equivalent to saying that V is a commutative group under addition.

II. Any vector α in V and any scalar s in **F** determine a vector $s\alpha$ in V, and for all α and β in V and all scalars r and s in **F** the following relations hold:

$$(i)\ (r+s)\alpha = r\alpha + s\alpha; \quad (ii)\ r(s\alpha) = (rs)\alpha;$$
$$(iii)\ r(\alpha+\beta) = (r\alpha) + (r\beta); \quad (iv)\ 0\cdot\alpha = O; \quad (v)\ 1\cdot\alpha = \alpha.$$

Note that in defining vectors to be n-tuples of scalars subject to the laws of combination specified in (1) and (2) of Section 4–1, the postulates I(*i*), (*ii*), (*iii*), (*iv*) and II(*i*), (*ii*), (*iii*), (*iv*), (*v*) were established as logical consequences of the definition we took and of the properties of the underlying field. That is, the set $V_n(\mathbf{F})$ of n-tuples satisfies the above formal axioms of abstract spaces; the point we are making now is that entities other than these n-tuples also satisfy the axioms I and II. One illustration of an abstract vector space is the set V of all polynomials, with complex coefficients, in a real variable x. As a second, consider the functions $f(x)$ whose domain is any set S whatever (as a plane region), with a field **F** as range, so that $f(x)$ assigns to each $x \subset S$ a value of $f(x) \subset \mathbf{F}$. Such a set of functions form a vector space in **F**, if the sum $h = f + g$ and the scalar k multiplied by f, $h' = kf$, are the functions defined for each $x \subset S$ by the relations

$$h(x) = f(x) + g(x) \qquad \text{and} \qquad h'(x) = kf(x).$$

In the theory of abstract vector spaces it is customary to say that a finite vector space V is of *dimension n* if it contains n linearly independent elements while every set of $n + 1$ elements is linearly dependent. In such theory it is proved that *every (general) finite vector space of dimension n can be put in a one-to-one correspondence (is isomorphic) with a vector space* V_n(**F**) *of n-tuples of numbers in* **F**.*

* See C. C. Macduffee, "Vectors and Matrices," *Carus Mathematical Monograph* **7**, The Mathematical Association of America, 1943, pp. 181–182.

CHAPTER 5

ELEMENTARY PROPERTIES OF MATRICES

5–1 The concept of a matrix. We have been studying vectors, which were defined to be one-way ordered sets of scalars which obey certain rules of operations. Our next step in the development of multiple algebra is the study of two-way ordered sets of scalars, whose elements are arranged in a square as

$$(1) \qquad A = [a]_2^2 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

or in a rectangle as

$$(2) \qquad B = [b]_3^2 = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix}.$$

Sets of mn scalars arranged in rectangular arrays with m rows and n columns and which obey certain rules of combination are called *matrices*. Let a_{ij} designate the scalar element in the ith row and the jth column, frequently called the *(i, j)th element*, of the matrix A with m rows and n columns; then A may be written

$$(3) \qquad A = [a]_n^m = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

In the (i, j)th term a_{ij}, i is called the *row index* and j is called the *column index*. A matrix with m rows and n columns is called a *matrix of order m by n*. If $m = n$ the matrix is a *square matrix of order n*. Thus the matrix of (1) above is a square matrix of order 2; that of (2) is a matrix of order 2 by 3.

We shall commonly use capital letters in italic type to represent matrices, and small italic letters to represent the scalar elements of the matrices. Capital letters in bold face type are used, as heretofore, to represent an underlying field or ring.

5–2 Addition of matrices. The now familiar operations with vectors may be used to suggest ways of defining operations with matrices. Recall from plane analytics that there are many ways of mapping or transforming a plane upon itself so that each point $\alpha =$

(x_1, x_2) of the plane is carried into a new point $\beta = (y_1, y_2)$, the coordinates of the new point β being related to those of the old point α by linear homogeneous functions as

$$(4) \quad \begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 \\ y_2 &= a_{21}x_1 + a_{22}x_2 \end{aligned} \quad \text{with } \textit{transformation matrix} \quad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

For example, a counterclockwise rotation about the origin is expressed by

$$\begin{aligned} y_1 &= x_1 \cos\theta + x_2 \sin\theta \\ y_2 &= -x_1 \sin\theta + x_2 \cos\theta \end{aligned}$$

with transformation matrix

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

Let another point $\gamma = (z_1, z_2)$ be related to α in the manner

$$(5) \quad \begin{aligned} z_1 &= b_{11}x_1 + b_{12}x_2 \\ z_2 &= b_{21}x_1 + b_{22}x_2 \end{aligned} \quad \text{with transformation matrix} \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}.$$

Denote the sum of β and γ by λ; that is, $\lambda = (w_1, w_2) = (y_1 + z_1, y_2 + z_2)$. Then we have

$$(6) \quad \begin{aligned} w_1 &= (a_{11} + b_{11})x_1 + (a_{12} + b_{12})x_2 \\ w_2 &= (a_{21} + b_{21})x_1 + (a_{22} + b_{22})x_2 \end{aligned}$$

with transformation matrix

$$C = \begin{bmatrix} a_{11} + b_{11}, & a_{12} + b_{12} \\ a_{21} + b_{21}, & a_{22} + b_{22} \end{bmatrix}.$$

We are thus led to the idea of defining the sum of two matrices A and B to be a matrix C, such that an element in C is the sum of the corresponding elements of A and B.

Another way of approaching addition of matrices is to consider the matrices A and B above as ordered sets of row vectors:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}.$$

Then

$$A + B = \begin{bmatrix} a_{11} + b_{11}, & a_{12} + b_{12} \\ a_{21} + b_{21}, & a_{22} + b_{22} \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} = C.$$

In general, the *sum* of two matrices of the same order $A = [a]_n^m$ and $B = [b]_n^m$ is the uniquely determined matrix $C = [c]_n^m$ such that

$$(7i) \quad c_{ij} = a_{ij} + b_{ij} \quad (i = 1, 2, \ldots, m)(j = 1, 2, \ldots, n)$$

or equivalently

$$(7ii)\qquad \gamma_i = \alpha_i + \beta_i \qquad (i = 1, 2, \ldots, m)$$

That is, *to add two matrices add corresponding elements or, equivalently, add corresponding vectors.*

Since any scalar element (or vector) in a matric sum is the algebraic sum of corresponding scalars (or vectors), it follows from the commutative and associative laws of addition for scalars (or vectors) that corresponding laws hold for the addition of matrices. Thus we have

THEOREM I. *The addition of matrices is commutative:* $A + B = B + A$.

THEOREM II. *The addition of matrices is associative:* $(A + B) + C = A + (B + C)$.

EXERCISES

1. If $R = \begin{bmatrix} 2 & 9 \\ 4 & 3 \end{bmatrix}$ and $S = \begin{bmatrix} 1 & 5 \\ 7 & 1 \end{bmatrix}$, verify that $R + S = \begin{bmatrix} 3 & 14 \\ 11 & 4 \end{bmatrix}$.

2. For

$$T = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} 6 & 2 & 3 \\ 4 & 4 & 6 \\ 7 & 8 & 8 \end{bmatrix},$$

find $T + U$.

3. Give a detailed proof of Theorem I for $A = [a]_2^2$ and $B = [b]_2^2$ based upon (*i*) the addition of scalars; (*ii*) the addition of vectors. An economical way to prove (*i*) is to show that the (i, j)th element in $(A + B)$ is equal to the (i, j)th element in $(B + A)$. Thus

$$\begin{aligned} \text{the } (i, j)\text{th element in } (A + B) &= a_{ij} + b_{ij} \quad \text{By } (7i) \\ &= b_{ij} + a_{ij} \quad \text{By } (SIa) \\ &= \text{the } (i, j)\text{th element in } (B + A). \end{aligned}$$

Observe that this form of proof is applicable to matrices of any order.

4. Give a proof of Theorem II for $A = [a]_2^3$, $B = [b]_2^3$, and $C = [c]_2^3$ based upon (*i*) the addition of scalars; (*ii*) the addition of vectors.

5–3 Multiplication of a matrix by a scalar.

Returning to equation (4) of the preceding section, if we multiply $\beta = (y_1, y_2)$ by the scalar k, we get the elements of $k\beta = (ky_1, ky_2)$ subjected to the transformation

$$(8)\qquad \begin{aligned} ky_1 &= ka_{11}x_1 + ka_{12}x_2 \\ ky_2 &= ka_{21}x_1 + ka_{22}x_2 \end{aligned}$$

with transformation matrix

$$B = \begin{bmatrix} ka_{11} & ka_{12} \\ ka_{21} & ka_{22} \end{bmatrix}.$$

The form of the matrix B suggests that we define the product of a matrix $A = [a]_n^m$ by a scalar k to be the matrix $B = [b]_n^m$ such that an element of B is the product of the corresponding element of A by k. That is, if $B = kA$, then

(9*i*) $$b_{ij} = ka_{ij},$$

or, equivalently, using vectors,

(9*ii*) $$\beta_i = k\alpha_i.$$

In general, *to multiply a matrix by a scalar, multiply each scalar element of the matrix by that scalar, or equivalently multiply each vector of the matrix by the scalar multiplier.*

From the definition of the multiplication of a matrix by a scalar and the basic laws for the combination of scalars, for any scalars r and s and any matrices A and B, it follows that

THEOREM III. $rA = Ar$.

THEOREM IV. $rA + sA = (r + s)A$.

THEOREM V. $(rs)A = r(sA)$.

THEOREM VI. $r(A + B) = rA + rB$.

The behavior of matrices as expressed in Theorems III, IV, V, and VI may be summarized in

THEOREM VII. *Linear combinations of matrices with scalar coefficients obey the laws of ordinary scalar algebra.*

Thus some of the laws of ordinary scalar algebra hold for matrices, but others do not, as we shall see. Significantly, those which do not hold for matrices are the commutative law and the cancellation law for the multiplication of matrices. If every law and theorem of scalar algebra would hold for matrices, there would be no reason for the existence of the theory of matrix algebra.

From Theorem IV it should be clear that we are justified in writing $4A$ for $A + A + A + A$, and $3A$ for $7A - 4A$.

We may write $A + (-1)B$ as $A - B$; we say that B is *subtracted* from A.

EXERCISES

1. Calculate $2R - S$, R and S being the numerical matrices of Ex. 1, Section 5–2.

2. Determine $U - T$, where U and T are the matrices of Ex. 2, Section 5–2.

3. Verify in detail Theorems III, IV, and V for the matrix $A = [a]_3^2$.

4. Prove Theorem VI for the matrices $A = [a]_4^2$ and $B = [b]_4^2$.

5–4 Equality of matrices. The null matrix. Two matrices are said to be *equal* if they are of the same order and have their corresponding scalar elements (or vectors) equal. So if $A = [a]_n^m$ and $B = [b]_n^m$ are equal, then

$$a_{ij} = b_{ij} \tag{10i}$$

or

$$\alpha_j = \beta_j. \tag{10ii}$$

The equality of two matrices of order m by n implies, by the above definition of equality, the satisfying of mn scalar equations. To illustrate, the single matrix equation $[a]_3^2 = [b]_3^2$ is equivalent to the six scalar equations

$$a_{11} = b_{11}, \quad a_{12} = b_{12}, \quad a_{13} = b_{13}, \quad a_{21} = b_{21}, \quad a_{22} = b_{22}, \quad a_{23} = b_{23}.$$

Each of the matric relations in Theorem I . . . VI for $A = [a]_n^m$, $B = [b]_n^m$, and $C = [c]_n^m$ is equivalent to mn scalar equations involving a_{ij}, b_{ij}, and c_{ij}, where i and j are constants for each particular scalar equation.

A matrix having every element zero is called a *null matrix*, and is written $O = [0]_n^m$. From the definition of O and the definition of equality of matrices, it follows that $A = B$ and $A - B = O$ (A, B, and O each of order m by n) mean the same thing, each element a_{ij} of A being equal to the corresponding element b_{ij} of B. Clearly $A + O = A$ and $A - O = A$.

5–5 Row matrices and column matrices. A matrix $A = [a]_n^1$ with n elements arranged in a single row is called a *row matrix*, and we write

$$[a]_n^1 = (a_1, a_2, \ldots, a_n). \tag{11}$$

A matrix $A = [a]_1^m$ with m elements arranged in a single column is called a *column matrix*, and we write

$$[a]_1^m = \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_m \end{bmatrix} = \{a_1, a_2, \ldots, a_m\}. \tag{12}$$

Analogous to thinking of a matrix $A = \|a_{ij}\|_1^m$ as an ordered set of m row vectors,

$$(13a) \qquad A = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_m \end{bmatrix} \qquad \text{where} \qquad \alpha_i = (a_{i1}, a_{i2}, \ldots, a_{in}),$$

we may think of A as an ordered set of n column vectors,

$$(13b) \qquad A = (\alpha_1', \alpha_2', \ldots, \alpha_n') \qquad \text{where} \qquad \alpha_j' = \{a_{1j}, a_{2j}, \ldots, a_{mj}\}.$$

We often refer to α_i as the ith row vector of A (j having the values $1, \ldots, n$) and α_j' as the jth column vector of A (i having the values $1, \ldots, m$).

5–6 Multiplication of matrices of the second order. Let the linear transformation

$$(14) \qquad \begin{aligned} x_1 &= a_{11}y_1 + a_{12}y_2 \\ x_2 &= a_{21}y_1 + a_{22}y_2 \end{aligned} \quad \text{with transformation matrix} \quad A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix},$$

which transforms $\beta = (y_1, y_2)$ into $\xi = (x_1, x_2)$, follow the linear transformation

$$(15) \qquad \begin{aligned} y_1 &= b_{11}z_1 + b_{12}z_2 \\ y_2 &= b_{21}z_1 + b_{22}z_2 \end{aligned} \quad \text{with transformation matrix} \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix},$$

which transforms $\gamma = (z_1, z_2)$ into $\beta = (y_1, y_2)$. The combined results of these transformations, obtained by substituting from (15) to (14), are

$$(16) \qquad \begin{aligned} x_1 &= (a_{11}b_{11} + a_{12}b_{21})z_1 + (a_{11}b_{12} + a_{12}b_{22})z_2 \\ x_2 &= (a_{21}b_{11} + a_{22}b_{21})z_1 + (a_{21}b_{12} + a_{22}b_{22})z_2 \end{aligned}$$

with matrix

$$C = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}.$$

Note that (16) is a linear transformation which maps γ directly into ξ with C as transformation matrix. The form of C suggests that we define the *product* of A and B to be $C = AB$. That is,

$$(17) \qquad AB = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix} = C.$$

Note that

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j}$$

If *row vectors* are used in writing the above transformations, we have correspondingly

(20)″ $$\alpha' = \beta' A'$$

and

(21)″ $$\beta' = \gamma' B',$$

the transpose of a column vector being a row vector. We thus see that it is natural and convenient to use *unprimed* small Greek letters as $\alpha, \beta, \gamma, \ldots$ to represent *column vectors* in connection with matrices, and relatedly to use *primed* small Greek letters as $\alpha', \beta', \gamma', \ldots$ to represent *row vectors;* subsequently we shall adhere to this practice. Integrating this convention and the notation of Section 5-5, we shall designate the *ith row of the matrix* $A = [a]_n^m$ by α_i' and the *jth column of A by* α_j.

We now utilize the convention just made as to the use of primed and unprimed vectors to give an alternate form of the typical element

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j}$$

of the matrix product $C = AB$. Let

$$\alpha' = (a_{i1}, a_{i2})$$

be the *ith row vector* in the matrix A of the product (17), and let

$$\beta = \{b_{1j}, b_{2j}\}$$

be the *jth column vector* in the matrix B of this product. We define the *inner product of the row vector* α' *and the column vector* β to be

$$\alpha'\beta = (a_{i1}, a_{i2})\{b_{1j}, b_{2j}\} = a_{i1}b_{1j} + a_{i2}b_{2j}.$$

Hence the typical element c_{ij} of the product $C = AB$ as defined by (17) may be written

$$c_{ij} = \alpha'\beta = (a_{i1}, a_{i2})\{b_{1j}, b_{2j}\}.$$

This result gives us a basis for saying that *the* (i, j) *element of the product of A and B is the inner product of the ith row vector of A and the jth column vector of B.*

Exercises

1. Verify that if $R = \begin{bmatrix} 2 & 9 \\ 4 & 3 \end{bmatrix}$ and $S = \begin{bmatrix} 1 & 5 \\ 7 & 2 \end{bmatrix}$, then

$$RS = \begin{bmatrix} (2, 9)\{1, 7\}, & (2, 9)\{5, 2\} \\ (4, 3)\{1, 7\}, & (4, 3)\{5, 2\} \end{bmatrix} = \begin{bmatrix} 65 & 28 \\ 25 & 26 \end{bmatrix}.$$

Find SR and note that $SR \neq RS$.

By $|A|$ we mean the usual *determinant* of A:

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Show that $|RS| = |SR| = |S| \cdot |R|$.

2. Test the associative law $(QR)S = Q(RS)$ for the matrix $Q = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}$ and the matrices R and S of Ex. 1 above.

3. Find BA when $B = [b]_2^2$ and $A = [a]_2^2$, and observe that in general $AB \neq BA$. Show, however, that $|AB| = |BA| = |A| \cdot |B|$.

4. If $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $O = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$, and $A = [a]_2^2$, show that

$$(i)\ IA = AI = A, \qquad (ii)\ OA = AO = O.$$

5. Using the definitions given by equations (17) and (18), and the convention at the end of Section 5-6 above for $A = [\alpha_1, \alpha_2] = [a]_2^2$ and $B = [\beta_1, \beta_2] = [b]_2^2$, show that $AB = [A\beta_1, A\beta_2]$. To prove this, note that $A\beta_1 = b_{11}\alpha_1 + b_{21}\alpha_2$ and $A\beta_2 = b_{12}\alpha_1 + b_{22}\alpha_2$. Thus the jth column in AB is $b_{1j}\alpha_1 + b_{2j}\alpha_2 = A\beta_j$; or the jth column in AB is a linear combination of the columns of A, the coefficients in this linear combination being the elements of the jth column of B.

6. In the manner of Ex. 5, for the same matrices show that $AB = \begin{bmatrix} \alpha_1' B \\ \alpha_2' B \end{bmatrix}$, and consequently that the ith row in $AB = \alpha_i' B$. Observe that this ith row in AB may be written $a_{i1}\beta_1' + a_{i2}\beta_2'$; thus the ith row in AB is a linear combination of the rows of B, the coefficients in this linear combination being the elements of the ith row of A.

7. For $A = [a]_2^2$, prove that $|kA| = k^2|A|$.

8. Using the result of Ex. 5 above and that of Ex. 7 of Section 2-9, prove that if $A = [a]_2^2$ and $B = [b]_2^2$, then $|AB| = |A| \cdot |B|$.

5-7 Multiplication of matrices in general. The multiplication of matrices differs in two important respects from scalar multiplication and from the multiplication of a matrix by a scalar: (1) matrix multiplication is in general not commutative; and (2) two matrices can be multiplied only when they satisfy a certain condition, namely when the number of columns in the first matrix A is equal to the number of rows in the second matrix B of the product AB. Matrices which satisfy the latter condition are said to be *conformable*. If A is an m by n matrix and B is an n by k matrix, the product AB is an m by k matrix.

The *product* AB of any two conformable matrices

$$A = [a]_n^m = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

and

$$B = [b]_k^n = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1k} \\ b_{21} & b_{22} & \dots & b_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ b_{n1} & b_{n2} & \dots & b_{nk} \end{bmatrix}$$

is defined to be the matrix

$$C = [c]_k^m = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1k} \\ c_{21} & c_{22} & \dots & c_{2k} \\ \cdot & \cdot & \cdot & \cdot \\ c_{m1} & c_{m2} & \dots & c_{mk} \end{bmatrix},$$

in which

(23) $$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}.$$

Here i takes the range $1, 2, \dots, m$, and j the range $1, 2, \dots, k$. It is sometimes convenient to represent this product in the form $[a]_n^m[b]_k^n = [c]_k^m$. We refer to the product AB as the *premultiplication of B by A*, and the product BA as the *postmultiplication of B by A*. Let

$$\alpha' = (a_{i1}, a_{i2}, \dots, a_{in})$$

be the *ith row vector* in the matrix A and let

$$\beta = \{b_{1j}, b_{2j}, \dots, b_{nj}\}$$

be the *jth column vector* in the matrix B. The *inner product* of the *row vector* α' and the *column vector* β is defined to be

$$\begin{aligned} \alpha'\beta &= (a_{i1}, a_{i2}, \dots, a_{in})\{b_{1j}, b_{2j}, \dots, b_{nj}\} \\ &= a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}. \end{aligned}$$

Consequently, the typical element c_{ij} of the product AB as defined by (23) can be written

$$c_{ij} = \alpha'\beta = (a_{i1}, a_{i2}, \dots, a_{in})\{b_{1j}, b_{2j}, \dots, b_{nj}\}.$$

In words, this process of *multiplication of matrices* may be defined as follows: *to obtain the* (i, j)*th element in the product* AB, *select the ith row of A and the jth column of B and take the inner product of these two vectors.*

EXERCISES

1. For

$$T = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \quad \text{and} \quad U = \begin{bmatrix} 6 & 2 & 3 \\ 4 & 4 & 6 \\ 7 & 8 & 8 \end{bmatrix},$$

find TU and UT and note that $TU \neq UT$. Show, however, that $|TU| = |UT| = |T| \cdot |U|$.

2. For

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

find AB. Is it possible to form the product BA, where A and B have these values? If so, find BA.

3. Determine the product of the matrices $(3, -2, 1)$ and $\{2, 5, 6\}$.

4. Find the product of the matrices $\{2, 5, 6\}$ and $(3, -2, 1)$.

5. For $A = [a]_3^3$, $B = [b]_1^3$, $C = [c]_3^1$:

(*i*) Determine AB, and note that the product of a square matrix and a column matrix is a column matrix;

(*ii*) Determine CA, and note that the product of a row matrix and a square matrix is a row matrix.

(*iii*) Determine CB, and note that the product of a row matrix and a column matrix is a scalar.

(*iv*) Determine BC, and note that the product of a column matrix and a row matrix is a matrix with proportional rows and proportional columns.

6. Prove that

$$(x, y, z) \begin{bmatrix} a & h & g \\ h & b & f \\ g & f & c \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = ax^2 + by^2 + cz^2 + 2fyz + 2gzx + 2hxy.$$

7. Show that $AB = [A\beta_1, A\beta_2, \ldots, A\beta_k]$, and consequently that the *j*th *column* in $AB = A\beta_j$. Note that we may write this *j*th column in AB as $b_{1j}\alpha_1 + b_{2j}\alpha_2 + \cdots + b_{nj}\alpha_n$; thus the *j*th column in AB is a linear combination of the columns of A and the coefficients of this linear combination are the elements of the *j*th column of B. (See Ex. 5, Section 5-6 for hints as how to fill in the details.)

8. Prove that the *i*th row in AB is a linear combination of the rows of B, and the coefficients in this linear combination are the elements of the *i*th row of A (see Ex. 6, Section 5-6).

9. For $A = [a]_n^n$, prove that $|kA| = k^n |A|$.

10. Using the result of Ex. 7 above and that of Ex. 10, Section 2-9, prove that if $A = [a]_3^3$ and $B = [b]_3^3$, then $|AB| = |A| \cdot |B|$. *Hint:* Write $AB = [b_{11}\alpha_1 + b_{21}\alpha_2 + b_{31}\alpha_3,\ b_{12}\alpha_1 + b_{22}\alpha_2 + b_{32}\alpha_3,\ b_{13}\alpha_1 + b_{23}\alpha_2 + b_{33}\alpha_3]$, and use $|\beta\gamma\delta| = (\beta \times \gamma) \cdot \delta$.

11. Using the result of Ex. 7 above, prove that if $A = [a]_n^n$ and $B = [b]_n^n$, then $|AB| = |A| \cdot |B|$.

12. It should be realized that when we defined the "inner product" of the "row vector" $\alpha' = (a_1, a_2)$ and the "column vector" $\beta = \{b_1, b_2\}$, we actually were giving a rule for the construction of a scalar from two particular kinds of matrices. Note carefully that $\alpha'\beta \neq \beta\alpha'$, for

$$\alpha'\beta = (a_1, a_2)\{b_1, b_2\} = (a_1, a_2)\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = a_1b_1 + a_2b_2,$$

but

$$\beta\alpha' = \{b_1, b_2\}(a_1, a_2) = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}(a_1, a_2) = \begin{bmatrix} b_1a_1, & b_1a_2 \\ b_2a_1, & b_2a_2 \end{bmatrix}.$$

Show, however, that $\alpha'\beta = \beta'\alpha$.

5–8 Commutative matrix products. The matrices $A = [a]_n^m$ and $B = [b]_k^n$ are conformable for the product $[a]_n^m[b]_k^n$; they are conformable for the product $[b]_k^n[a]_n^m$ only if $m = k$. So for $A = [a]_n^m$ and $B = [b]_m^n$ we may form both products AB and BA. In Ex. 2 of Section 5–7 we saw that $[a]_3^2[b]_2^3 = [c]_2^2$, while $[b]_2^3[a]_3^2 = [d]_3^3$. Clearly, a 2 by 2 matrix $[c]_2^2$ cannot equal a 3 by 3 matrix $[d]_3^3$. So even though the products AB and BA both exist for $A = [a]_n^m$ and $B = [b]_m^n$, a necessary but not sufficient condition for AB and BA to be equal is that $m = n$. If $AB = BA$, the product of the matrices A and B is said to be *commutative*. *In order for two matrices to be commutative they must be both conformable and square.* However, the product of square matrices of the same order is not commutative except in very special cases. In Ex. 3, Section 5–6, we saw that $AB \neq BA$ for $A = [a]_2^2$ and $B = [b]_2^2$. More generally, if

$$[c]_m^m = [a]_m^m[b]_m^m \qquad \text{and} \qquad [d]_m^m = [b]_m^m[a]_m^m,$$

we have

$$\begin{aligned} c_{ij} &= (a_{i1}, a_{i2}, \ldots, a_{im})\{b_{1j}, b_{2j}, \ldots, b_{mj}\} \\ &= a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{im}b_{mj}, \\ d_{ij} &= (b_{i1}, b_{i2}, \ldots, b_{im})\{a_{1j}, a_{2j}, \ldots, a_{mj}\} \\ &= b_{i1}a_{1j} + b_{i2}a_{2j} + \cdots + b_{im}a_{mj}. \end{aligned}$$

From these values we see in general that $c_{ij} \neq d_{ij}$, and $[c]_m^m \neq [d]_m^m$.

However, there are some special products of matrices which are commutative. Thus for

$$A = \begin{bmatrix} 4 & 2 \\ 7 & -1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -1 & -2 \\ -7 & 4 \end{bmatrix}, \quad AB = BA = -18\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Note that the latter relation may be written $AB = BA = k[1]_2^2$, where $k = |A| = |B|$, and $[1]_2^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Exercises

1. For

$$A = \begin{bmatrix} 2 & 5 & 3 \\ 3 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -3 & 1 & 7 \\ -1 & -1 & 5 \\ 5 & 1 & -13 \end{bmatrix}$$

show that $AB = BA = k[1]_3^3$. Determine the relation of the scalar k to the determinant of A.

2. For $A = [a]_2^2$, $B = [b]_2^2$, $C = [c]_2^2$, prove that $A(B + C) = AB + AC$.

5–9 Null matrices. The division law. A matrix having every one of its scalar elements zero is called a *zero matrix* or a *null matrix* and is represented by O or $[0]_n^m$. The relations

$$OA = AO = O$$

are true for a zero matrix of any order and a conformal matrix A. But if $AB = O$, it does not necessarily follow that either A or B is a null matrix. In other words, the *cancellation law of multiplication* (if $xy = 0$, then either x or y must be zero) *does not hold for matrices.* For example,

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

and

$$\begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix} \begin{bmatrix} b & b \\ -a & -a \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

are two products of matrices being the null matrix without either factor in one of the products being the null matrix.

However, the cancellation law of multiplication does hold for certain matrices; it will be proved later that if $|B| \neq 0$ and if $AB = O$, then $A = O$.

5–10 The summation convention. A most useful convention in mathematics, particularly in dealing with products of matrices, is *the summation convention: the appearance of an index twice in a single term indicates that a summation is to be made over the range of the index repeated.* Thus

$$a_ib_i \text{ for } i = 1, 2, 3, 4$$

means

$$\sum_{i=1}^{4} a_ib_i = a_1b_1 + a_2b_2 + \cdots + a_4b_4.$$

$$a_{kj}x_j \text{ for } j = 1, 2, \ldots, n$$

means

$$\sum_{j=1}^{n} a_{kj}x_j = a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kn}x_n.$$

$$a_{kj}b_{hk} \text{ for } k = 1, 2, \ldots, n$$

means

$$\sum_{k=1}^{n} a_{kj}b_{hk} = a_{1j}b_{h1} + a_{2j}b_{h2} + \cdots + a_{nj}b_{hn}.$$

In order to apply the convention, one must have given the range of values of the repeated index, or this range must be implied by the context. The repeated index is called a *dummy index;* the significance of this appellation is that the particular letter used for the repeated or dummy index is immaterial. That is, $a_{ij}x_j$ and $a_{ik}x_k$ stand for the same thing, assuming of course that the repeated index in each expression will take the same range. An index that is not dummy or repeated is called a *free index;* no summation is implied in connection with a free index. Suppose the range of all indices in $a_{ij}x_j$ is 1, 2, 3. By the summation convention this expression stands for

$$a_{i1}x_1 + a_{i2}x_2 + a_{i3}x_3,$$

in which the free index i is free to take on any one of the values 1, 2, 3. So $a_{ij}x_j$ stands for

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3, \quad a_{21}x_1 + a_{22}x_2 + a_{23}x_3, \quad a_{31}x_1 + a_{32}x_2 + a_{33}x_3.$$

Using the summation convention, the (i, j)th element in $C = AB$, or in $[c]^m_k = [a]^m_n[b]^n_k$, which is given in detailed expanded form by (23), may be written

$$c_{ij} = a_{ih}b_{hj},$$

with the indices i and j taking the ranges stipulated in Section 5–7, and h taking the range $1, 2, \ldots, n$.

EXERCISES

1. If each index in $a_{ij}x_ix_j$ has the range of values 1, 2, 3, what does this expression represent in expanded form comparable to that of the above illustration?

2. If each index has the range 1, 2, 3, 4, write out in detail what $a_{ij}x_i$ represents.

5–11 The distributive laws for multiplication of matrices. We propose to show that matrices satisfy distributive laws corresponding to those for scalars:

$$a(b + c) = ab + ac \qquad \text{and} \qquad (a + b)c = ac + bc.$$

We may prove that $A(B + C) = AB + AC$ by showing that the (i, j)th element of the matrix on the left is equal to the (i, j)th element of the matrix on the right. Let $A = [a]^m_n$, $B = [b]^n_k$, $C = [c]^n_k$.

Let x_{ij} be the (i, j)th element of $A(B + C)$ and y_{ij} the (i, j)th element of $AB + AC$. Then

$$x_{ij} = a_{ih}(b_{hj} + c_{hj}) = a_{ih}b_{hj} + a_{ih}c_{hj} = y_{ij}.$$

We have proved

THEOREM VIII(i). $A(B + C) = AB + AC$.

Similarly, we may prove

THEOREM VIII(ii). $(A + B)C = AC + BC$.

These two theorems may be combined into

THEOREM VIII. *Multiplication of matrices is distributive with respect to addition.*

5-12 The associative law for multiplication of matrices. Just as scalar multiplication satisfies the associative law $a(bc) = (ab)c$, so does matric multiplication satisfy the associative law $A(BC) = (AB)C$. Let $A = [a]_n^m$, $B = [b]_k^n$, $C = [c]_p^k$, and let x_{ij} be the (i, j)th element of $(AB)C$ and y_{ij} the (i, j)th element of $A(BC)$. Then

$$x_{ij} = (a_{ih}b_{hk})c_{kj} = a_{ih}(b_{hk}c_{kj}) = y_{ij}.$$

We have proved

THEOREM IX. *The product of three matrices A, B, C is associative.*

From Theorem IX there follows

THEOREM X. *The product of any number of matrices A_1, A_2, $\ldots$, A_p is associative; that is, in such a product the factors may be grouped in any manner provided the sequence is not changed.*

5-13 Powers of matrices. If A is a square matrix of order n, then the continued product $AAA \ldots A$ to p factors is written A^p. By the associative law of multiplication,

(24) $$A^rA^s = A^sA^r = A^{r+s}$$

and

(25) $$(A^r)^s = (A^s)^r = A^{rs}.$$

From (24) we have

THEOREM XI. *Positive integral powers of a square matrix are permutable.*

It is customary to interpret A^0 to mean I, the *unit matrix* described in the next section. From (24) and (25) we see that in the multiplication of powers of matrices, the usual index laws of scalar algebra hold for positive integral and zero indices.

EXERCISES

1. For $A = \begin{bmatrix} 3 & 4 \\ -2 & 7 \end{bmatrix}$ find A^2, and verify that $A^2A = AA^2$.

2. Why is $A^2 - B^2 \neq (A - B)(A + B)$ for matrices A and B, in general?

3. Expand $(A + B)(A + B)$. Note that the expanded form has four terms, and not three, as has the expansion of the corresponding expression in scalars.

4. Expand $(A + B)^4$.

5. If

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

show that

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

5-14 Diagonal, scalar, and unit matrices. In a square matrix of order n,

$$A = [a]_n^n = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix}.$$

The elements of the type a_{ii}, where $i = 1, 2, \ldots, n$, are said to lie in the *principal diagonal* of the matrix. A square matrix in which all the elements are zero except those along the principal diagonal is called a *diagonal matrix*. Such a matrix has the property that $a_{ij} = 0$, $i \neq j$, and is of the form

$$D = \begin{bmatrix} d_{11} & 0 & \ldots & 0 \\ 0 & d_{22} & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & d_{nn} \end{bmatrix}.$$

A diagonal matrix for which the diagonal elements are equal is called a *scalar matrix*. A scalar matrix for which the diagonal elements are all unity is called the *unit matrix*, and is represented by

$$I = [1]_n^n = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & 1 \end{bmatrix}.$$

From the definition of the multiplication of a matrix by a scalar it follows that a scalar matrix with each diagonal element equal to k may be written $K = k[1]_n^n$. To indicate that $[a]_n^n$ is a unit matrix of order n, we would write $[a]_n^n = [1]_n^n$, and similarly to indicate that $[a]_n^n$ is a scalar matrix of order n, we would write $[a]_n^n = k[1]_n^n$.

The properties of diagonal matrices, and in particular of scalar and unit matrices, are quite important in matrix algebra, and we now consider some of them.

Theorem XII. *Premultiplication DA of a square matrix A with a diagonal matrix D multiplies each row of A by the corresponding diagonal element in D.*

Let d_{ii} denote the element of D in the ith row and the ith column. Then if x_{ij} be the (i, j)th element of the product DA, we have

$$x_{ij} = (0, 0, \ldots, d_{ii}, 0, \ldots, 0)\{a_{1j}, a_{2j}, \ldots, a_{ij}, \ldots, a_{nj}\} = d_{ii}a_{ij}.$$

Hence

$$[x_{i1}, x_{i2}, \ldots, x_{ij}, \ldots, x_{in}] = d_{ii}[a_{i1}, a_{i2}, \ldots, a_{ij}, \ldots, a_{in}].$$

Thus the theorem is proved. To illustrate,

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

A similar proof gives us

Theorem XIII. *Postmultiplication AD of a square matrix A with a diagonal matrix D multiplies each column of A by the corresponding diagonal element in D.* (In Theorem XII and in related theorems when we speak of the product DA it is to be understood that D and A are conformable for multiplication.)

From Theorems XII and XIII and the definition of a scalar matrix, we have

Theorem XIV. *Premultiplication or postmultiplication of a square matrix A with a scalar matrix K multiplies each of the elements of A by the constant diagonal element in K.*

Thus

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

As a direct consequence of Theorem XIV, we have

THEOREM XV. *Any square matrix is commutative with a scalar matrix.*

And by direct multiplication or as a special case of Theorem XV, we have

THEOREM XVI. *Premultiplication or postmultiplication of a square matrix A by the unit matrix I leaves the matrix A unaltered.* $AI = IA = A$.

Also there readily follows

THEOREM XVII. $I^m = I$.

The properties of I expressed by the last two theorems justifies its being called the *unit matrix*. In the index notation which represents A by $[a_{ij}]$, writers often use $[\delta_{ij}]$ for the unit matrix; in this notation the properties of the unit matrix may be symbolized by saying that

$$\delta_{ij} \begin{cases} = 0, \text{ when } i \neq j, \\ = 1, \text{ when } i = j. \end{cases}$$

5–15 Correspondence between scalars and scalar matrices. If

$$K_1 = k_1[1]_n^n \qquad \text{and} \qquad K_2 = k_2[1]_n^n$$

are two scalar matrices of the same order n, then

$$K_1 + K_2 = (k_1 + k_2)[1]_n^n \qquad \text{and} \qquad K_1K_2 = k_1k_2[1]_n^n$$

are the scalar matrices corresponding to the scalars $k_1 + k_2$ and k_1k_2. Therefore there exists a one-to-one correspondence between scalars and scalar matrices with respect to the operations of addition and multiplication; we say that scalars and scalar matrices are "simply isomorphic" with respect to these operations. Because of this correspondence, as long as we are concerned with matrices of a *given order*, each scalar matrix may be represented by its corresponding scalar. For example, let $A = \begin{bmatrix} 1 & 3 \\ 2 & 1 \end{bmatrix}$. We may replace the equation

$$A^2 - 2A - 5[1]_2^2 = [0]_2^2 \qquad \text{by} \qquad A^2 - 2A - 5 = 0.$$

5–16 Summary of the laws of matrices. We have seen that square matrices of order n obey the following laws:

$$
\begin{aligned}
A + B &= B + A, & A + (B + C) &= (A + B) + C,\\
A(B + C) &= AB + AC, & (B + C)A &= BA + CA,\\
AI = IA &= A, & A(BC) &= (AB)C.
\end{aligned}
$$

These are the laws of scalar algebra, except for the omission of the commutative law of multiplication and the cancellation law of multiplication. That is, the main points of difference between scalars a and b and matrices A and B are:

(*i*) While for scalars $ab = ba$, this commutative law of multiplication does not hold for matrices, for $AB \neq BA$ in general.

(*ii*) While for scalars the relation $ab = 0$ implies that a or b is zero, the matrix equation $AB = O$ does not necessarily imply that either A or B is the null matrix.

Exercises

1. Verify that the A specified in Section 5–15 does satisfy the equation given at the end of that section.

2. Show that if k_1, k_2 are real or complex numbers, and A is a square matrix of order n, then

$$A^2 - (k_1 + k_2)A + k_1k_2I = (A - k_1I)(A - k_2I),$$

where I is the unit matrix of order n.

3. If $B = rA + sI$, show that $AB = BA$.

4. Given that

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

show that $A^3 = [1]_3^3$.

5. Prove that

$$\begin{bmatrix} 0 & c & -b \\ -c & 0 & a \\ b & -a & 0 \end{bmatrix}\begin{bmatrix} a^2 & ab & ac \\ ab & b^2 & bc \\ ac & bc & c^2 \end{bmatrix} = [0]_3^3.$$

The (i, j)th element of $(AB)'$

$$\begin{aligned} &= \text{the } (j, i)\text{th element of } (AB) \\ &= (j\text{th row of } A)\{i\text{th column of } B\} \\ &= (a_{j1}, a_{j2}, \ldots, a_{jn})\{b_{1i}, b_{2i}, \ldots, b_{ni}\} \\ &= (b_{1i}, b_{2i}, \ldots, b_{ni})\{a_{j1}, a_{j2}, \ldots, a_{jn}\} \\ &= (i\text{th row of } B')\{j\text{th column of } A'\}. \end{aligned}$$

It is of course understood in this and other theorems that the matrices involved are conformable for the indicated products.

THEOREM III. *The transpose of the product of any number of matrices is equal to the product of their transposes in reverse order:* $(AB \ldots G)' = G' \ldots B'A'$.

Theorem III is an obvious generalization of Theorem II.

THEOREM IV. *The transpose of the pth power of the matrix A is equal to the pth power of the transpose of A:* $(A^p)' = (A')^p$.

For $(A')^p = A'A' \ldots A'$ to p factors $= (AA \ldots A$ to p factors$)' = (A^p)'$.

6-2 **Symmetric and skew symmetric matrices.** When transposition leaves a given matrix unchanged, that matrix is said to be *symmetric*; that is, A is *symmetric* if $A = A'$. In the index notation, if $A = [a_{ij}]$ is symmetric, then $a_{ij} = a_{ji}$. The matrix $A = \begin{bmatrix} a & h & g \\ h & b & f \\ g & f & c \end{bmatrix}$ is symmetric.

If $A = A'$, then $A^p = (A')^p$. But $(A')^p = (A^p)'$ by Theorem IV. Therefore $A^p = (A^p)'$, so A^p is symmetric if A is symmetric. Clearly the unit matrix I is symmetric. Also kA is symmetric if A is symmetric. Further, if A and B are symmetric, then $(A + B)' = A' + B' = A + B$, or the sum of two symmetric matrices is symmetric. These facts establish

THEOREM V. *If A is symmetric, so is any polynomial in A with scalar coefficients.*

THEOREM VI. *The product of any matrix A and its transpose A′ is symmetric.*

For if $P = AA'$, then $P' = (AA')' = (A')'A' = AA'$. So $P = P'$.

THEOREM VII(*i*). *The sum of any matrix and its transpose A′ is symmetric.*

If $S = A + A'$, then $S' = (A + A')' = (A')' + A' = A + A'$. Hence $S = S'$.

If transposition changes the sign of all the elements of a given matrix, that matrix is said to be *skew symmetric*. If $A = [a_{ij}]$ is skew symmetric, then $a_{ij} = - a_{ji}$. The matrix

$$A = \begin{bmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{bmatrix}$$

is skew symmetric. For a matrix A to be symmetric or skew symmetric it must necessarily be square. As we have noted, if $A = [a_{ij}]$ is skew symmetric, then $a_{ij} = -a_{ji}$; in particular for $i = j$, $a_{ii} = -a_{ii}$, therefore $a_{ii} = 0$. That is, a skew symmetric matrix necessarily has zero elements in its principal diagonal.

Similar to Theorem VII(i), we have

THEOREM VII(ii). *The difference of any matrix and its transpose is skew symmetric.*

Note that the skew symmetric matrix of the third order given above has $1 + 2 = 3$ independent scalar elements, namely, those lying above the principal diagonal. By a similar reasoning we conclude that the number of independent scalar elements in the general skew symmetric matrix of order n is

$$1 + 2 + 3 + \cdots + (n - 1) = \tfrac{1}{2}n(n - 1).$$

EXERCISES

1. Form the product AA' of $A = (a_1, a_2, a_3)$ and $A' = \{a_1, a_2, a_3\}$. From the appearance of AA' and the definition of a symmetric matrix, does it follow that this product is a symmetric matrix? This product is a matrix of what order?

2. Find the product $A'A$ of the matrices given in Ex. 1. Is this product symmetric? It is of what order?

3. Form the sum of the matrix $A = [a]_2^2$ and its transpose A'. From the appearance of this sum and the definition of a symmetric matrix, does it follow that $A + A'$ is symmetric?

4. The general matrix of order n has n^2 arbitrary scalar elements. We have seen that the general skew symmetric matrix of order n has $\frac{1}{2}n(n - 1)$ independent scalar elements. Combine these results to show that the general symmetric matrix of order n has $\frac{1}{2}n(n + 1)$ independent scalar elements.

5. Using Theorems VII(i) and VII(ii) prove that every matrix A can be

expressed uniquely as the sum of a symmetric and a skew symmetric matrix, that is,

$$A = \tfrac{1}{2}(A + A') + \tfrac{1}{2}(A - A').$$

6-3 Minors and cofactors. Let us consider the matrix of order 3 by 3,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad \text{and its determinant} \quad |A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$

The determinant obtained from $|A|$ by deleting the ith row and the jth column is called the *minor* of the element a_{ij}, and we represent this minor by M_{ij}; thus

M_{11}, the minor of a_{11}, is $\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$; M_{12}, the minor of a_{12}, is $\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}$.

In some circumstances we prefer to work not with minors, but with expressions related to them, called cofactors. The *cofactor* of the element a_{ij} in $|A|$ is designated by A_{ij} and is defined by

$$A_{ij} = (-1)^{i+j} M_{ij}.$$

That is, the cofactor of an element a_{ij} is the minor of a_{ij} with a sign, $+$ or $-$, prefixed. To illustrate,

$$A_{21}, \quad \text{the cofactor of } a_{21}, \quad \text{is } (-1)^{2+1} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} = -M_{21}.$$

Note that we may expand $|A|$, the determinant of the third order, by any row, getting respectively for the expansions by the first, second, and third row,

$$|A| = a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13}, \quad |A| = a_{21}A_{21} + a_{22}A_{22} + a_{23}A_{23},$$
$$|A| = a_{31}A_{31} + a_{32}A_{32} + a_{33}A_{33}.$$

These may be combined into the single relation

(1) $$|A| = a_{h1}A_{j1} + a_{h2}A_{j2} + a_{h3}A_{j3} \quad (h = j = 1, 2, 3).$$

Similarly, expanding $|A|$ by the first, second, and third column, respectively, we have

$$|A| = a_{11}A_{11} + a_{21}A_{21} + a_{31}A_{31}, \quad |A| = a_{12}A_{12} + a_{22}A_{22} + a_{32}A_{32},$$
$$|A| = a_{13}A_{13} + a_{23}A_{23} + a_{33}A_{33}.$$

These may be combined into the single relation

(2) $$|A| = a_{1i}A_{1k} + a_{2i}A_{2k} + a_{3i}A_{3k} \quad (i = k = 1, 2, 3).$$

The relation (1) pertaining to the expansion of $|A|$ by rows and the relation (2) for the expansion of $|A|$ by columns may be expressed in the following words:

A determinant of the third order is equal to the sum of the products obtained by multiplying the elements of any row (or column) by the corresponding cofactors of these elements.

On the contrary, note that

$$\begin{vmatrix} a_{21} & a_{22} & a_{23} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = 0, \quad \text{and} \quad \begin{vmatrix} a_{31} & a_{32} & a_{33} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = 0,$$

since each is a determinant with two rows equal. From these zero determinants we have, respectively,

$$a_{21}A_{11} + a_{22}A_{12} + a_{23}A_{13} = 0, \quad \text{and} \quad a_{31}A_{11} + a_{32}A_{12} + a_{33}A_{13} = 0.$$

These two relations may be written together as

(3) $$a_{h1}A_{j1} + a_{h2}A_{j2} + a_{h3}A_{j3} = 0 \quad (h = 2, 3; j = 1).$$

In like manner we may establish

(4) $$a_{h1}A_{j1} + a_{h2}A_{j2} + a_{h3}A_{j3} = 0 \quad (h = 1, 3; j = 2),$$

(5) $$a_{h1}A_{j1} + a_{h2}A_{j2} + a_{h3}A_{j3} = 0 \quad (h = 1, 2; j = 3).$$

Still more compactly we may combine (3), (4), and (5) in the single relation

(6) $$a_{h1}A_{j1} + a_{h2}A_{j2} + a_{h3}A_{j3} = 0 \quad (h \neq j).$$

It should be clear that the free indices h and j each take the permissible range of values 1, 2, 3 except for the restriction $h \neq j$ specifically designated. If further we should use the summation convention explained in Section 5–10, we may write the relation (6) as $a_{hm}A_{jm} = 0$, $h \neq j$.

The relations (1) and (6) may be combined in the form

(7) $$a_{h1}A_{j1} + a_{h2}A_{j2} + a_{h3}A_{j3} \begin{cases} = 0 \text{ when } h \neq j, \\ = |A| \text{ when } h = j; \end{cases}$$

or

$$a_{hm}A_{jm} = |A|\,\delta_{hj}.$$

Here the indicial range is of course 1, 2, 3.

Analogous to the procedure for establishing (6), expanding appropriately chosen determinants by columns, we may establish

(8) $$a_{1i}A_{1k} + a_{2i}A_{2k} + a_{3i}A_{3k} = 0 \quad (i \neq k).$$

The relation symbolized by (6) for rows and (8) for columns may be stated in words as follows:

The algebraic sum of the products obtained by multiplying the elements of the hth row (or column) of a third order determinant by the cofactors of the corresponding elements of the jth row (or column), where $h \neq j$, is zero.

Analogous to the manner of combining (1) and (6) into (7), the relations (2) and (8) may be written compositely as

$$(9) \qquad a_{1i}A_{1k} + a_{2i}A_{2k} + a_{3i}A_{3k}\begin{cases} = 0 \text{ when } i \neq k, \\ = |A| \text{ when } i = k, \end{cases}$$

or

$$a_{mi}A_{mk} = |A|\,\delta_{ik}.$$

Next, consider a square matrix A of order n, and its determinant $|A|$:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}, \quad \text{and} \quad |A| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}.$$

Proceeding as above for the determinant of the third order, we may establish the following results:

A determinant of the nth order is equal to the sum of the products obtained by multiplying the elements of any row (or column) by the corresponding cofactors of these elements; that is,

$$(10) \quad |A| = a_{h1}A_{j1} + a_{h2}A_{j2} + \cdots + a_{hn}A_{jn} \quad (h = j = 1, 2, \ldots, n),$$

$$(11) \quad |A| = a_{1i}A_{1k} + a_{2i}A_{2k} + \cdots + a_{ni}A_{nk} \quad (i = k = 1, 2, \ldots, n).$$

The algebraic sum of the products obtained by multiplying the elements of the hth row (or column) of a determinant of the nth order by the cofactors of the corresponding elements of the jth row (or column), where $h \neq j$, is zero; in symbols

$$(12) \qquad a_{h1}A_{j1} + a_{h2}A_{j2} + \cdots + a_{hn}A_{jn} = 0 \quad (h \neq j),$$

$$(13) \qquad a_{1i}A_{1k} + a_{2i}A_{2k} + \cdots + a_{ni}A_{nk} = 0 \quad (i \neq k).$$

The relations (10) and (12) may be combined into

$$(14) \qquad a_{h1}A_{j1} + a_{h2}A_{j2} + \cdots + a_{hn}A_{jn}\begin{cases} = 0 \text{ when } h \neq j, \\ = |A| \text{ when } h = j; \end{cases}$$

or

$$a_{hm}A_{jm} = |A|\,\delta_{hj}.$$

Similarly, the relations (11) and (13) may be written compositely as

$$(15) \qquad a_{1i}A_{1k} + a_{2i}A_{2k} + \cdots + a_{ni}A_{nk}\begin{cases} = 0 \text{ when } i \neq k, \\ = |A| \text{ when } i = k; \end{cases}$$

or

$$a_{mi}A_{mk} = |A|\,\delta_{ik}.$$

It is of course understood that any index appearing in connection with a determinant of the nth order has the range $1, 2, \ldots, n$.

6–4 The adjoint and inverse of a square matrix of order three. We know that the inverse, or reciprocal, of a scalar a is defined if $a \neq 0$, and satisfies the relation

$$aa^{-1} = a^{-1}a = 1.$$

In a similar manner there is associated with a square matrix A, if $|A| \neq 0$, a unique matrix, written A^{-1} by analogy to the inverse of a scalar, and satisfying the relation

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

I being the unit matrix of the same order as A.

For a square matrix A of the third order the nine scalar equations represented by (7) are together equivalent to the single matrix equation

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{bmatrix} = |A| \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{16}$$

We introduce the concept of the *adjoint of a matrix* A, represented by $\tilde{A}$, and defined by

$$\tilde{A} = [\tilde{A}_{ij}] = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} & \tilde{A}_{13} \\ \tilde{A}_{21} & \tilde{A}_{22} & \tilde{A}_{23} \\ \tilde{A}_{31} & \tilde{A}_{32} & \tilde{A}_{33} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{bmatrix} \tag{17}$$

or

$$\tilde{A}_{ij} = A_{ji}.$$

Note that the indices i and j are interchanged in $\tilde{A}_{ij}$ as compared with A_{ji}; the adjoint of A is the transpose of the matrix of the cofactors of A. Then we have from (16) and (17)

$$A\tilde{A} = |A|I. \tag{18}$$

Dividing equation (18) by $|A|$, we obtain $A\tilde{A}/|A| = I$, when $|A| \neq 0$. The matrix $\dfrac{\tilde{A}}{|A|}$ is called the *inverse* of A, and we denote it by A^{-1}, that is,

$$A^{-1} = \frac{\tilde{A}}{|A|}. \tag{19}$$

Then we have

$$AA^{-1} = I. \tag{20}$$

Similarly, for the matrix A of the third order the nine scalar equations represented by (9) are together equivalent to the single matrix equation

$$(21) \qquad \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} = |A| \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

or

$$A'(\tilde{A})' = |A|I.$$

Taking the transpose of both sides of the last equation and using Theorem II, we have

$$(22) \qquad \tilde{A}A = |A|I.$$

Dividing both sides of this equation by $|A|(|A| \neq 0)$ and using (19), we get

$$(23) \qquad A^{-1}A = I.$$

If the determinant of a square matrix A, $|A|$, does not vanish, the matrix A is said to be *nonsingular;* if $|A|$ vanishes, A is said to be *singular*. We have found that the inverse A^{-1} of a nonsingular square matrix A of order 3 exists and is commutative with A, the product being the unit matrix I; thus

$$(24) \qquad AA^{-1} = A^{-1}A = I.$$

If we use the index notation in which $A = [a_{ij}]$ and let $A^{-1} = [\mathrm{a}_{ij}]$, then the relation comparable to (24) is

$$(25i) \qquad [a_{ik}][\mathrm{a}_{kj}] = [\mathrm{a}_{ik}][a_{kj}] = [\delta_{ij}]$$

or

$$(25ii) \qquad a_{ik}\mathrm{a}_{kj} = \mathrm{a}_{ik}a_{kj} = \delta_{ij};$$

the first of the relations (25) is a matrix equation, the second is, for a designated pair of values of i and j, one of the nine scalar equations involved. In (25) the range of each index is 1, 2, 3, since here $n = 3$.

Note that the adjoint of A exists whether A is singular or nonsingular. *If A is singular so that $|A| = 0$, then*

$$(26) \qquad A\tilde{A} = \tilde{A}A = 0.$$

To calculate the inverse of a square matrix of order 3 it may be convenient for the novice to proceed as follows: first, construct the matrix whose elements are the cofactors of the elements of A; this we term the *cofactor matrix* of A and denote it by Co A. Secondly, construct $\tilde{A}$, the adjoint of A, by taking the transpose of Co A; that is, $\tilde{A} = (\text{Co } A)'$. Thirdly, calculate $|A|$, and multiply $\tilde{A}$ by $1/|A|$, and thereby construct A^{-1}, the inverse of A. To summarize,

$$(27) \qquad A^{-1} = \frac{\tilde{A}}{|A|} = \frac{(\text{Co } A)'}{|A|}.$$

To illustrate, if

$$A = \begin{bmatrix} 2 & 5 & 3 \\ 3 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix}, \qquad \text{then} \qquad \text{Co } A = \begin{bmatrix} -3 & -1 & 5 \\ 1 & -1 & 1 \\ 7 & 5 & -13 \end{bmatrix},$$

$$\tilde{A} = \begin{bmatrix} -3 & 1 & 7 \\ -1 & -1 & 5 \\ 5 & 1 & -13 \end{bmatrix}, \quad |A| = 4.$$

Therefore

$$A^{-1} = \frac{1}{4}\begin{bmatrix} -3 & 1 & 7 \\ -1 & -1 & 5 \\ 5 & 1 & -13 \end{bmatrix}.$$

Recall that $|A|$ may be obtained by multiplying any row (or column) in A by the corresponding row (or column) in the matrix of the cofactors of the elements of A. As a check on the value of A just obtained, we may verify by direct multiplication that $AA^{-1} = A^{-1}A = I$.

Exercises

1. If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then Co $A = \begin{bmatrix} d & -c \\ -b & a \end{bmatrix}$, and $\tilde{A} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. In this manner, for $A = \begin{bmatrix} 3 & 4 \\ 1 & -2 \end{bmatrix}$, find $\tilde{A}$ and A^{-1}. Check your result by showing that the relation (24) is satisfied.

2. Combine the scalar equations $3x + 4y = 7$, $x - 2y = 9$ into the single matrix equation $\begin{bmatrix} 3 & 4 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 7 \\ 9 \end{bmatrix}$. Solve the latter equation for the column vector $\{x, y\}$ by premultiplying both sides of this equation by the matrix A^{-1} found in Ex. 1 above.

3. Combine the scalar equations $5x - 2y = 10$, $3x + y = 17$ into a single matrix equation as in Ex. 2, and solve the latter for the column vector $\{x, y\}$ by premultiplying both sides of this equation by the inverse of the *coefficient matrix* $\begin{bmatrix} 5 & -2 \\ 3 & 1 \end{bmatrix}$.

4. For the matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, show directly that: (*i*) $|\tilde{A}| = |A|$; (*ii*) adjoint of the adjoint of $A = A$; (*iii*) the determinant of the inverse of A

is equal to the reciprocal of the determinant of A, that is, $|A^{-1}| = 1/|A|$. Do this two ways, first by taking the determinant of

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

and using the fact that $|kA| = k^2|A|$ (see Ex. 7, Section 5–6); and secondly by taking the determinant of both sides of $AA^{-1} = I$ and using the fact that $|AB| = |A| \cdot |B|$.

5. If

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{bmatrix},$$

what is A^{-1}?

6. For $A = [a]_2^2$ and $B = [b]_2^2$, show that $(AB)^{-1} = B^{-1}A^{-1}$. Do this two ways, first by using the expression for the inverse of each matrix as that given in Ex. 4(*iii*) and thus forming the products and inverses indicated in a direct manner; secondly, form the product $(B^{-1}A^{-1})(AB)$, and recall that multiplication of matrices is associative.

7. Combine the scalar equations $x_1 + x_2 + x_3 = 6$, $x_1 + 2x_2 + 3x_3 = 14$, $x_1 + 4x_2 + 9x_3 = 36$ into a single matrix equation as in Ex. 2, and solve the latter for the column vector $\{x_1, x_2, x_3\}$ by premultiplying this equation by the inverse of the coefficient matrix (which you have found in Ex. 5).

8. In the manner of Ex. 5, solve $2x_1 + 3x_2 + x_3 = 2$, $x_1 + x_2 + x_3 = 4$, $3x_1 - 2x_2 + 2x_3 = 6$ for the unknown vector.

9. For $A = [a]_3^3$, take the determinant of both sides of $A\tilde{A} = |A|I$, and using $|AB| = |A| \cdot |B|$, show that $|\tilde{A}| = |A|^2$.

10. For $A = [a]_3^3$, take the determinant of both sides of $AA^{-1} = I$, and show that $|A^{-1}| = 1/|A|$.

11. For $A = [a]_3^3$ and $B = [b]_3^3$ show that $(AB)^{-1} = B^{-1}A^{-1}$. *Hint:* Proceed as in the second part of Ex. 6.

6–5 The adjoint and inverse of a square matrix of order n. For the matrix $A = [a]_n^n$ the n^2 scalar equations (14) are together equivalent to the single matrix equation

$$(28)\quad \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}\begin{bmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \cdot & \cdot & \cdot & \cdot \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{bmatrix} = |A|\begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

We define the *adjoint* of the matrix A, designated by $\tilde{A}$ (occasionally by adj A), by the equation

(29) $$\tilde{A} = [\tilde{A}_{ij}] = \begin{bmatrix} \tilde{A}_{11} & \tilde{A}_{12} & \dots & \tilde{A}_{1n} \\ \tilde{A}_{21} & \tilde{A}_{22} & \dots & \tilde{A}_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \tilde{A}_{n1} & \tilde{A}_{n2} & \dots & \tilde{A}_{nn} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \cdot & \cdot & \cdot & \cdot \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{bmatrix},$$

or

$$\tilde{A}_{ij} = A_{ji}.$$

Note that the indices i and j are interchanged in $\tilde{A}_{ij}$ as compared with A_{ji}; the adjoint of A is the transpose of the cofactor matrix of A. Accordingly, we have $A\tilde{A} = |A|I$. Dividing this equation by $|A|$ we get

(30) $$A\frac{\tilde{A}}{|A|} = I \qquad (|A| \neq 0).$$

The matrix

(31) $$A^{-1} = \frac{\tilde{A}}{|A|}$$

is called the *inverse* of A. *The inverse of a matrix A is equal to the adjoint of A (transpose of the cofactor matrix of A) divided by the determinant of A.* From (30) and (31), we have

(32) $$AA^{-1} = I.$$

In like manner, for $A = [a]_n^n$ the n^2 scalar equations (15) are together equivalent to the single matrix equation

(33) $$\begin{bmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \cdot & \cdot & \cdot & \cdot \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{bmatrix} = |A| \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & 1 \end{bmatrix} \quad \text{or} \quad A'(\tilde{A})' = |A|\,I.$$

Taking the transpose of both sides of the last equation and using Theorem II, we have

$$\tilde{A}A = |A|I \qquad \text{and hence} \qquad \frac{\tilde{A}}{|A|}A = I \quad (|A| \neq 0),$$

or

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

Therefore the inverse A^{-1} of a nonsingular square matrix A of order n

exists and is commutative with A, this product being the unit matrix of order n; that is,

$$AA^{-1} = A^{-1}A = I. \tag{35}$$

Equation (25) has the same form for a square matrix of order n as for a square matrix of order 3, the range of each index now being 1, 2, . . . , n.

THEOREM VIII. *If A is a nonsingular square matrix, there is only one matrix which when multiplied by A gives the unit matrix, and that is A^{-1}.*

To prove Theorem VIII, let B be any matrix such that $AB = I$. Since $|A| \neq 0$, A^{-1} exists, and

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

and similarly if $BA = I$, then $A^{-1} = B$. These results may be alternately stated in

THEOREM IX. *The necessary and sufficient condition for two nonsingular square matrices A and B to be inverses of each other is that AB or BA be the identity matrix.*

Taking the determinant of both sides of (34) and using the fact that the determinant of the product of two matrices is equal to the product of the determinants of the two matrices (Appendix I), we get $|A^{-1}| \cdot |A| = |I|$. Since $|I| = 1$, we have proved

THEOREM X. *The determinant of the inverse of a matrix A is the reciprocal of the determinant of A.*

From Theorem IX and the fact that a matrix and its inverse are commutative, there follows

THEOREM XI. *If the product of two square matrices of the same order is the identity matrix, that product is commutative.*

More generally, we have

THEOREM XII. *If the product of two square matrices of the same order is the non-null scalar matrix $K = k[1]_n^n$, then the product is commutative.*

To prove the latter theorem, suppose we have

$$AB = kI, \quad k \neq 0. \tag{36}$$

Then $(k^{-1}A)B = I$. Consequently, by Theorem XI, $B(k^{-1}A) = I$. From these relations it follows that $BA = kI = AB$. To illustrate, if

$$\begin{bmatrix} 3 & 4 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} 4 & 8 \\ 2 & -6 \end{bmatrix} = 20\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

we know that

$$\begin{bmatrix} 4 & 8 \\ 2 & -6 \end{bmatrix}\begin{bmatrix} 3 & 4 \\ 1 & -2 \end{bmatrix} = 20\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

However, if the product of two square matrices of the same order is the null matrix, the product is not necessarily commutative. To illustrate

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

but

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

Yet if one of the matrices is the adjoint of the other, then the product is commutative. Thus

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

and

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

From this we see that the matrix B for which $AB = O$ (A singular) is not unique.

THEOREM XIII. *The inverse of the product of two nonsingular square matrices is the product of their inverses in reverse order:* $(AB)^{-1} = B^{-1}A^{-1}$.

We want to prove that $(AB)(B^{-1}A^{-1}) = I$. Now $(AB)B^{-1}A^{-1} = [(AB)B^{-1}]A^{-1} = A(BB^{-1})A^{-1} = (AI)A^{-1} = AA^{-1} = I$, and the theorem is established.

THEOREM XIV. *The inverse of the transpose of a nonsingular square matrix A is equal to the transpose of the inverse of A:* $(A')^{-1} = (A^{-1})'$.

This theorem follows from Theorem VIII and the fact that

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

THEOREM XV. *If the matrix A is nonsingular, the equation $AB = O$ necessitates that $B = O$.*

Since A is nonsingular, it has an inverse A^{-1}, and $A^{-1}A = I$. This relation combined with $AB = O$ gives $A^{-1}AB = O$; but $A^{-1}AB = IB = B$; so $B = O$. Similarly, we may prove

THEOREM XVI. *If the matrix B is nonsingular, the equation $AB = O$ implies $A = O$.*

The last two theorems may be combined into the "cancellation law of multiplication for matrices":

THEOREM XVII. *If $AB = O$, then either $A = O$, or $B = O$, or both A and B are singular matrices.*

6–6 Solution of n linear nonhomogeneous equations in n unknowns. Cramer's rule. The system of n nonhomogeneous linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdot \qquad \cdot \qquad \cdot \qquad & \cdot \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned} \tag{37}$$

or

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ b_n \end{bmatrix}$$

may be written

$$A\delta = \beta \qquad \text{or} \qquad a_{ij}x_j = b_i, \tag{38}$$

where A is the square *coefficient matrix* of order n, and δ and β are the column vectors $\delta = \{x_1, x_2, \ldots, x_n\}$ and $\beta = \{b_1, b_2, \ldots, b_n\}$. As previously indicated in Ex. 2, 3, 7, and 8 of Section 6–4, if A is nonsingular ($|A| \neq 0$), we may premultiply both sides of (38) by A^{-1} and thereby solve for the unknown vector δ; for $A^{-1}A\delta = A^{-1}\beta$; $I\delta = A^{-1}\beta$, and so we get

$$\delta = A^{-1}\beta. \tag{39}$$

Let H_i be the matrix obtained from A by replacing the ith column by β. *The solution* (39) *may be written in the form*

$$\delta = \frac{1}{|A|}(|H_1|, |H_2|, \ldots, |H_n|). \tag{40}$$

Let $\{x_1, x_2, \ldots, x_n\}$ be a solution of (38). Then we have

$$a_{ij}x_j = b_i. \tag{41}$$

Multiplying equation (41) by the cofactors $A_{11}, A_{21}, \ldots, A_{n1}$ of the elements $a_{11}, a_{21}, \ldots, a_{n1}$ in the first column of the coefficient matrix A and adding the results, we get

$$|A|\, x_1 = b_1A_{11} + b_2A_{21} + \cdots + b_nA_{n1}, \qquad \text{or} \qquad x_1 = \frac{|H_1|}{|A|}.$$

If we use as multipliers the cofactors $A_{12}, A_{22}, \ldots, A_{n2}$ of the elements in the second column of the coefficient matrix, we get

$$|A|\, x_2 = b_1A_{12} + b_2A_{22} + \cdots + b_nA_{n2}, \qquad \text{or} \qquad x_2 = \frac{|H_2|}{|A|}.$$

If we use as multipliers the cofactors $A_{1i}, A_{2i}, \ldots, A_{ni}$ of the elements in the ith column of the coefficient matrix, we get

$$|A|\, x_i = b_1A_{1i} + b_2A_{2i} + \cdots + b_nA_{ni}, \qquad \text{or} \qquad x_i = \frac{|H_i|}{|A|}.$$

We thus see that if $|A| \neq 0$, our assumed solution can be nothing other than that given by (40). That this is a solution can be verified by direct substitution.

The rule given by (40) for writing down the solution of n linear nonhomogeneous equations in n unknowns, whose coefficient matrix is nonsingular, is known as *Cramer's rule*.

6–7 Negative powers of matrices. General index laws for matrices. We have already used A^p to stand for $AAA \ldots A$ to p factors, and have stated in Section 5–13 the index laws for matrices when the indices are positive integers or zero.

We have seen that the inverse A^{-1} of the matrix A exists provided A is nonsingular. If A is nonsingular ($|A| \neq 0$) we may define higher negative integral powers of A as powers of the inverse of A:

$$A^{-m} = (A^{-1})^m,$$

where m is a positive integer. To illustrate, if $A = \begin{bmatrix} 1 & -2 \\ -2 & -2 \end{bmatrix}$, then

$$A^{-1} = -\frac{1}{6}\begin{bmatrix} -2 & 2 \\ 2 & 1 \end{bmatrix}, \quad A^{-2} = \frac{1}{36}\begin{bmatrix} 8 & -2 \\ -2 & 5 \end{bmatrix},$$
$$A^{-3} = -\frac{1}{216}\begin{bmatrix} -20 & 14 \\ 14 & 1 \end{bmatrix}.$$

It follows that the laws of exponents for matrices

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

hold for negative integral indices, as well as for positive integral and zero indices, provided A is nonsingular.

6–8 Orthogonal matrices. Clearly, the familiar scalar relations for representing the rotation of cartesian coordinate axes OX_1 and OX_2 through an angle θ (see Ex. 1, Section 2–8 for derivation)

$$y_1 = x_1 \cos\theta + x_2 \sin\theta$$
$$y_2 = -x_1 \sin\theta + x_2 \cos\theta,$$

may be put in the matrix form

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \text{or} \qquad [y]_1^2 = [m]_2^2[x]_1^2,$$

where

$$M = [m]_2^2 = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

Now $|M| = 1$ and $\tilde{M} = M'$; consequently $M^{-1} = \tilde{M} = M'$. Similarly, the formulas for the rotation in three-dimensional geometry of the cartesian coordinate axes OX_1, OX_2, OX_3 to new positions OY_1, OY_2, OY_3 in matrix form (see Section 2–8) are

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \qquad \text{or} \qquad [y]_1^3 = [m]_3^3[x]_1^3.$$

For the matrix $[m]_3^3$ the elements of each row are the direction cosines of one of the new coordinate axes relative to the old axes, so that the sum of the squares of the elements of each row is 1. Since the coordinate axes are mutually perpendicular, the inner product of any row vector of this matrix by any other row vector of it is zero. Moreover, the elements of each column are the direction cosines of one of the old coordinate axes relative to the new system, so that the properties just stated for rows hold also for columns. From these facts and the properties of direction cosines of line vectors, the reader can verify directly that $|M| = 1$ and $M^{-1} = \tilde{M} = M'$, where $M = [m]_3^3$. Alternately, one may note that $MM' = I$, from which we get $M^{-1} = M'$. Thus associated with a rotation of coordinate axes in two or three dimensions there is a matrix M with the property that $M^{-1} = M'$. A square matrix A with the property that its inverse is equal to its transpose is called an *orthogonal matrix*. An orthogonal matrix is necessarily nonsingular.

Since $|A'| = |A|$ and $|A^{-1}| = 1/|A|$ for any nonsingular square matrix A, it follows that if A is orthogonal, then $|A|^2 = 1$. We have proved

THEOREM XVIII. *The determinant of any orthogonal matrix is* $+1$ *or* -1.

The two types of orthogonal matrices are distinguished from each other by the values of their determinants. Those with determinants $+1$ are called *proper;* those with determinants -1 are called *improper.*

THEOREM XIX. *The product of two orthogonal matrices is orthogonal.*

For if $AA' = I$ and $BB' = I$, then $AB(AB)' = ABB'A' = AA' = I$.

THEOREM XX. *The inverse of an orthogonal matrix is orthogonal.*

We want to show that if $A'A = I$, then $(A^{-1})'A^{-1} = I$. Now $(A^{-1})'A^{-1} = (A')'(A^{-1}) = AA^{-1} = I$, and the theorem is proved.

A necessary and sufficient condition for a square matrix A to be orthogonal is for the product of A and its transpose A' to be the unit matrix:

(42) $$AA' = I.$$

For $n = 2$ this matrix condition on A may be written

(43) $$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

which is equivalent to the scalar conditions

(44) $$(a_{11})^2 + (a_{12})^2 = 1, \quad (a_{21})^2 + (a_{22})^2 = 1, \quad a_{11}a_{21} + a_{12}a_{22} = 0.$$

Note that the scalar product of the first row in A and the first column in A' yields the first of these conditions; the scalar product of the second row in A and the second column in A' yields the second condition; however, either the scalar product of the first row in A and the second column in A' or that of the second row in A and the first column in A' yields the third scalar condition.

The matrix condition $A'A = I$ is equivalent to the matrix condition $AA' = I$ just considered, and yields equivalent scalar conditions.

Since the general square matrix of order 2 has $2^2 = 4$ independent scalar elements, when A is orthogonal this number 4 is reduced by the number of the independent scalar conditions on these elements as a consequence of the matrix being orthogonal, namely by 3, since in (44) we have 3 independent scalar conditions. Hence there is only one arbitrary scalar element in an orthogonal square matrix of order 2.

By a similar argument we can show that the matrix condition $AA' = I$ for the square matrix of order 3 to be orthogonal yields 6 independent scalar conditions, 3 of these being conditions for the rows in A to be normalized vectors, and 3 conditions for any two of the three rows in A to be orthogonal. The general square matrix of order 3 has $3^2 = 9$ independent scalar elements; therefore the general orthogonal square matrix of order 3 has 3 independent scalar elements.

For the square matrix $A = [a]_n^n$ to be orthogonal, the matrix condition (42) is equivalent to the scalar conditions

$$a_{im}a_{jm} = \delta_{ij}. \tag{45}$$

The scalar conditions (45) mean that the sum of the squares of the elements of each row vector of A is 1, which gives n scalar conditions, and that the inner product of any two different row vectors of A is zero, which gives ${}_nC_2 = \frac{1}{2}n(n-1)$ scalar conditions. Altogether there are then $\frac{1}{2}n(n+1)$ scalar conditions on $A = [a]_n^n$ for it to be orthogonal. Since $A = [a]_n^n$ has n^2 independent scalar elements, it follows that there are $\frac{1}{2}n(n-1)$ independent scalar elements in the general orthogonal matrix of order n.

Theorem XXI. *The product of two proper or of two improper orthogonal matrices is a proper orthogonal matrix.*

In Theorem XIX we showed that if A and B are orthogonal matrices of the same order, then their product AB is likewise orthogonal. Also $|AB| = |A| \cdot |B|$; if A and B are proper orthogonal matrices, each has determinant equal to $+1$, and consequently $|AB|$ is equal to $+1$. A similar statement holds if A and B are both improper, each having determinant equal to -1.

Theorem XXII. *The inverse of a proper (improper) orthogonal matrix is a proper (improper) orthogonal matrix.*

Since $|A^{-1}| = 1/|A|$, $|A^{-1}|$ and $|A|$ must be of the same sign; from this fact and Theorem XX there follows Theorem XXII.

Theorem XXIII. *If the rows or columns of an orthogonal matrix are permuted, the resulting matrix is still orthogonal.*

Let $\alpha_1', \alpha_2', \ldots, \alpha_n'$ be the rows in the orthogonal matrix A; then $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the columns in $A' = A^{-1}$. The condition for A to be orthogonal, $AA' = I$, is equivalent to requiring $\alpha_i' \cdot \alpha_i = 1$, and $\alpha_i' \cdot \alpha_j = 0$ for $i \neq j$. If the rows of A are subjected to a given permutation, the columns of A' are subjected to the same permutation. Suppose the permutation takes $1, 2, \ldots, n$ into $\rho_1, \rho_2, \ldots, \rho_n$. Consider the product

$$\begin{bmatrix} \alpha'_{\rho_1} \\ \alpha'_{\rho_2} \\ \cdot \\ \cdot \\ \cdot \\ \alpha'_{\rho_n} \end{bmatrix} [\alpha_{\rho_1}, \alpha_{\rho_2}, \ldots, \alpha_{\rho_n}].$$

Since $\alpha'_{\rho_1} \cdot \alpha_{\rho_1} = 1$ and $\alpha'_{\rho_1} \cdot \alpha_{\rho_2} = 0$, etc., the matrix must be orthogonal, and the theorem is proved.

Exercises

1. Solve each of the following systems of equations by Cramer's rule:

(i)
$$\begin{aligned} x_1 + x_2 - 2x_3 + 4x_4 &= 1, \\ -x_1 + x_2 + 3x_3 + 2x_4 &= 2, \\ 2x_1 - x_2 + 2x_3 - 3x_4 &= -3, \\ 2x_1 \qquad - x_3 - 2x_4 &= -2. \end{aligned}$$

(ii)
$$\begin{aligned} x_1 - 3x_2 + 4x_3 - 2x_4 &= 6, \\ x_1 \qquad - 2x_3 + 3x_4 &= 9, \\ x_2 - 3x_3 + 2x_4 &= 1, \\ x_1 - 2x_2 + x_3 + x_4 &= 10. \end{aligned}$$

2. Calculate the inverse of the following matrices A (if it exists) and check your results by showing that $AA^{-1} = I$.

$$(i) \begin{bmatrix} 1 & 1 & 1 & 0 \\ -2 & 1 & -1 & -2 \\ -4 & -2 & -3 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}. \qquad (ii) \begin{bmatrix} 2 & 0 & -1 & 1 \\ -1 & 1 & 1 & 0 \\ 1 & 2 & 1 & 1 \\ 6 & 1 & -2 & 3 \end{bmatrix}.$$

3. For

$$A = \begin{bmatrix} k & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & n \end{bmatrix}$$

show that

$$A^2 = \begin{bmatrix} k^2 & 0 & 0 \\ 0 & m^2 & 0 \\ 0 & 0 & n^2 \end{bmatrix} \quad A^{-1} = \begin{bmatrix} 1/k & 0 & 0 \\ 0 & 1/m & 0 \\ 0 & 0 & 1/n \end{bmatrix}.$$

4. If the general orthogonal matrix of order n is to be expressed in terms of some other type of matrix, it must be a matrix that has $\frac{1}{2}n(n-1)$ independent scalar elements. Such is the skew symmetric matrix of order n, as noted in Section 6–2. To construct the general orthogonal matrix of order 2, consider the general skew symmetric matrix of order 2,

$$S = \begin{bmatrix} 0 & a \\ -a & 0 \end{bmatrix}.$$

Construct the matrix

$$M = (I - S)(I + S)^{-1} = \frac{1}{1 + a^2}\begin{bmatrix} 1 - a^2, & -2a \\ 2a, & -a^2 + 1 \end{bmatrix}$$

and show that it is orthogonal.

Note that the determinant of the orthogonal matrix M just given is $+1$; so it is not the general orthogonal matrix of order 2. But by modifying it with a numerical factor we can construct from it the general orthogonal matrix of order 2. In general, *the orthogonal matrix M of order n with determinant ± 1 according as p is even or odd is given by*

$$M = J(I - S)(I + S)^{-1},$$

where J is the diagonal matrix with p negative 1's and $n - p$ positive 1's for diagonal elements, I being the unit matrix of order n, and S the general skew symmetric matrix of order n.

5. Using the general skew symmetric matrix of order 3, as given near the end of Section 6–2, show as in Ex. 4 that

$$M = (I - S)(I + S)^{-1}$$
$$= k\begin{bmatrix} 1 + c^2 - a^2 - b^2, & -2(a + bc), & 2(ac - b) \\ 2(a - bc), & 1 - a^2 + b^2 - c^2, & -2(c + ab) \\ 2(ac + b), & 2(c - ab), & 1 + a^2 - b^2 - c^2 \end{bmatrix},$$

where $k = 1/(1 + a^2 + b^2 + c^2)$. The result in this exercise may be thought of as finding a set of rational values of direction cosines of three mutually perpendicular lines in space. It is difficult to solve the latter problem by any other means.

6. By direct multiplication and use of trigonometric formulas, show that the product of two proper orthogonal matrices of the form

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

is a proper orthogonal matrix of the same form, and that the inverse of M is likewise a proper orthogonal matrix of the same form as M. Do these matrices form a group under matrix multiplication?

7. Verify that

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

is an orthogonal matrix.

8. Given that

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

show that $A^3 = A^2A = AA^2 = [1]_3^3$.

9. If $A = [a]_3^3$ has all of its elements in the field **R** of real numbers, and $AA' = 0$, prove that $A = 0$.

6–9 Matrices and rings. If we compare the laws for matrices of Section 5-16 with the postulates for an integral domain, which were stated in Section 1–4, we see that the commutative law for multiplication ($ab = ba$) and the cancellation law for multiplication (if $ab = 0$ then either $a = 0$ or $b = 0$) are missing. Hence square matrices of order n with elements in a field **F** do not form an integral domain. Neither do such matrices form a field, for to do so the two properties just mentioned would be required, and in addition every matrix except zero would require an inverse. We have seen in this chapter that that is not true, but on comparison of the laws which these matrices satisfy and the properties of a ring, which are stated in Section 1–6, we see that the matrices do form a ring. Moreover, they constitute a ring with unit element, for $IA = AI = A$. Therefore, we have

THEOREM XXIV. *The set of all square matrices of order n with elements in a field* **F** *is a ring with unit element.* This ring has divisors of zero (zero being the null matrix); for instance

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

Most of the existing theory of matrices is concerned with matrices with elements in a given field or in an integral domain. But some of the more advanced theory of matrices deals with matrices with elements in an arbitrary commutative ring **R** with unit element. In such theory it is shown that the set of all square matrices of order n with elements in a ring **R** themselves *constitute* a ring $\mathbf{R}_n$.*

* See N. H. McCoy, "Rings and Ideals," Chapter 8.

CHAPTER 7

GROUPS, MATRICES, AND TRANSFORMATIONS

7–1 Groups of matrices. Consider the set S of all *nonsingular square matrices of order n with elements in a given field.* Clearly the set is closed under the operation of matrix multiplication, for if A and B are elements of S, then $C = AB$ is also an element of S. By Theorem IX of Chapter 5 the associative law holds. The set contains a single identity element, the unit matrix I, such that for every element A of S we have $AI = IA = A$. Since the set S is limited to nonsingular square matrices, every element A of S has a unique inverse A^{-1} such that $A^{-1}A = AA^{-1} = I$. Thus the four conditions for a group are satisfied, and we have proved

THEOREM I. *All nonsingular square matrices of order n with elements in a field* **F** *form a group with respect to matrix multiplication.*

The group of all nonsingular square matrices of order n may have subgroups. Thus a subgroup of the group of all nonsingular square matrices of order 2 is (the group operation of course being matrix multiplication)

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

Another subgroup of all nonsingular square matrices of order 2 is the set of all orthogonal matrices of order 2.

A matrix is said to be *complex* if its elements are complex scalars and *real* if its elements are real scalars. As a particularization of Theorem I we have

THEOREM II. *All nonsingular square matrices of order n with elements in the complex field* **C** *form a group.*

THEOREM III. *All nonsingular real square matrices of order n form a group, which is a subgroup of the group of the nonsingular square complex matrices of that order.*

In Theorems XIX and XX of Chapter 6 we showed that the product of two orthogonal matrices of order n is an orthogonal matrix of the same order, and the inverse of an orthogonal matrix is an orthogonal matrix. From these facts and Theorem III there follows

THEOREM IV. *All (real) orthogonal matrices of order n form a group, this being a subgroup of the group of all nonsingular real square matrices of order n.*

EXERCISES

1. Prove that the set of all nonsingular diagonal matrices of order n forms a commutative group with respect to matrix multiplication.

2. Show that the set of all scalar matrices of order n forms a group with respect to matrix multiplication.

7–2 Simple isomorphism of groups. Let $a \longleftrightarrow a'$ represent a *one-to-one correspondence* of a set S to a set S'; by this we mean that every element a in S has one and only one correspondent a' in S', and every element a' in S' is the correspondent of one and only one element in S. Two groups G and G' are said to be *simply isomorphic* if there is a one-to-one correspondence between their elements which preserves group multiplication; that is, if $a \longleftrightarrow a'$ and $b \longleftrightarrow b'$, then $ab \longleftrightarrow a'b'$.

The two permutation groups I, (ab), (cd), $(ab)(cd)$; I, $(ab)(cd)$, $(ac)(bd)$, $(ad)(bc)$ are simply isomorphic, and this simple isomorphism may be signified by the correspondence $I \longleftrightarrow I'$, $(ab) \longleftrightarrow (ab)(cd)$, $(cd) \longleftrightarrow (ac)(bd)$, $(ab)(cd)$ $(ad) \longleftrightarrow (bc)$.

The six matrices

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

are simply isomorphic with the elements of the symmetric group on three letters

$$I, \quad (abc), \quad (acb), \quad (ac), \quad (ab), \quad (bc)$$

in the given order, so that like products always correspond.

7–3 Matric representation of groups. As far as abstract properties are concerned, such as subgroups of various kinds and a multiplication table, two simply isomorphic groups are identical. If the elements of a group are just symbols which have no interpretation except as elements of the group, the group is called an *abstract group*. If the elements of a group are quantities or operations of a special kind, such as matrices or permutations, the group is called a *special group*, and the special group is said to form a *representation* of the simply isomorphic abstract group or of any simply isomorphic special group. In particular, the representation of an abstract group or of a given special group by a group of matrices is called a *matric representation* of the group. It is convenient to let $M(s_1)$, $M(s_2)$, . . . , $M(s_m)$ designate the matrices which form a matric representation of a group of order m with elements $s_1, s_2, \ldots, s_m$. If we can associate with each element s_i of the group G, which has the elements $s_1, s_2, \ldots, s_m$, a square matrix $M(s_i)$ of order n such that

$$M(s_i)M(s_j) = M(s_is_j) \tag{1}$$

for every s_i, s_j in G, then the matrices $M(s_1)$, $M(s_2)$, . . . , $M(s_m)$ constitute a *matric representation of order n* of the group G; we sometimes speak of this as an *n-dimensional matric representation* of the group. A set of matrices with the property (1) relative to a group G necessarily forms a group.

7–4 Linear transformations. Let $\alpha = \{x_1, x_2, \ldots, x_n\}$ be an arbitrary vector of an n-dimensional vector space $V_n(\mathbf{F})$. Suppose the scalar coordinates of α are related to the coordinates of another vector $\bar{\alpha} = \{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n\}$ by means of the scalar relations

$$\begin{aligned} \bar{x}_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n, \\ \bar{x}_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n, \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \\ \bar{x}_n &= a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n. \end{aligned} \tag{2}$$

Equivalently we may say that the vector α is related to the vector $\bar{\alpha}$ by the transformation

$$\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \cdot \\ \bar{x}_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ x_n \end{bmatrix} \quad \text{or} \quad \bar{\alpha} = A\alpha. \tag{3}$$

We say that the transformation (2) or (3) is a *linear transformation* in the n scalars $x_1, x_2, \ldots, x_n$. The matrix equation (3) may be interpreted as defining a transformation $\mathbf{T}$ conveniently signified by*

$$\alpha \rightarrow \mathbf{T}\alpha = \bar{\alpha}, \tag{4}$$

* In this section we depart from our usual practice of reserving the use of capital letters in bold face type to represent an underlying field or ring, for here we find it convenient to represent transformations somewhat abstractly by such bold face type.

which carries the vector α into the vector $\bar{\alpha}$. The matrix A of (3) is sometimes called the *presenting matrix* of the linear transformation (4).

The linear transformation **T** has two important properties:

(*i*) The linear transformation of $r\alpha$ is effected by

$$A(r\alpha) = r(A\alpha). \tag{5}$$

That is, the transform $\mathbf{T}r\alpha$ of the product of the scalar r and the vector α is equal to $r\mathbf{T}\alpha$, and the latter is r times the transform of α.

(*ii*) Consider the sum $\alpha + \beta$. The linear transformation of this sum is given by

$$A(\alpha + \beta) = A\alpha + A\beta. \tag{6}$$

The relation (6) says that the linear transform of the sum of two vectors is the sum of their linear transforms. These results may be expressed by saying that the linear transformation **T** has the properties

$$\mathbf{T}r\alpha = r\mathbf{T}\alpha, \tag{7}$$

$$\mathbf{T}(\alpha + \beta) = \mathbf{T}\alpha + \mathbf{T}\beta; \tag{8}$$

or, compositely written, *the linear transformation* **T** *has the property*

$$\mathbf{T}(r\alpha + s\beta) = r\mathbf{T}\alpha + s\mathbf{T}\beta. \tag{9}$$

Conversely, it can be shown that the property (9) of a linear transformation **T** implies that this transformation is of the form (3).

The linear transformation (3) may be interpreted as effecting a transition from one nonhomogeneous coordinate systm α to another such coordinate system $\bar{\alpha}$; that is, the vector coordinates α and $\bar{\alpha}$ of the same point in two coordinate systems are linked by the transformation $\mathbf{T}: \bar{\alpha} = A\alpha$. Another interpretation of the transformation **T** is that it describes a renaming of the vectors with respect to a fixed coordinate system, or that it is a linear transformation of the vector space into itself; such a transformation is termed a *linear mapping* of the vector space on itself, and we call the linear vector space itself the *carrier space*.

The *product of two linear transformations*

$$\mathbf{T}: \bar{\alpha} = A\alpha \qquad \text{and} \qquad \mathbf{U}: \bar{\bar{\alpha}} = B\bar{\alpha} \tag{10}$$

is defined to be their successive applications **UT**; that is, as in group theory, first apply **T** and then apply **U**.

THEOREM V. *The product of two linear transformations is a linear transformation.*

By definition, a product transformation **UT** effects the transformation

$$\alpha \to (\mathbf{UT})\alpha = \mathbf{U}[\mathbf{T}\alpha],$$

whence

$$\begin{aligned}\mathbf{UT}[r\alpha + s\beta] &= \mathbf{U}[\mathbf{T}(r\alpha + s\beta)] = \mathbf{U}[\mathbf{T}r\alpha + \mathbf{T}s\beta] \\ &= \mathbf{U}[\mathbf{T}r\alpha] + \mathbf{U}[\mathbf{T}s\beta],\end{aligned}$$

or

$$\mathbf{UT}[r\alpha + s\beta] = \mathbf{UT}r\alpha + \mathbf{UT}s\beta = r[(\mathbf{UT})\alpha] + s[(\mathbf{UT})\beta].$$

Therefore the product **UT** has the characteristic property (9) of a linear transformation.

It follows that the product transformation **UT** may be represented by

(11) $$\mathbf{UT}: \bar{\bar{\alpha}} = B(A\alpha) \qquad \text{or} \qquad \bar{\bar{\alpha}} = BA\alpha.$$

We then have

THEOREM VI. *If we pass from the coordinate vector α to the coordinate vector $\bar{\alpha}$ by a linear transformation* **T** *with matrix A, and from the coordinate vector $\bar{\alpha}$ to the coordinate vector $\bar{\bar{\alpha}}$ by another linear transformation* **U** *with matrix B, then the linear transformation* **UT** *with matrix BA will transform α directly into $\bar{\bar{\alpha}}$.*

The linear transformation $\mathbf{T}: \bar{\alpha} = A\alpha$ which takes the vector α into the vector $\bar{\alpha}$ is said to be *reversible* if there exists a related transformation $\mathbf{T}^{-1}: \alpha = A^{-1}\bar{\alpha}$ which takes $\bar{\alpha}$ into α. The transformation $\mathbf{T}^{-1}$ is called the *inverse* of **T**. The combined effect of a transformation **T** followed by its inverse $\mathbf{T}^{-1}$ of a vector α is to leave that vector unchanged.

Proceeding as in the proof of Theorem V, we may establish

THEOREM VII. *The inverse of a linear transformation is a linear transformation.*

It should now be clear that all reversible linear transformations of an n-dimensional vector space into itself have the properties that the product of two such transformations is a linear transformation, and that the inverse of a linear transformation is also a linear transformation; moreover, we have seen that there exists an identity element $\mathbf{T}^{-1}\mathbf{T}$, and it should be apparent that the associative law holds. These facts establish

THEOREM VIII. *The set of all reversible linear transformations* $\mathbf{T}_1$, $\mathbf{T}_2$, . . . *of an n-dimensional vector space into itself form a group, the* *group operation being the sequential performance of two transformations as* $\mathbf{T}_j\mathbf{T}_i$.

Further we have observed the existence of a one-to-one correspondence of the group of all reversible linear transformations $\mathbf{T}_1, \mathbf{T}_2, \ldots$, of an n-dimensional vector space into itself and the group of all nonsingular square matrices of order n, $A_1, A_2, \ldots$, such that if

$$\mathbf{T}_i \longleftrightarrow A_i \quad \text{and} \quad \mathbf{T}_j \longleftrightarrow A_j, \quad \text{then} \quad \mathbf{T}_j\mathbf{T}_i \longleftrightarrow A_jA_i.$$

We have then

THEOREM IX. *The group of all reversible linear transformations of an n-dimensional vector space* $V_n(\mathbf{F})$ *into itself is simply isomorphic with the group of all nonsingular square matrices of order n with elements in the field* $\mathbf{F}$.

It is worthy of emphasis that when we speak of the existence of a simple isomorphism between linear transformations and matrices, we have reference to *linear transformations of a vector space into itself;* that is, the coordinate system is fixed. For if we pass from one coordinate system to another, the same linear transformation may correspond to several matrices, and one matrix may correspond to many linear transformations.

The group of all reversible linear transformations in the n-dimensional vector space $V_n(\mathbf{C})$ is called the *general linear group* and is usually denoted by $GL(n, \mathbf{C})$. Because of the simple isomorphism of linear transformations and related groups of matrices, the same symbols and terminology are sometimes used for linear transformation groups and matric groups; subsequently we shall adhere to that custom. Both the special terminology of linear transformations and that of matrices contribute to this common terminology. For example, we shall speak of "linear groups of matrices"; and we shall call a reversible linear transformation a "nonsingular transformation," its presenting matrix being nonsingular. Accordingly, we refer to the group of all n by n nonsingular matrices with complex elements as the *general linear group of order n,* and denote this group of matrices by $GL(n, \mathbf{C})$.

We noted in Theorem III that all nonsingular real square matrices of order n form a group; this group is called the *real linear group of order n,* and is commonly designated by $RL(n)$. From Theorem IV we know that all (real) orthogonal matrices form a subgroup of $RL(n)$; this is termed the *real orthogonal group* $0(n)$.

7-5 Cogredient and contragredient vectors. Sometimes when a given vector α transforms by a linear transformation with matrix A, another vector transforms by a linear transformation with matrix $(A^{-1})'$. We now study such situations. Let the column vector α undergo the nonsingular linear transformation

$$\alpha = A\bar{\alpha}. \tag{12}$$

Introduce the row vector $\xi' = (u_1, u_2, \ldots, u_n)$ and consider the effect of the transformation (12) on the product

$$\begin{aligned} \xi'\alpha &= (u_1, u_2, \ldots, u_n)\{x_1, x_2, \ldots, x_n\} \\ &= u_1x_1 + u_2x_2 + \cdots + u_nx_n. \end{aligned} \tag{13}$$

In this connection recall that near the end of Section 5-6 we agreed to use unprimed Greek letters to represent column vectors. Also, recall that A' means the transpose of the matrix A, and α' is the row vector which is the transpose of the column vector α. Under the influence of the transformation (12) the product (13) is transformed as follows:

$$\xi'\alpha = \xi'A\bar{\alpha}. \tag{14}$$

If we denote the product on the right side of equation (14) by $\bar{\xi}'\bar{\alpha}$, then $\bar{\xi}' = \xi'A$. Taking the transpose of both sides of the latter, we get $\bar{\xi} = A'\xi$. Solving the last equation by premultiplying both sides by the inverse of the transpose of A, we get

$$\xi = (A^{-1})'\bar{\xi}. \tag{15}$$

We thus see that if the vector α of the *form* or product (13) is transformed by the matrix A, then the vector ξ, whose transpose appears in the form (13), is transformed by the transpose of the inverse of A if the product $\xi'\alpha$ remains unchanged. In general, two n-dimensional vectors are called *contragredient* if, when one is subjected to a nonsingular linear transformation, the other is subjected to the transformation which has for its matrix the transpose of the inverse of the matrix of the first transformation.

The product or form (13) may be interpreted in different ways. If we interpret the u's as parametric coefficients in a linear form in the variables x's, we have

THEOREM X. *The coefficients of a linear form transform contragrediently to the variables in that form.*

However, it is not necessary to consider the u's of (13) as parametric constants and the x's as variables. When the u's do not all vanish, the equation

$$(16) \qquad \xi'\alpha = 0 \qquad \text{or} \qquad u_1x_1 + u_2x_2 + \cdots + u_nx_n = 0,$$

for constant u's, defines a hyperplane, that is, an $(n - 1)$ subspace in the vector space V_n. A vector α lies in the hyperplane if its elements satisfy the equation (16). On the other hand, if the x's are interpreted as constants, it is appropriate and usual to consider the u's as the scalar coordinates of a vector ξ in a second n-dimensional vector space, the *dual space* of V_n, which we represent by V_n^*. Duality of vector spaces is a reciprocal relationship; a change of coordinates in one of two dual spaces is automatically connected with the contragredient change of the coordinates in the other space. Examples of contragredient variables abound in analytic projective geometry where homogeneous coordinates are commonly used.

When two vectors $\alpha = \{x_1, x_2, \ldots, x_n\}$ and $\beta = \{y_1, y_2, \ldots, y_n\}$ are related to the vectors $\bar{\alpha} = \{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n\}$, $\bar{\beta} = \{\bar{y}_1\ \bar{y}_2, \ldots, \bar{y}_n\}$ by the same linear transformations

$$\alpha = A\bar{\alpha} \qquad \text{and} \qquad \beta = A\bar{\beta},$$

the two vectors α and β (or equally well $\bar{\alpha}$ and $\bar{\beta}$) are said to be *cogredient* vectors. The coordinates of two points $\alpha = \{x_1, x_2, x_3\}$ and $\beta = \{y_1, y_2, y_3\}$ in ordinary 3-space furnish an illustration of cogredient vectors.

7–6 Similar matrices. Matric transformations. The same linear transformation **T** may be represented by different matrices, the particular matric representation depending on the choice of coordinates. Thus for $n = 2$, the transformation

$$\begin{aligned} x_1 &= 3\bar{x}_1 + 2\bar{x}_2 \\ x_2 &= 2\bar{x}_1 + 3\bar{x}_2 \end{aligned} \qquad \text{or} \qquad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}$$

is represented by the matrix

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

but relative to the coordinates $y_1 = x_1 + x_2$, $y_2 = x_1 - x_2$ the same transformation is effected by

$$\begin{aligned} y_1 &= 5\bar{y}_1 \\ y_2 &= \bar{y}_2 \end{aligned} \qquad \text{or} \qquad \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \end{bmatrix}, \quad \text{which is represented by the matrix} \quad B = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}.$$

In general, two n by n matrices are called *similar* if they represent the same linear transformation of a vector space $V_n(\mathbf{F})$ relative to different bases.

In Section 4 6 we saw that the unit vectors

$$\epsilon_1 = \{1, 0, 0, \ldots, 0\}, \quad \epsilon_2 = \{0, 1, 0, \ldots, 0\}, \quad \ldots, \quad \epsilon_n = \{0, 0, 0, \ldots, 1\}$$

constitute a basis in which the scalar coordinates of the vector $\alpha = \{x_1, x_2, x_3, \ldots, x_n\}$ are the coefficients in the linear expression of the vector in terms of the basis:

$$(17) \qquad \alpha = \epsilon_1 x_1 + \epsilon_2 x_2 + \cdots + \epsilon_n x_n = E'X,$$

where

$$E' = [\epsilon_1, \epsilon_2, \ldots, \epsilon_n], \quad X = \{x_1, x_2, \ldots, x_n\}.$$

Let the scalar coordinates $x_1, x_2, \ldots, x_n$ of the vector α be subjected to the transformation

$$(18) \qquad X = A\bar{X}, \qquad \text{where} \qquad A = [a]_n^n.$$

If we substitute from (18) in (17), we get $\alpha = E'A\bar{X}$. Renaming the right side of the relation by $\bar{E}'\bar{X}$, that is, letting $\bar{E}'\bar{X} = E'A\bar{X}$, then $\bar{E}' = E'A$. Taking the transpose of the latter relation, we have $\bar{E} = A'E$; solving this for E, we obtain

$$(19) \qquad E = (A^{-1})'\bar{E}.$$

The transformation matrix $(A^{-1})'$ of equation (19) is the inverse of the transpose (or equally the transpose of the inverse) of the transformation matrix A of equation (18).

It should be recognized that E is a square matrix of order n with the ϵ's for columns. So (19) is not the transformation of a vector, as we have been accustomed to use that term, but the transformation of a matrix. Let us broaden our definition of contragredience of the preceding section to say that if a vector β is transformed in the manner

$$(20) \qquad \beta = A\bar{\beta} \quad \text{with } A \text{ as transformation matrix,}$$

and a matrix C is transformed in the manner

$$(21) \qquad C = (A^{-1})'\bar{C} \quad \text{with } (A^{-1})' \text{ as transformation matrix,}$$

then the *matrix transformation* (21) *is contragredient to the vector transformation* (20). From this convention and the facts of the preceding paragraph there follows

THEOREM XI. *The transformation of the coordinates of a vector space V_n is contragredient to the transformation of the bases for V_n.*

From Theorem XI we see that a change of basis in an n-dimensional vector space V_n with nonsingular matrix P as (note the transposition of the matric entities)

$$\bar{E}' = E'P \qquad \text{or} \qquad \bar{E} = P'E \tag{22}$$

induces the transformation

$$\bar{X} = P^{-1}X \qquad \text{or} \qquad X = P\bar{X} \tag{23}$$

on the coordinates of a vector of V_n. Let

$$Y = AX \tag{24}$$

be a transformation of the coordinates $X = \{x_1, x_2, \ldots, x_n\}$ of a vector of V_n when the *fixed basis* of V_n is $E' = [\epsilon_1, \epsilon_2, \ldots, \epsilon_n]$. Let us determine the form of the coordinate transformation (24) when the basis is changed from $E' = [\epsilon_1, \epsilon_2, \ldots, \epsilon_n]$ to $\bar{E}' = [\bar{\epsilon}_1, \bar{\epsilon}_2, \ldots, \bar{\epsilon}_n]$ by (22). In the old coordinate transformation (24), relative to the basis E', we must replace X by $P\bar{X}$ and Y by $P\bar{Y}$. These substitutions in (24) give

$$P\bar{Y} = AP\bar{X} \qquad \text{or} \qquad \bar{Y} = P^{-1}AP\bar{X}. \tag{25}$$

Hence a change of basis of V_n with matrix P of the transformation (22) replaces the matrix A of the transformation (24) by

$$B = P^{-1}AP. \tag{26}$$

This proves

THEOREM XII. *Two n by n matrices A and B are similar if and only if $B = P^{-1}AP$ for some nonsingular matrix P.*

It is now evident that in order to study those properties of linear transformations of a vector space V_n into itself which do not depend on the basis of V_n over a given field **F**, we need only to study those properties of related square matrices which are invariant under the matric transformation $B = P^{-1}AP$, that is, properties of similar matrices.

When $B = P^{-1}AP$, B is called the *transform* of A by P under the *similarity transformation*. Some properties of the matrix A are unchanged when it is transformed by the matrix P, and if the transform B is simpler than the original matrix A, we may advantageously study B rather than A. Certain very simple forms of matrices are called *canonical forms*, and an important problem is to determine such forms for a given matrix, and to find a matrix P which will transform a matrix A to a given canonical form.

THEOREM XIII. *Similar matrices have the same determinant.*

Clearly, if $B = P^{-1}AP$, then

$$|B| = |P^{-1}AP| = |P^{-1}| \cdot |A| \cdot |P| = |P^{-1}| \cdot |P| \cdot |A|$$
$$= \frac{1}{|P|} \cdot |P| \cdot |A| = |A|.$$

CHAPTER 8

THE CHARACTERISTIC EQUATION OF A MATRIX

8–1 The characteristic matrix of a given matrix. Associated with every square matrix of order n there is the matrix $(A - \lambda I)$, where λ is a variable scalar and I is the unit matrix of order n, called the *characteristic matrix* of A. This matrix is obtained by subtracting λ from each element of the leading diagonal of A; thus for $A = [a]_2^2$,

$$A - \lambda I = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{bmatrix}.$$

8–2 The characteristic function of a matrix. The determinant of the characteristic matrix of A, $|A - \lambda I|$, is called the *characteristic determinant* of A; when this determinant is expanded in powers of $-\lambda$ we obtain a scalar polynomial of degree n in the scalar λ,

$$(1) \quad f(\lambda) = (-1)^n[\lambda^n - p_1\lambda^{n-1} + p_2\lambda^{n-2} - p_3\lambda^{n-3} + \cdots + (-1)^n p_n].$$

This scalar polynomial $f(\lambda) = |A - \lambda I|$ is called the *characteristic function* of the matrix A. From (1), $f(0) = (-1)^n(-1)^n p_n$, and so $f(0) = p_n$ for *n even* or *odd*. From $f(\lambda) = |A - \lambda I|$, we have $f(0) = |A|$. Therefore, $p_n = |A|$. To illustrate, for $n = 2$, $f(\lambda) = \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21}$.

Exercises

1. Show for $A = [a]_3^3$ that $f(\lambda) = (-1)^3[\lambda^3 - p_1\lambda^2 + p_2\lambda - p_3]$, where p_1 is the sum of the elements of the principal diagonal of A, p_2 is the sum of the two-rowed principal minors of A, and $p_3 = |A|$.

2. Show for $A = [a]_4^4$ that $f(\lambda) = (-1)^4[\lambda^4 - p_1\lambda^3 + p_2\lambda^2 - p_3\lambda + p_4]$, where p_1 is the sum of the elements of the principal diagonal of A, p_2 is the sum of the two rowed principal minors, p_3 is the sum of the three rowed principal minors, and $p_4 = |A|$.

It is a fact * that in general the coefficient p_h of the characteristic function $f(\lambda)$ is the sum of the h-rowed principal minors of the matrix A. Recall that in a square matrix of order n, $A = [a]_n^n$, the elements of the type a_{ii}, where

---

* See A. K. Mitchel, "A note on the characteristic determinant of a matrix," *American Mathematical Monthly*, **38** (1931), pp. 386–388; or C. E. Cullis, *Matrices and Determinoids*, **III**, pp. 307–308.

$i = 1, 2, \ldots, n$, are said to lie in the *principal diagonal* of the matrix. A minor obtained from $|A|$ by the deletion of the ith row and the ith column is called an $(n-1)$th rowed *principal minor;* a minor obtained by the deletion of the ith and jth rows and the ith and jth columns is called an $(n-2)$th rowed principal minor; and generally the minor formed by the deletion of the i_1th, i_2th, ..., i_rth rows and the same columns is called an *$(n-r)$th rowed principal minor* of A. These principal minors of A are situated symmetrically with respect to the principal diagonal of A. In particular, the elements of the principal diagonal are 1-rowed principal minors, their cofactors are $(n-1)$th rowed principal minors, and $|A|$ itself is the one n-rowed principal minor.

8–3 The characteristic equation of a matrix. The polynomial equation of degree n in the indeterminate scalar obtained by setting the characteristic function of the matrix $A = [a]_n^n$ equal to zero, $|A - \lambda I| = 0$, or using the expanded form (1) and dividing by $(-1)^n$,

$$\lambda^n - p_1\lambda^{n-1} + p_2\lambda^{n-2} - p_3\lambda^{n-3} + \cdots + (-1)^n p_n = 0, \tag{2}$$

is called the *characteristic equation* of the matrix A. The n roots $\lambda_1, \lambda_2, \ldots, \lambda_n$ of the characteristic equation of a matrix A are called the *characteristic roots* of the matrix A. To illustrate, for the square matrix of the second order with $\alpha = \{5, -2\}$ and $\beta = \{-2, 2\}$ as column vectors, the characteristic equation is $\lambda^2 - 7\lambda + 6 = 0$, and the characteristic roots of the given matrix are $\lambda_1 = 1, \lambda_2 = 6$.

THEOREM I. *Similar matrices have the same characteristic function and the same characteristic equation.*

Recall that A and B are similar if $B = P^{-1}AP$. Then

$$\begin{aligned}|B - \lambda I| &= |P^{-1}AP - \lambda I| = |P^{-1}AP - \lambda P^{-1}IP| \\ &= |P^{-1}(A - \lambda I)P| = |A - \lambda I|.\end{aligned}$$

We have shown that the characteristic function of two similar matrices are equal; it follows that they have the same characteristic equations and the same characteristic roots. As an immediate consequence of Theorem I, we have

THEOREM II. *The coefficients of the characteristic equation of a matrix A are invariants under any similarity transformation of that matrix.*

As an illustration of Theorem II, for $A = [a]_2^2$ and $B = [b]_2^2$, $B = P^{-1}AP$, we have

$$\lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21}$$
$$= \lambda^2 - (b_{11} + b_{22})\lambda + b_{11}b_{22} - b_{12}b_{21}.$$

Since this relation holds for every value of the indeterminate λ, we have

$$a_{11}a_{22} - a_{12}a_{21} = b_{11}b_{22} - b_{12}b_{21} \quad \text{and} \quad a_{11} + a_{22} = b_{11} + b_{22}.$$

The first of these two relations is simply a verification of the fact that $|A| = |B|$, which was proved generally in Theorem XIII of the last chapter. But the second relation is new. The sum of the diagonal elements of any square matrix is called its *trace*, and is written tr A. Similarly, we see from Theorem II and Ex. 1 of Section 8–2 that the square matrix of order 3, $A = [a]_3^3$, has three invariants under the similarity transformation, its trace $= p_1$, the sum of the two-rowed principal minors $= p_2$, and its determinant $= p_3$.

Of the n invariant coefficients of the characteristic equation of a matrix $A = [a]_n^n$ under a similarity transformation, tr A and $|A|$ are commonly given special attention. We have previously emphasized the invariance of $|A|$ in this connection. We now give like emphasis to tr A in

THEOREM III. *The traces of similar matrices are equal.*

The concept of the trace of a matrix is of fundamental importance in the theory of group representations by matrices.

EXERCISES

Find the characteristic equations and the characteristic roots of each of the following matrices.

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

4. Corresponding to the scalar equation $\lambda^2 - 7\lambda + 6 = 0$, the characteristic equation of the numerical matrix A given immediately preceding Theorem I, there is the matric polynomial equation $X^2 - 7X + 6I = O$, where I and O are the unit matrix and the zero matrix, respectively, with order equal to that of A. Show that the matrix A in question satisfies the latter equation.

5. Show that the matrix $A = [a]_2^2$ satisfies the matrix polynomial equation $F(X) = X^2 - (a_{11} + a_{22})X + (a_{11}a_{22} - a_{12}a_{21})I = O$, I and O being the second order unit and null matrices.

6. From Ex. 5 we have that any second order square matrix A satisfies

the relation $A^2 - (a_{11} + a_{22})A + (a_{11}a_{22} - a_{12}a_{21})I = O$. Solving the latter relation for I, and then multiplying by A^{-1}, we obtain ($|A| \neq 0$)

$$A^{-1} = -\frac{1}{|A|}[A - (a_{11} + a_{22})I],$$

which is a formula for A^{-1} in terms of A and the coefficients of the characteristic function of A. Using this formula, calculate the inverse of each of the following matrices. In each case verify the correctness of your result by showing that $A^{-1}A = I$.

$$(i)\ A = \begin{bmatrix} 3 & 4 \\ 1 & -2 \end{bmatrix}, \qquad (ii)\ A = \begin{bmatrix} 5 & 1 \\ 2 & 4 \end{bmatrix}.$$

8-4 The Cayley-Hamilton Theorem. In Ex. 6 of the last section we saw that if $f(X)$ is the matrix polynomial corresponding to the characteristic function $f(\lambda)$ of the matrix $A = [a]_2^2$, then $F(A) = O$. This is a special instance of one of the most famous theorems of matrix algebra,

THE CAYLEY-HAMILTON THEOREM. *Let*

(3) $f(\lambda) = (-1)^n[\lambda^n - p_1\lambda^{n-1} + p_2\lambda^{n-2} - p_3\lambda^{n-3} + \cdots + (-1)^n p_n]$

be the characteristic function of a square matrix A of order n. Then the matric polynomial equation

(4) $F(X) = X^n - p_1X^{n-1} + p_2X^{n-2} - p_3X^{n-3} + \cdots + (-1)^n p_n I = O$

is satisfied by $X = A$.

The Cayley-Hamilton Theorem was proved for $n = 2$ and $n = 3$ by Cayley in 1858; he also stated the theorem in the general case with the comment that he thought it unnecessary to undertake a proof of it. Hamilton's share in the title of the theorem is based on the fact that he had established a comparable result for quaternions in 1853. Many proofs of the theorem for the general case appear in the literature on matrix algebra. One form of such proof depends on the adjoint of the matrix $(A - \lambda I)$, and that we now give.

Let B denote the adjoint of the characteristic matrix $(A - \lambda I)$. Since the elements of $(A - \lambda I)$ are at most of the first degree in λ, their cofactors in $|A - \lambda I|$ are at most of degree $(n - 1)$ in λ. Hence a typical element b_{ij} of the adjoint matrix B can be written as

$$b_{ij} = k_0 + k_1\lambda + \cdots + k_{n-1}\lambda^{n-1}$$

where the coefficients $k_0, k_1, \ldots, k_{n-1}$ are polynomial functions of the elements a_{ij} of A. Consequently the matrix B itself can be written as the matrix polynomial

$$B = \operatorname{adj}(A - \lambda I) = B_0 + B_1\lambda + \cdots + B_{n-1}\lambda^{n-1}.$$

Here the B_i's are matrices whose elements are polynomials in the a_{ij}'s.

In Section 6–5 we saw that for any matrix C

$$C \cdot \operatorname{adj} C = |C| \cdot I.$$

Hence

$$(A - \lambda I)B = |A - \lambda I| \cdot I = f(\lambda)I,$$

since

$$f(\lambda) = |A - \lambda I|.$$

Substituting in this relation the above polynomial expression for B and the expanded expression (3) for $f(\lambda)$, we get

$$\begin{aligned}(A - \lambda I)&(B_0 + B_1\lambda + B_2\lambda^2 + \cdots + B_{n-2}\lambda^{n-2} + B_{n-1}\lambda^{n-1}) \\ &= (-1)^n[\lambda^n - p_1\lambda^{n-1} + p_2\lambda^{n-2} + \cdots + (-1)^{n-1}p_{n-1}\lambda + (-1)^n p_n]I.\end{aligned}$$

This is an identity which is true for all values of λ. Equating corresponding coefficients of the powers $\lambda^n, \lambda^{n-1}, \ldots, \lambda^1, \lambda^0$, we obtain

$$\begin{aligned}
-B_{n-1} &= (-1)^n I, \quad (\text{since } IB_{n-1} = B_{n-1}) \\
-B_{n-2} + AB_{n-1} &= -(-1)^n p_1 I, \\
-B_{n-3} + AB_{n-2} &= (-1)^n p_2 I, \\
&\cdots \\
-B_1 + AB_2 &= (-1)^n(-1)^{n-2}p_{n-2}I, \\
-B_0 + AB_1 &= (-1)^n(-1)^{n-1}p_{n-1}I, \\
AB_0 &= (-1)^n(-1)^n p_n I.
\end{aligned}$$

If we premultiply these matric relations by $A^n, A^{n-1}, A^{n-2}, \ldots, A^2, A, I$, respectively, we get

$$\begin{aligned}
-A^nB_{n-1} &= (-1)^nA^n, \\
-A^{n-1}B_{n-2} + A^nB_{n-1} &= -(-1)^n p_1 A^{n-1}, \\
-A^{n-2}B_{n-3} + A^{n-1}B_{n-2} &= (-1)^n p_2 A^{n-2}, \\
&\vdots \\
-A^2B_1 + A^3B_2 &= (-1)^n(-1)^{n-2}p_{n-2}A^2, \\
-B_0A + A^2B_1 &= (-1)^n(-1)^{n-1}p_{n-1}A, \\
AB_0 &= (-1)^n(-1)^n p_n I.
\end{aligned}$$

On addition of these matric equalities, the terms on the left annul one another. We then have

$$(-1)^n[A^n - p_1A^{n-1} + p_2A^{n-2} + \cdots + (-1)^{n-1}p_{n-1}A + (-1)^n p_n I] = O,$$

or simply,

$$A^n - p_1A^{n-1} + p_2A^{n-2} + \cdots + (-1)^{n-1}p_{n-1}A + (-1)^n p_n I = O,$$

and the Cayley-Hamilton Theorem is proved. Note that this proof puts no restriction on the matrix A (A may be singular or nonsingular), and no restriction on the characteristic roots of A (these roots may be distinct, or some may be repeated).

The Cayley-Hamilton Theorem is often stated in the abbreviated and implicit form, "A matrix satisfies its own characteristic equation." Such a statement makes sense only if we keep in mind the correspondence between scalars and scalar matrices discussed in Section 5-15. It is usual to refer to either the scalar polynomial $f(\lambda)$ given by (3) or the matrix polynomial $F(X)$ given by (4) as the characteristic function of the matrix A, the context reflecting to which reference is being made.

Since we can factorize $f(\lambda)$ in terms of the characteristic roots λ_i in the manner

$$f(\lambda) = (-1)^n(\lambda - \lambda_1)(\lambda - \lambda_2)\ldots(\lambda - \lambda_n),$$

it follows that

$$F(A) = (A - \lambda_1 I)(A - \lambda_2 I)\ldots(A - \lambda_n I),$$

where the order of the matric factors is immaterial. Therefore by the Cayley-Hamilton Theorem the product of the n matrices $A - \lambda_i I$ is the null matrix, although it does not follow that any one of the matrices $A - \lambda_i I$ is the null matrix. To illustrate, for the matrix

$$(5) \qquad A = \begin{bmatrix} 11 & -6 & 2 \\ -6 & 10 & -4 \\ 2 & -4 & 6 \end{bmatrix},$$

the characteristic equation is $\lambda^3 - 27\lambda^2 + 180\lambda - 324 = 0$, with 3, 6, and 18 for roots. Then

$$(6) \qquad A^3 - 27A^2 + 180A - 324I = O,$$

or

$$(7) \qquad (A - 3I)(A - 6I)(A - 18I) = O.$$

The relation (7) displayed in detail is

$$(8) \quad \begin{bmatrix} 8 & -6 & 2 \\ -6 & 7 & -4 \\ 2 & -4 & 3 \end{bmatrix}\begin{bmatrix} 5 & -6 & 2 \\ -6 & 4 & -4 \\ 2 & -4 & 0 \end{bmatrix}\begin{bmatrix} -7 & -6 & 2 \\ -6 & -8 & -4 \\ 2 & -4 & -12 \end{bmatrix} = O.$$

Let A be a square matrix of order n with elements in a given field $\mathbf{F}$. The successive powers $I, A, A^2, A^3, \ldots, A^i$ may or may not be linearly independent; that depends, for one thing, on the value of i. If m is the smallest integer for which the powers $I, A, A^2, \ldots, A^m$ are linearly dependent, then there is an equation

$$(9) \quad M(X) = X^m + c_1X^{m-1} + c_2X^{m-2} + \cdots + c_{m-1}X + c_mI = O$$

with coefficients in $\mathbf{F}$ which is satisfied by the matrix A. This equation is called the *minimum equation* of A, the integer m is called the *index* of A, and the function $M(X)$ is called the *minimum function* of the matrix A. The index of a scalar matrix is 1; every matrix except the null matrix has an index. The Cayley-Hamilton Theorem tells us that $m \leqq n$. That is, as a consequence of the Cayley-Hamilton Theorem, we have

THEOREM IV. *Any positive integral power A^p of the square matrix A of order n for $p \geqq n$ is linearly expressible in terms of the unit matrix of order n and the first $(n - 1)$ powers of A.*

To illustrate, for the numerical matrix given by equation (5), from equation (6) we may obtain A^3 in terms of lower powers of A and I; multiplying the expression just referred to by A, we obtain a like expression for A^4, and so on. Further, since any positive integral power of A is expressible in terms of I and the first $(n - 1)$ powers of A, so is any rational integral function of A.

For a given matrix A the minimum function $M(X)$ is unique up to a nonzero scalar factor. We have seen that the Cayley-Hamilton Theorem tells us that the index $m \leqq n$, and clearly when $m = n$, $M(X) = kF(X)$; that is, when $m = n$ the minimum function differs from the characteristic function only by a scalar multiplier. We agree to choose the constant scalar k, so that the coefficient of the highest degree term is 1. Thus for the numerical matrix A given by (5) above, we see from (6) that the minimum function $M(X)$ is given by $M(X) = X^3 - 27X^2 + 180X - 324I$. So for this matrix A the index m is 3.

A matrix A for which $m < n$ is called *derogatory*. For such a matrix it is proved in advanced matrix theory* that the minimum function is a factor of the characteristic function of A, that is,

* See H. W. Turnbull and A. C. Aitken, *An Introduction to the Theory of Canonical Matrices*, Blackie and Son, London (1932), pp. 46–48.

$F(X) = M(X)G(X)$. For this reason the minimum function of a matrix A is sometimes called the *reduced characteristic function* of A; the modifier "reduced" is interpreted to include the possibility that $F(X)$ and $M(X)$ are identical.

Further, it is established in matrix theory that a matrix A is derogatory ($m < n$) only when the characteristic function of A is not a product of distinct irreducible factors.* That is, for the matrix A to be derogatory, the characteristic equation of A must have multiple roots. To illustrate derogatory matrices, consider the matrix

$$A = \begin{bmatrix} 7 & 4 & -4 \\ 4 & -8 & -1 \\ -4 & -1 & -8 \end{bmatrix} \tag{10}$$

with the characteristic equation

$$f(\lambda) = \lambda^3 + 9\lambda^2 - 81\lambda - 729 = 0. \tag{11}$$

The characteristic roots are $-9, -9, 9$; and hence the Cayley-Hamilton Theorem tells us that $A^3 + 9A^2 - 81A - 729I = O$. But since $f(\lambda)$ has a double root, the matrix A may be derogatory with a minimum equation of degree lower than three. That such is the case is readily and explicitly verified, for

$$(A + 9I)(A - 9I) = \begin{bmatrix} 16 & 4 & -4 \\ 4 & 1 & -1 \\ -4 & -1 & 1 \end{bmatrix} \begin{bmatrix} -2 & 4 & -4 \\ 4 & -17 & -1 \\ -4 & -1 & -17 \end{bmatrix} = O. \tag{12}$$

So for this matrix A the minimum function is not the characteristic function $F(X) = X^3 + 9X^2 - 81X - 729I$, but the reduced characteristic function $M(X) = X^2 - 81I$, and the index of A is 2. From (12), $A^2 - 81I = O$, $I = A^2/81$; so $A^{-1} = A/81$.

The Cayley-Hamilton Theorem holds for a square matrix A of order n whether or not A is singular. If A is nonsingular then $|A| \neq 0$, and A^{-1} exists. In these circumstances we have the following symbolic statement of the Cayley-Hamilton Theorem:

$$A^n - p_1A^{n-1} + p_2A^{n-2} - p_3A^{n-3} + \cdots + (-1)^n p_n I = O, \tag{13}$$

and the consequent solution for I,

$$I = -\frac{1}{(-1)^n p_n}[A^n - p_1A^{n-1} + p_2A^{n-2} + \cdots + (-1)^{n-1}p_{n-1}A]. \tag{14}$$

* See C. C. MacDuffee, *Vectors and Matrices*, The Mathematical Association of America (1943), p. 79.

Multiplying both sides of equation (14) by A^{-1}, we see that the inverse of the matrix A can be computed by the formula

$$(15)\quad A^{-1} = -\frac{1}{(-1)^n p_n}[A^{n-1} - p_1 A^{n-2} + p_2 A^{n-3} + \cdots + (-1)^{n-1} p_{n-1} I].$$

This relation leads us to

THEOREM V. *Any negative integral power of the nonsingular square matrix of order n is linearly expressible in terms of the unit matrix and the first $(n-1)$ powers of A.*

To illustrate, for the numerical matrix A given by (5), we may obtain from (6)

$$I = \tfrac{1}{324}[A^3 - 27A^2 + 180A], \quad A^{-1} = \tfrac{1}{324}[A^2 - 27A + 180I],$$
$$A^{-2} = \tfrac{1}{324}[A - 27I + 180A^{-1}].$$

Using the second of these relations, A^{-2} may be expressed in terms of positive integral powers of A.

Clearly, if a matrix A is derogatory, its minimum function is a simpler medium than the characteristic function for expressing positive integral powers, A^{-1}, and negative integral power in terms of the first $(m-1)$ powers of A.

EXERCISES

Find the characteristic equation and the characteristic roots of each of the following matrices, and give in an explicit display form analogous to (8) or (12) the related product which equals the null matrix. Compute the inverse of the given matrix (if it exists) with the aid of its minimum function.

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

4. If

$$B = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix},$$

where λ_1, λ_2, λ_3, are the characteristic roots of the matrix $A = [a]_3^3$, prove that B also satisfies the characteristic equation of A.

8–5 The characteristic equation and rotation of coordinate axes. Consider the central conic whose equation is

$$(16)\qquad 5(x_1)^2 - 4x_1x_2 + 2(x_2)^2 = 6.$$

or

(17) $$(x_1, x_2)\begin{bmatrix} 5 & -2 \\ -2 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 6,$$

referred to OX_1 and OX_2 as axes. Recall from plane analytics that to eliminate the product term x_1x_2 from $a(x_1)^2 + bx_1x_2 + c(x_2)^2 = d$, we rotate the coordinate axes through an angle θ where $\tan 2\theta = b/(a - c)$. For equation (16), $\tan 2\theta = -\frac{4}{3}$; so $\sin \theta = 2/\sqrt{5}$, $\cos \theta = 1/\sqrt{5}$, and the desired rotation which is expressed generally by

(18) $$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}$$

in this case is

(19) $$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix}.$$

Substituting from (19) in (17), we get

$$(\bar{x}_1, \bar{x}_2) \cdot \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} 5 & -2 \\ -2 & 2 \end{bmatrix}\frac{1}{\sqrt{5}}\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = 6,$$

or

(20) $$(\bar{x}_1, \bar{x}_2)\begin{bmatrix} 1 & 0 \\ 0 & 6 \end{bmatrix}\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = 6,$$

or

(21) $$\frac{(\bar{x}_1)^2}{6} + \frac{(\bar{x}_2)^2}{1} = 1.$$

The effect of the rotation (19) on the equation (17) with symmetric coefficient matrix $A = \begin{bmatrix} 5 & -2 \\ -2 & 2 \end{bmatrix}$ is to transform the latter to a new form in which the symmetric coefficient matrix

$$B = \begin{bmatrix} 1 & 0 \\ 0 & 6 \end{bmatrix}$$

is a *diagonal matrix*, and this diagonal matrix B has its diagonal elements equal to the characteristic roots of the matrix A. Note that A and B are similar, for $B = R^{-1}AR$, where R is the orthogonal matrix of the rotation (19); so $|A| = |B|$, and $\operatorname{tr} A = \operatorname{tr} B$.

The process by which a given matrix is put in diagonal form is important, not only in pure but also in applied mathematics. In Chapter 11 we discuss some applications of this procedure.

Let R represent the orthogonal matrix of the general rotation of the plane, given by (18), and subject the central conic

$$(22) \qquad (x_1, x_2)\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 1 \quad \text{or} \quad \alpha' A \alpha = 1 \qquad (A = A')$$

to this rotation. Substituting from (18) in (22), we obtain

$$(23) \qquad (\bar{x}_1, \bar{x}_2)\begin{bmatrix} r_{11} & r_{21} \\ r_{12} & r_{22} \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix}\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = 1,$$

or

$$\bar{\alpha}' R^{-1} A R \bar{\alpha} = 1,$$

where $R = [r]_2^2$. The result is to transform the symmetric coefficient matrix A of (22) in the manner $B = R^{-1}AR$; so A and B are similar matrices. Therefore, under any rotation R, the symmetric matrix A and its transform B have the same characteristic equation and the same characteristic roots. The significant point is that in order to eliminate the product term from (22), R must be a specially chosen orthogonal matrix so that $B = R^{-1}AR$ is in diagonal form.

Exercises

1. Write each of the following equations in the form $\alpha' A \alpha = k$, comparable to (22), where A is a symmetric matrix. Then subject it to the appropriate rotation of the type (18) to eliminate the xy term of the given equation, that is, diagonalize the coefficient matrix of the given form. In each problem find the characteristic roots of the symmetric coefficient matrix in the given equation, and also those of the diagonal matrix in the transformed equation, and verify that they are the same.

(*i*) $25x^2 + 14xy + 25y^2 = 288$. (*ii*) $x^2 - 24xy - 6y^2 = 1$.
(*iii*) $xy = 12$. (*iv*) $x^2 - 2xy + y^2 = 12$.

2. Verify that the equation $3x^2 - 8xy + 3y^2 - 5z^2 = -5$ may be written in the form

$$(x, y, z)\begin{bmatrix} 3 & -4 & 0 \\ -4 & 3 & 0 \\ 0 & 0 & -5 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = -5.$$

Effect the rotation

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & \sqrt{2} \end{bmatrix}\begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{bmatrix}$$

on the given equation. Is the coefficient matrix of the transformed equation diagonal? How do its characteristic roots compare with those of the coefficient matrix of the given equation? What kind of surface does the given equation represent?

3. Write the equation of the surface $103x^2 + 125y^2 + 66z^2 - 48xy - 12xz - 60yz = 294$ in the form $\alpha' A \alpha = 294$, where $\alpha = \{x, y, z\}$ and A is a symmetric matrix. Subject the coordinate axes to the rotation

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{1}{7}\begin{bmatrix} 2 & 6 & -3 \\ 3 & 2 & 6 \\ 6 & -3 & -2 \end{bmatrix}\begin{bmatrix} \bar{x} \\ \bar{y} \\ \bar{z} \end{bmatrix},$$

using matrix algebra, and determine the equation into which $\alpha' A \alpha = 294$ is transformed. Compare the characteristic roots of A and those of the new coefficient matrix.

8–6 The characteristic vectors of a matrix. Let $A = [a]_n^n$ be a square matrix of order n with the n *distinct* characteristic roots $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$. We know that $[A - \lambda_h I]$, $h = 1, 2, \ldots, n$, is a singular matrix, since each of the characteristic roots λ_h has been found subject to the condition that $|A - \lambda_h I| = 0$. Since $[A - \lambda_h I]$ is singular, we can find (Appendix 2) for each characteristic root λ_h (*h fixed now*) a column vector

$$\beta_h = \{b_{ih}\} = \{b_{1h}, b_{2h}, \cdots, b_{nh}\}$$

such that

$$(24) \qquad [A - \lambda_h I]\{b_{ih}\} = O$$

or

$$(25) \qquad A\{b_{ih}\} = \lambda_h\{b_{ih}\}.$$

It is essential that we clearly realize that there is an equation of the type (24), and equivalently (25), for each characteristic root λ_h; thus

$$(26) \quad \begin{cases} \text{for } h = 1, & [A - \lambda_1 I]\{b_{i1}\} = O, & \text{or} & A\{b_{i1}\} = \lambda_1\{b_{i1}\}; \\ \text{for } h = 2, & [A - \lambda_2 I]\{b_{i2}\} = O, & \text{or} & A\{b_{i2}\} = \lambda_2\{b_{i2}\}; \\ \text{for } h = 3, & [A - \lambda_3 I]\{b_{i3}\} = O, & \text{or} & A\{b_{i3}\} = \lambda_3\{b_{i3}\}; \\ \cdot & \cdot & \cdot & \cdot \\ \text{for } h = n, & [A - \lambda_n I]\{b_{in}\} = O, & \text{or} & A\{b_{in}\} = \lambda_n\{b_{in}\}. \end{cases}$$

The n vectors $\{b_{ih}\}$, $h = 1, 2, \ldots, n$, are linearly independent. For if they were linearly dependent we could find scalars $k_1, k_2, \ldots, k_n$ not all zero such that

$$(27) \qquad k_1\{b_{i1}\} + k_2\{b_{i2}\} + k_3\{b_{i3}\} + \cdots + k_n\{b_{in}\} = O.$$

Now suppose that k's did exist so that (27) were true; premultiply both sides of (27) by $(A - \lambda_2 I) \cdot (A - \lambda_3 I)(A - \lambda_4 I) \ldots (A - \lambda_n I)$, and we would have

$$(28) \qquad (A - \lambda_2 I)(A - \lambda_3 I) \ldots (A - \lambda_n I) k_h \{b_{ih}\} = O,$$

where here *h is summed from* 1 *to n*. But in virtue of the last $(n - 1)$ equations of the set (26), the relation (28) becomes

$$(29) \qquad (A - \lambda_2 I)(A - \lambda_3 I) \ldots (A - \lambda_n I) k_1 \{b_{i1}\} = O$$

Thus for the numerical matrix A given above, we might have well chosen for the modal matrix

$$\bar{B} = \begin{bmatrix} m & 0 & p \\ 2m & n & p \\ 2m & n & 2p \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 2 & 1 & 2 \end{bmatrix} \begin{bmatrix} m & 0 & 0 \\ 0 & n & 0 \\ 0 & 0 & p \end{bmatrix} = [b_{ij}] \begin{bmatrix} m & 0 & 0 \\ 0 & n & 0 \\ 0 & 0 & p \end{bmatrix}.$$

But the transform of A by $\bar{B}$ is the same as the transform of A by B; this we leave for the reader to verify.

Exercises

For each of the following matrices A find the characteristic vectors and construct the modal matrix B, such that $B^{-1}AB$ is a diagonal matrix. Lastly, find this diagonal form of A as we have indicated above.

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

8–7 The characteristic vectors of a symmetric matrix. Diagonalization of symmetric matrices. The symmetric matrix

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

has

$$\text{characteristic root } \lambda_1 = 1$$

with corresponding

$$\text{characteristic vector } \{b_{11}, b_{21}\} = \{1, 2\},$$

and

$$\text{characteristic root } \lambda_2 = 6$$

with corresponding

$$\text{characteristic vector } \{b_{12}, b_{22}\} = \{-2, 1\}.$$

From these values we have

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

and

$$B^{-1}AB = \frac{1}{5}\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}\begin{bmatrix} 5 & -2 \\ -2 & 2 \end{bmatrix}\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 6 \end{bmatrix}.$$

Note that the characteristic vectors of the symmetric matrix A are orthogonal. That is no coincidence; in more advanced books on matrix algebra it is shown that in general *any two characteristic vectors* *of a real symmetric matrix belonging to different characteristic roots of that matrix are orthogonal.** By normalizing each vector of the modal matrix B belonging to the symmetric matrix A we construct an orthogonal modal matrix for A. If the characteristic roots of the symmetric matrix A are distinct, and if their order is specified, the orthogonal modal matrix B belonging to the symmetric matrix A is unique except that the signs of the elements of any column may be reversed; if the characteristic roots of A are not distinct, then B is not unique. For the numerical symmetric matrix A given at the beginning of this section, $1/\sqrt{5}$ is a normalizing factor for each of the characteristic vectors of A, and the unique orthogonal modal matrix for A (unique except for the choice of the signs of the elements of each column) is $B = \dfrac{1}{\sqrt{5}}\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}$.

Symmetric matrices occur quite often in applied problems which lead to multivariate algebra, and usually in such problems it is important to know how to diagonalize such matrices. To make the procedure clear, we consider two more illustrations.

Associated with the quadric surface

$$(x, y, z)\begin{bmatrix} 11 & -6 & 2 \\ -6 & 10 & -4 \\ 2 & -4 & 6 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = 18$$

is the symmetric coefficient matrix

$$A = \begin{bmatrix} 11 & -6 & 2 \\ -6 & 10 & -4 \\ 2 & -4 & 6 \end{bmatrix}.$$

The process of subjecting the symmetric coefficient matrix A to a similarity transformation so that the transform of A is diagonal will effect a transformation of the surface to its principal axes. The matrix A has 3, 6, and 18 for characteristic roots and $\frac{1}{3}\{1, 2, 2\}$, $\frac{1}{3}\{2, 1, -2\}$, and $\frac{1}{3}\{2, -2, 1\}$ as corresponding orthogonal normalized characteristic vectors. Constructing the orthogonal modal matrix B with these characteristic vectors as column vectors, we find that

---

* See C. C. MacDuffee, *Vectors and Matrices*, p. 170; R. A. Frazer, W. J. Duncan, and A. R. Collar, *Elementary Matrices and Some Applications to Dynamics and Differential Equations*, p. 77.

$B^{-1}AB$ is the diagonal matrix with 3, 6, and 18 as diagonal elements. Therefore, under a rotation with B as transformation matrix the given quadric surface is transformed into $3\bar{x}^2 + 6\bar{y}^2 + 18\bar{z}^2 = 18$.

If two of the characteristic roots of the symmetric coefficient matrix A associated with a quadric surface are equal, the orthogonal matrix for transforming A into diagonal form is not unique but can be constructed.

Consider the quadric surface

$$2x^2 + 2y^2 + 2z^2 - 2xy + 2xz - 2yz = 4,$$

or

$$(x, y, z)\begin{bmatrix} 2 & -1 & 1 \\ -1 & 2 & -1 \\ 1 & -1 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = 4, \qquad \text{or} \qquad \alpha' A\alpha = 4.$$

The coefficient matrix A has characteristic roots $\lambda_1 = 1$, $\lambda_2 = 1$, and $\lambda_3 = 4$. Corresponding to the characteristic root $\lambda_3 = 4$ there is the normalized characteristic vector $\{b_{13}, b_{23}, b_{33}\} = 1/\sqrt{3}\,\{1, -1, 1\}$. Corresponding to each of the double characteristic roots $\lambda_1 = \lambda_2 = 1$, we have the equations

$$[A - I]\{b_{i1}\} = O, \qquad \text{or} \qquad \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix}\begin{bmatrix} b_{11} \\ b_{21} \\ b_{31} \end{bmatrix} = O,$$

or

$$\begin{cases} b_{11} - b_{21} + b_{31} = 0, \\ -b_{11} + b_{21} - b_{31} = 0, \\ b_{11} - b_{21} + b_{31} = 0. \end{cases}$$

These three scalar equations in the b_{i1}'s are equivalent; otherwise stated, the rows of the matrix $[A - I]$ are linearly connected by two independent relations, that is, $[A - I]$ is of nullity 2. We may choose for $\{b_{11}, b_{21}, b_{31}\}$ and $\{b_{12}, b_{22}, b_{32}\}$ any two vectors subject to the conditions that they are orthogonal and that they satisfy the above equations. We choose

$$\{b_{11}, b_{21}, b_{31}\} = \frac{1}{\sqrt{2}}\{0, 1, 1\}$$

and

$$\{b_{12}, b_{22}, b_{32}\} = \frac{1}{\sqrt{6}}\{2, 1, -1\};$$

in making these choices we select two normalized vectors which are orthogonal to each other and each orthogonal to the characteristic vector $\{b_{13}, b_{23}, b_{33}\} = 1/\sqrt{3}\,\{1, -1, 1\}$ determined above.

Then

$$B = \begin{bmatrix} 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}, \quad B^{-1} = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix},$$

and

$$B^{-1}AB = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

Under a rotation of axes with B as transformation matrix the given quadric form becomes

$$\bar{x}^2 + \bar{y}^2 + 4\bar{z}^2 = 4.$$

Note that in this case the orthogonal matrix B is not uniquely determined, and the quadric is a surface of revolution.

EXERCISES

Write each of the following equations in the form $\alpha' A \alpha = k$, where $\alpha = \{x, y, z\}$, A is a numerical symmetric matrix, and k is a scalar. Using the method of this section, find an orthogonal matrix B such that $B^{-1}AB$ is diagonal, and thereby transform the surface to its principal axes.

1. $7x^2 - 8y^2 - 8z^2 + 8xy - 8xz - 2yz = -9.$
2. $2xy + 2xz + 2yz = 4.$
3. For the symmetric matrix

$$A = \begin{bmatrix} 9 & 3 & -7 & 9 \\ 3 & -6 & 0 & 6 \\ -7 & 0 & -8 & -2 \\ 9 & 6 & -2 & 0 \end{bmatrix},$$

find an orthogonal matrix B such that $B^{-1}AB$ is a diagonal matrix.

CHAPTER 9

RANK OF A MATRIX

9–1 Relations connecting the rows and columns of a singular matrix. It is instructive to compare certain algebraic situations in which a nonsingular matrix, say

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

occurs, and like situations in which singular matrices occur. Note that the vector $\alpha = \{0, 0, 0\}$, where $\alpha = \{x_1, x_2, x_3\}$, is the only solution of the matrix equation

$$A\alpha = O, \qquad \text{or} \qquad \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = O,$$

or of the scalar equations

$$\begin{cases} x_1 + 2x_2 + 2x_3 = 0, \\ 2x_1 + x_2 - 2x_3 = 0, \\ 2x_1 - 2x_2 + x_3 = 0. \end{cases}$$

Otherwise stated, the only relation connecting the rows and the columns of A are

for the rows*

$$0(1, 2, 2) + 0(2, 1, -2) + 0(2, -2, 1) = O,$$

or

$$0(a_{1i}) + 0(a_{2i}) + 0(a_{3i}) = O;$$

for the columns

$$0\{1, 2, 2\} + 0\{2, 1, -2\} + 0\{2, -2, 1\} = O,$$

or

$$0\{a_{i1}\} + 0\{a_{i2}\} + 0\{a_{i3}\} = O.$$

---

* Here, and elsewhere in this chapter, we find it convenient to represent the rth row vector α'_r of the matrix $A = [a]_n^m$ by

$$\alpha'_r = (a_{ri}) = (a_{r1}, a_{r2}, \ldots, a_{rn}).$$

and the sth column vector α_s of $A = [a]_n^m$ by

$$\alpha_s = \{a_{is}\} = \{a_{1s}, a_{2s}, \ldots, a_{ms}\}.$$

Recall from Section 4-5 that the vectors $\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_m$ are said to be linearly dependent only if we can find scalars $k_1, k_2, k_3, \ldots, k_m$, not all zero, such that

$$k_1\alpha_1 + k_2\alpha_2 + k_3\alpha_3 + \cdots + k_m\alpha_m = O.$$

So the vectors of the nonsingular matrix A are linearly independent.

Of a different nature is the situation for the singular matrix

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

A single relation connects the rows (or the columns) of B. For rows the relation is

$$1(0, -1, 1) + 1(4, -1, -1) - 1(4, -2, 0) = O,$$

or

$$(b_{1i}) + (b_{2i}) - (b_{3i}) = O.$$

When we speak of a single linear relation connecting the rows of B, we mean that any other linear relation connecting the rows is obtainable from this one by multiplying it through by a constant scalar. For example, if we multiply the above relation through by 2 we get $2(b_{1i}) + 2(b_{2i}) - 2(b_{3i}) = O$, which is not independent of the first relation. For columns the relation is

$$1\{0, 4, 4\} + 2\{-1, -1, -2\} + 2\{1, -1, 0\} = O,$$

or

$$1\{b_{i1}\} + 2\{b_{i2}\} + 2\{b_{i3}\} = O.$$

From the last relation we see that the equation $B\alpha = O$ has the particular solution $\alpha = \{1, 2, 2\}$ and the general solution $k\alpha = \{k, 2k, 2k\}$, where k is an arbitrary scalar constant. The rows and the columns of the singular matrix B are linearly dependent.

Further, consider the matrix

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

Observe that the equation

$$C\alpha = O, \qquad \text{or} \qquad \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 2 & -2 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = O,$$

or the scalar equations

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

can be satisfied by the two particular solutions $\alpha_1 = \{0, 1, 1\}$ and $\alpha_2 = \{2, 1, -1\}$; the most general solution is $\alpha = k_1\alpha_1 + k_2\alpha_2$, where k_1 and k_2 are arbitrary scalars. In this case the rows (or the columns) of C are connected by two linearly independent relations. For the rows $(c_{2i}) = -(c_{1i})$ and $(c_{3i}) = 2(c_{1i})$; for the columns $\{c_{i2}\} = -\{c_{i1}\}$ and $\{c_{i3}\} = \{c_{i1}\}$. This matrix C has only one linearly independent row and only one linearly independent column; we say the rows and the columns are proportional. Such a matrix is expressible as the product of a column vector and a row vector; here $C = \{1, -1, 2\}(1, -1, 1)$.

Our work up to this point has been almost entirely with square matrices or matrices of order n by n, and with row and column vectors or matrices of order n by 1 and 1 by n. In this chapter we shall deal considerably with *rectangular matrices* or matrices of order m by n, as

$$F = \begin{bmatrix} 3 & 2 & -1 & 5 \\ 5 & 1 & 4 & -2 \\ 1 & -4 & 11 & -19 \end{bmatrix}.$$

Just as for a square matrix, the rows (and columns) of a rectangular matrix may be linearly independent or linearly dependent. In particular, the rows of the numerical matrix F just given are linearly dependent, for

$$3(f_{1i}) = 2(f_{2i}) - (f_{3i}).$$

EXERCISES

Determine whether the rows of each of the following matrices are connected by none, one, or two linearly independent relations. If such relations exist, give them.

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

9–2 Submatrices. Minors of a matrix. The rectangular matrix F of the preceding section has no unique determinant associated with it as a function of all its scalar elements, as does a square matrix. However, by striking from F certain rows and columns, we may associate with F a number of other matrices, called *submatrices* of F, some of which may be square and have determinants. By striking from F in turn the first, second, third, and fourth columns, we get four submatrices; note that the determinant of each of these square submatrices of the third order is zero. If we strike from F the third row and the last two columns, we get the square submatrix

$$\begin{bmatrix} 3 & 2 \\ 5 & 1 \end{bmatrix},$$

with its determinant equal to -7.

Consider the matrix A with m rows and n columns (or of order m by n):

$$(1) \qquad A = [a]_n^m = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix}.$$

Construct the matrix B of order s by t, by selecting s rows and t columns from A, where $s \leq m$ and $t \leq n$; we call B an s by t *submatrix* of A. If $s < m$ and $t < n$, the elements in the remaining $m - s$ rows and $n - t$ columns form an $(m - s)$ by $(n - t)$ submatrix C of A; we call B and C *complementary submatrices*. Related to the matrix $A = [a]_5^4$ are the complementary submatrices

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} a_{33} & a_{34} & a_{35} \\ a_{43} & a_{44} & a_{45} \end{bmatrix};$$

likewise

$$[a_{11}] \quad \text{and} \quad \begin{bmatrix} a_{22} & a_{23} & a_{24} & a_{25} \\ a_{32} & a_{33} & a_{34} & a_{35} \\ a_{42} & a_{43} & a_{44} & a_{45} \end{bmatrix}$$

are complementary matrices. In particular, every element of a matrix A is a one-rowed and one-columned square submatrix of A.

In some circumstances it will be desirable to regard a given matrix as being made up of certain of its submatrices, particularly when A has a number of zero or unit elements symmetrically placed, or has some other special property. Thus, using the indicial symbolism of the preceding paragraph, we may write

$$(2) \qquad A = [A_{ij}] \qquad (i = 1, \dots, s; j = 1, \dots, t),$$

where it is now understood that the symbols A_{ij} themselves represent rectangular matrices. We assume that for any fixed i the matrices $A_{i1}, A_{i2}, \dots, A_{it}$ all have the same number of rows, and for any fixed k the matrices $A_{1k}, A_{2k}, \dots, A_{sk}$ have the same number of columns. By this scheme we accomplish a *partitioning* of A by what amounts to drawing lines parallel to the rows and columns of A and between them, and representing the submatrices so formed by A_{ij}. Our main use of such partitioning will be where we shall regard A as a 2 by 2 matrix

$$(3) \qquad A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

whose elements A_{11}, A_{12}, A_{21}, A_{22} are themselves rectangular matrices.

Suppose two matrices of the same order to be partitioned in a corresponding way. Then the submatrices occupying corresponding positions will be of the same order, and may be added in the regular manner. Thus

$$\left[\begin{array}{cc|c} 3 & 0 & 5 \\ 0 & 3 & 1 \\ \hline 1 & 2 & 4 \end{array}\right] + \left[\begin{array}{cc|c} 4 & 0 & -2 \\ 0 & 4 & 1 \\ \hline 2 & 3 & -1 \end{array}\right]$$

$$= \left[\begin{array}{cc|c} \multicolumn{2}{c|}{3I_2} & \begin{array}{c} 5 \\ 1 \end{array} \\ \hline 1 & 2 & 4 \end{array}\right] + \left[\begin{array}{cc|c} \multicolumn{2}{c|}{4I_2} & \begin{array}{c} -2 \\ 1 \end{array} \\ \hline 2 & 3 & -1 \end{array}\right] = \left[\begin{array}{cc|c} \multicolumn{2}{c|}{7I_2} & \begin{array}{c} 3 \\ 2 \end{array} \\ \hline 3 & 5 & 3 \end{array}\right].$$

Concerning the multiplication of two partitioned matrices, for every partitioning line between columns of the matrix on the left there must be a partitioning line between the corresponding rows of the matrix on the right; if this condition is satisfied, the two matrices are said to be *conformably partitioned*. To illustrate, let

$$A = [a]_5^5 = \left[\begin{array}{cc|ccc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{array}\right] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

and

$$B = [b]_5^5 = \left[\begin{array}{ccc|cc} b_{11} & b_{12} & b_{13} & b_{14} & b_{15} \\ b_{21} & b_{22} & b_{23} & b_{24} & b_{25} \\ \hline b_{31} & b_{32} & b_{33} & b_{34} & b_{35} \\ b_{41} & b_{42} & b_{43} & b_{44} & b_{45} \\ b_{51} & b_{52} & b_{53} & b_{54} & b_{55} \end{array}\right] = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}.$$

Then

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

where $C = AB$.

In general, let A and B be two conformably partitioned matrices with their submatrices denoted respectively by A_{ij} and B_{ij}, where i denotes the order of occurrence in row groups and j such order in column groups. Also let the product $C = AB$ have the same row partitioning as A and the same column partitioning as B. If C_{ij} is the (i, j)th submatrix of C, then it can be proved that

$$C_{ij} = A_{ik}B_{kj}. \tag{4}$$

In this respect the multiplication of submatrices is like the ordinary multiplication of matrices. In fact, the ordinary multiplication of two matrices A and B may be considered as the multiplication of certain submatrices of A, the rows of A, and certain submatrices of B, the columns of B. As another illustration, we give

$$\left[\begin{array}{rr|rr} 1 & -2 & 0 & 0 \\ -2 & 5 & 0 & 0 \\ \hline 0 & 0 & 2 & -5 \\ 0 & 0 & -3 & 8 \end{array}\right]\left[\begin{array}{rr|rr} 5 & 2 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ \hline 0 & 0 & 8 & 5 \\ 0 & 0 & 3 & 2 \end{array}\right] = \begin{bmatrix} \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix}\begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix} & O_2 \\ O_2 & \begin{bmatrix} 2 & -5 \\ -3 & 8 \end{bmatrix}\begin{bmatrix} 8 & 5 \\ 3 & 2 \end{bmatrix} \end{bmatrix} = I_4.$$

Let B be a square submatrix of A, of order k; the determinant of B we call a *k-rowed minor* of A. Thus

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad \begin{vmatrix} a_{12} & a_{13} & a_{15} \\ a_{22} & a_{23} & a_{25} \\ a_{42} & a_{43} & a_{45} \end{vmatrix}$$

are a 2-rowed minor and a 3-rowed minor, respectively, of $A = [a]_5^5$.

What we have been calling the minor (Section 6–3) of a designated element of a square matrix A of order n is more generally termed an $(n - 1)$-rowed minor. The elements of a matrix A constitute 1-rowed minors. Also, for a square matrix A of order n there is only one n-rowed determinant associated with A, and this is what we have been calling *the determinant* of the square matrix A.

If M_1 and M_2 are complementary square submatrices of the matrix A, then $|M_1|$ and $|M_2|$ are complementary minors of A; either is the complement of the other. The *algebraic complement* of a minor $|M_1|$ of A is equal to its complementary minor $|M_2|$ multiplied by that power of -1 whose exponent is equal to the sum of the indices of the rows and columns of A used in the formation of the minor $|M_1|$. To relate these general ideas with those of Sections 6–3 and 8–2, we note the following. The one-rowed principal minors of a square matrix are the elements of its principal diagonal; the complement of a single element is its minor; the algebraic complement of a single element is its cofactor. For the matrix $A = [a]_4^4$ the algebraic complements of

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \quad \text{and} \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix}$$

are, respectively,

$$\begin{vmatrix} a_{33} & a_{34} \\ a_{43} & a_{44} \end{vmatrix} \quad \text{and} \quad (-1)^{1+3+1+2} \begin{vmatrix} a_{23} & a_{24} \\ a_{43} & a_{44} \end{vmatrix}.$$

Exercises

1. Write out all the square submatrices of the matrix A and their algebraic complements, where

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{bmatrix}.$$

2. For the matrix A of Ex. 1, find the algebraic sum of the products obtained by multiplying each of the 2-rowed minors of A that can be formed from the first two columns of A and their respective algebraic complements. Show that the sum so obtained is equal to $|A|$.

3. For the matrix A of Ex. 1, find the algebraic sum of the products obtained by multiplying each of the 2-rowed minors of A that can be formed from the first and third rows of A and their respective algebraic complements. Show that the sum so obtained is equal to $|A|$.

4. Show that if

$$A = \left[\begin{array}{cc|cc} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ \hline a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{array}\right] = \begin{bmatrix} A_{11} & O_2 \\ A_{21} & A_{22} \end{bmatrix},$$

then $|A| = |A_{11}| \cdot |A_{22}|$.

5. If

$$C = \begin{bmatrix} a_{11} & a_{12} & a_{13} & x_1 \\ a_{21} & a_{22} & a_{23} & x_2 \\ a_{31} & a_{32} & a_{33} & x_3 \\ u_1 & u_2 & u_3 & 0 \end{bmatrix} = \left[\begin{array}{c|c} A & \alpha \\ \hline \beta' & 0 \end{array}\right],$$

prove that

$$C^2 = \left[\begin{array}{c|c} A^2 + \alpha\beta' & A\alpha \\ \hline \beta' A & \beta'\alpha \end{array}\right].$$

What is the order of $\alpha\beta'$? Of $\beta'\alpha$?

9-3 Laplace's development of a determinant. In Ex. 2 of the preceding section we saw that for a particular numerical matrix the algebraic sum of the products obtained by multiplying each of the 2-rowed minors of A that can be formed from the first two columns of A and their respective algebraic complements is equal to $|A|$. Verify that the same is true for the general square matrix of order 4; that is, show that

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \cdot \begin{vmatrix} a_{33} & a_{34} \\ a_{43} & a_{44} \end{vmatrix} - \begin{vmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{vmatrix} \cdot \begin{vmatrix} a_{23} & a_{24} \\ a_{43} & a_{44} \end{vmatrix}$$
$$+ \begin{vmatrix} a_{11} & a_{12} \\ a_{41} & a_{42} \end{vmatrix} \cdot \begin{vmatrix} a_{23} & a_{24} \\ a_{33} & a_{34} \end{vmatrix} + \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \cdot \begin{vmatrix} a_{13} & a_{14} \\ a_{43} & a_{44} \end{vmatrix}$$
$$- \begin{vmatrix} a_{21} & a_{22} \\ a_{41} & a_{42} \end{vmatrix} \cdot \begin{vmatrix} a_{13} & a_{14} \\ a_{33} & a_{34} \end{vmatrix} + \begin{vmatrix} a_{31} & a_{32} \\ a_{41} & a_{42} \end{vmatrix} \cdot \begin{vmatrix} a_{13} & a_{14} \\ a_{23} & a_{24} \end{vmatrix}.$$

This expansion is a particular instance of an interesting theorem due to Laplace:

LAPLACE'S THEOREM. *The value of a determinant of a matrix A is equal to the algebraic sum of the products obtained by multiplying each of the k-rowed minors that can be formed from any k rows (or k columns) of A by their algebraic complements.*

This theorem dates from 1772, and is given in Laplace's *Œuvres, VIII*, pp. 365–406. Proofs are commonly given in complete treatises on determinants; in particular see T. Muir and W. H. Metzler, *A Treatise on the Theory of Determinants*, Longmans (1933), pp. 80–88.

The number of different k-rowed minors that can be formed from k selected rows (or k selected columns) is the number of sets of k things that can be formed out of n (n being the order of the given square matrix A), and therefore is equal to

$$_nC_k = \frac{n!}{k!(n-k)!}.$$

For $n = 4$ and $k = 2$ the number of Laplace expansions of $|A|$ using two columns (or using two rows) is 6.

For any matrix A, the Laplace expansion of $|A|$ by 1-rowed minors is the commonplace and usual expansion of $|A|$ by some one row or some one column.

Exercises

1. For $A = [a]_4^4$, write out the six Laplace expansions of $|A|$ by 2-rowed minors.

2. Give the Laplace expansion of $|A|$ for $A = [a]_5^5$ by 3-rowed minors from the first three rows of A.

3. Using a Laplace expansion, show that

$$\begin{vmatrix} a & b & c & d \\ e & f & g & h \\ 0 & 0 & j & k \\ 0 & 0 & l & m \end{vmatrix} = \begin{vmatrix} a & b \\ e & f \end{vmatrix} \cdot \begin{vmatrix} j & k \\ l & m \end{vmatrix}.$$

4. Using 2-rowed minors from the first two rows, show that

$$\frac{1}{2}\begin{vmatrix} a & b & c & d \\ e & f & g & h \\ a & b & c & d \\ e & f & g & h \end{vmatrix} = \begin{vmatrix} a & b \\ e & f \end{vmatrix} \cdot \begin{vmatrix} c & d \\ g & h \end{vmatrix} - \begin{vmatrix} a & c \\ e & g \end{vmatrix} \cdot \begin{vmatrix} b & d \\ f & h \end{vmatrix} + \begin{vmatrix} a & d \\ e & h \end{vmatrix} \cdot \begin{vmatrix} b & c \\ f & g \end{vmatrix} = 0.$$

5. By means of a Laplace expansion using 2-rowed minors from the first two columns, show that

$$\begin{vmatrix} 0 & a & b & c \\ -a & 0 & d & e \\ -b & -d & 0 & f \\ -c & -e & -f & 0 \end{vmatrix} = (af - be + cd)^2.$$

9–4 The concept of rank. We have seen that a matrix A with m rows and n columns (of order m by n) has *determinants* or *k-rowed minors* of all orders from 1 (the elements of A themselves) to the smaller of the two integers m and n inclusive. It is often of importance to specify the order of the highest order nonvanishing determinant or k-rowed minor of a given matrix.

A matrix has *rank* r if and only if it has at least one r-rowed minor which is not zero while all minors of order greater than r are zero. That is, the *rank* r of a matrix of order m by n is the order of a nonsingular minor of A of maximum order.

If A is a matrix of order m by n, then its rank r satisfies the inequalities $r \leq n$, $r \leq m$. It follows from the definition of rank that if the rank of A is r, then every minor of A of order greater than r will be zero. If the rank of a matrix is zero, all of its elements are zero.

If A is a nonsingular square matrix of order n, then rank of $A = r = n$. If A is a singular square matrix, then $r < n$. The quantity $n - r$ is called the *nullity* of the square matrix A. If A is a rectangular matrix of order m by n there are two nullities, a *row-nullity* $m - r$ and a *column nullity* $n - r$.

We give the following illustrations of rank and nullity. Of the matrices in Section 9–1, A has rank 3 and nullity 0; B has rank 2 and nullity $3 - 2 = 1$; C has rank 1 and nullity $3 - 1 = 2$; F has rank 2, row-nullity $3 - 2 = 1$, and column-nullity $4 - 2 = 2$. A row or column vector with at least one nonzero element has rank 1; the unit matrix of order n has rank n and nullity 0; the matrix of order m by n in which every element is unity has rank 1, row-nullity $m - 1$, and column-nullity $n - 1$.

Exercises

Give the rank and nullity (or nullities) of each of the following matrices.

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

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

9–5 Rank and linear dependence. The illustrations of the preceding section, with the work of Section 9–1, point up the following facts pertaining to the numerical matrices of Section 9–1:

(*i*) The square matrix A is of order 3 and rank 3; its three rows and three columns are linearly independent.

(*ii*) The square matrix B is of order 3 and rank 2; it has two linearly independent rows and two linearly independent columns.

(*iii*) The square matrix C is of order 3 and rank 1; it has only one linearly independent row and one linearly independent column.

(*iv*) The rectangular matrix F is of order 3 by 4 and has rank 2; it has 2 linearly independent rows and 2 linearly independent columns. We have previously given the one linear combination of its three rows: $3(f_{1i}) = 2(f_{2i}) - (f_{3i})$. Of the four columns of F only two are linearly independent; two independent linear combinations connecting the columns of F are

$$9\{f_{i1}\} - 17\{f_{i2}\} - 7\{f_{i3}\} = O \qquad \text{and} \qquad -2\{f_{i2}\} + \{f_{i3}\} + \{f_{i4}\} = O.$$

The maximum number of linearly independent rows in a rectangular matrix B is called its *row rank*, and the maximum number of linearly independent columns in B is called its *column rank*. To contrast these with the rank of the preceding section, which we related to an r-rowed determinant of B, we call the latter the *determinant rank* of B. The above examples furnish illustrations of the equality of row rank, column rank, and determinant rank of the given matrices. That these ranks are in general equal we prove below in Theorem X. However, before proceeding with this theorem we relate certain aspects of linear dependence of vectors with matrix algebra.

Recall from Section 4–5 that the condition for the vectors $\alpha_1, \alpha_2, \ldots, \alpha_m$ to be linearly dependent is that m scalars $s_1, s_2, \ldots, s_m$ of the underlying field, not all zero, exist such that

$$(5) \qquad s_1\alpha_1 + s_2\alpha_2 + \cdots + s_m\alpha_m = O.$$

If the α's are n-dimensional vectors, the vector equation (5) is equivalent to the n scalar equations

$$(6) \qquad \begin{cases} s_1a_{11} + s_2a_{21} + \cdots + s_ma_{m1} = 0, \\ s_1a_{12} + s_2a_{22} + \cdots + s_ma_{m2} = 0, \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ s_1a_{1n} + s_2a_{2n} + \cdots + s_ma_{mn} = 0. \end{cases}$$

These scalar equations are equivalent to the matrix equation

$$(7) \qquad \begin{bmatrix} a_{11} & a_{21} & \ldots & a_{m1} \\ a_{12} & a_{22} & \ldots & a_{m2} \\ \cdot & \cdot & \cdot & \cdot \\ a_{1n} & a_{2n} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \cdot \\ s_m \end{bmatrix} = O, \qquad \text{or} \qquad A\gamma = O,$$

where A is the matrix of which the vectors $\alpha_i = \{a_{i1}, a_{i2}, \ldots, a_{in}\}$, $i = 1, 2, \ldots, m$, are the columns, and the vector $\gamma = \{s_1, s_2, \ldots, s_m\}$ is not the null vector. These observations lead us to

THEOREM I. *If certain column vectors $\alpha_i = \{a_{i1}, a_{i2}, \ldots, a_{in}\}$, $i = 1, 2, \ldots, m$, are linearly dependent, then the homogeneous scalar equations (6) represented by $A\gamma = O$, where the column vectors α_i are the columns of the matrix A, have a nontrivial solution $\gamma \neq O$.*

Of interest to us next is

THEOREM II. *If the columns of a square matrix A are linearly dependent then $|A| = 0$.*

Suppose the condition for linear dependence is given by (6), with $s_i \neq 0$. Then we can solve equations (6) for $a_{i1}, a_{i2}, \ldots, a_{in}$ and obtain the ith column of A as a linear combination of the other columns; this linear combination will furnish an operation on the ith columns of A which will replace all elements of that column by zeros without altering the value of A; hence $|A| = 0$. To illustrate, suppose that the columns of the square matrix A of order 3 satisfy the linear relation $\text{col}_1 + 3\,\text{col}_2 - 2\,\text{col}_3 = O$, then the operation of replacing col_1 by $\text{col}_1 + 3\,\text{col}_2 - 2\,\text{col}_3$ applied to A will give all zero elements in the first row of $|A|$; so $|A| = 0$.

By a similar argument, we can establish

THEOREM III. *If the rows of a square matrix A are linearly dependent, then $|A| = 0$.*

THEOREM IV. *If m and n are positive integers such that $m \leq n$, and if m n-dimensional column vectors $\alpha_i = \{a_{i1}, a_{i2}, \ldots, a_{in}\}$, $i = 1, 2, \ldots, m$, are linearly dependent, then $r < m$, r being the rank of the matrix A with the m vectors α_i as columns.*

To prove Theorem IV we note that the condition for such linear dependence may be represented by (5), (6), or (7). It seems desirable that we fix our attention on (6). Suppose that $r = m$. Then by a rearrangement, if necessary, of the equations (6), or equivalently of the rows of A, we can partition the rearranged matrix (call it B) so that its submatrix B_1, with the first m rows of B, is nonsingular, with the remaining $n - m$ rows of B constituting another submatrix B_2. From the designated rearranged pattern of (6) we have, relative to the specified partition of B,

$$\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \cdot \\ \cdot \\ \cdot \\ s_m \end{bmatrix} = O, \quad \text{so that} \qquad B_1\gamma = O \qquad \text{and} \qquad B_2\gamma = O.$$

Under the supposition that $r = m$, the square matrix B_1 of order m by m (or r by r) is nonsingular, and so B_1^{-1} exists. Premultiplying $B_1\gamma = O$ by B_1^{-1}, we get

$$B_1^{-1}B_1\gamma = O, \qquad \text{or} \qquad I\gamma = O,$$

where I is the unit matrix of order m by m. The relation $I\gamma = O$, or

$$\begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & 1 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \cdot \\ s_m \end{bmatrix} = O \quad \text{requires that} \quad \gamma = \{s_1, s_2, \ldots, s_m\} = O,$$

which is contrary to the hypothesis that the rows of A be linearly dependent. Hence $r < m$, and the theorem is proved.

Exercises

1. If $\alpha_1 = \{2, 2, 0\}$ and $\alpha_2 = \{0, 2, -1\}$, find the vector α_3 such that $\alpha_3 = 2\alpha_1 - \alpha_2$. Find the rank of the matrix with α_1, α_2, and α_3 as columns, and thus verify Theorem IV.

2. If $\alpha_1 = \{3, 2, -1, 5\}$ and $\alpha_2 = \{5, 1, 4, -2\}$, find the vector α_3 such that $3\alpha_1 - 2\alpha_2 + \alpha_3 = O$. Find the rank of the matrix with α_1, α_2, and α_3 as columns, and thus verify Theorem IV.

THEOREM V. *If m and n are positive integers such that $m \leq n$ and if the rank r of the matrix A is less than m, where A is the matrix with the m n-dimensional vectors $\alpha_i = \{a_{i1}, a_{i2}, \ldots, a_{in}\}$, $i = 1, 2, \ldots, m$, as columns, then these m n-dimensional vectors are linearly dependent.*

To prove Theorem V it is convenient to let the rows and the columns of A be rearranged so that there is a square nonsingular submatrix of order r as the *leading square submatrix*, that is, a nonsingular submatrix of rank r is in the upper left-hand corner. Designate the rearranged form of A by

$$B = \begin{bmatrix} b_{11} & b_{21} & \cdots & b_{r1} & b_{p1} & \cdots & b_{m1} \\ b_{12} & b_{22} & \cdots & b_{r2} & b_{p2} & \cdots & b_{m2} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ b_{1r} & b_{2r} & \cdots & b_{rr} & b_{pr} & \cdots & b_{mr} \\ b_{1p} & b_{2p} & \cdots & b_{rp} & b_{pp} & \cdots & b_{mp} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ b_{1n} & b_{2n} & \cdots & b_{rn} & b_{pn} & \cdots & b_{mn} \end{bmatrix}. \tag{8}$$

Denote the m column vectors $\beta_1, \beta_2, \ldots, \beta_r, \beta_p, \ldots, \beta_m$ of the matrix B by

$$\beta_1 = \{b_{1i}\}, \beta_2 = \{b_{2i}\}, \ldots, \beta_r = \{b_{ri}\}, \beta_p = \{b_{pi}\}, \ldots, \beta_m = \{b_{mi}\},$$

in each of which i takes the value $1, 2, \ldots, n$. We want to prove that the column vectors $\{b_{1i}\}, \{b_{2i}\}, \ldots, \{b_{ri}\}, \{b_{pi}\}, \ldots, \{b_{mi}\}$, are linearly dependent. To do this we want to show the existence of scalars $s_1, s_2, \ldots, s_p, \ldots, s_m$ (not all zero) such that

$$s_1\{b_{1i}\} + s_2\{b_{2i}\} + \cdots + s_r\{b_{ri}\} + s_p\{b_{pi}\} + \cdots + s_m\{b_{mi}\} = O \quad (i = 1, 2, \ldots, r, p, \ldots, n). \tag{9}$$

In particular the first p columns of B are linearly dependent if there exist scalars $s_1, s_2, \ldots, s_p$ (not all zero) such that

$$
(10)\qquad \begin{cases} s_1b_{11} + s_2b_{21} + \cdots + s_rb_{r1} + s_pb_{p1} = 0, \\ s_1b_{12} + s_2b_{22} + \cdots + s_rb_{r2} + s_pb_{p2} = 0, \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ s_1b_{1r} + s_2b_{2r} + \cdots + s_rb_{rr} + s_pb_{pr} = 0, \\ s_1b_{1p} + s_2b_{2p} + \cdots + s_rb_{rp} + s_pb_{pp} = 0, \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ s_1b_{1n} + s_2b_{2n} + \cdots + s_rb_{rn} + s_pb_{pn} = 0. \end{cases}
$$

Let us fix our attention on a particular submatrix of B, designated by

$$
(11)\qquad B_p = \begin{bmatrix} b_{11} & b_{21} & \ldots & b_{r1} & b_{p1} \\ b_{12} & b_{22} & \ldots & b_{r2} & b_{p2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ b_{1r} & b_{2r} & \ldots & b_{rr} & b_{pr} \\ b_{1p} & b_{2p} & \ldots & b_{rp} & b_{pp} \end{bmatrix}, \quad p = r + 1.
$$

This square matrix B_p is of course singular, since $|B_p|$ is an $(r + 1)$-rowed minor of B, a matrix of rank r. Let $B_{1p}, B_{2p}, \ldots, B_{rp}, B_{pp}$ denote the cofactors of the last row, that is, the cofactors of the pth $=$ $(r + 1)$th row of $|B_p|$. Note that this makes $B_{pp} \neq 0$, since the leading square submatrix of order r is nonsingular. Expansion of $|B_p|$ by the last row gives

$$
(12)\qquad |B_p| = b_{1p}B_{1p} + b_{2p}B_{2p} + \cdots + b_{rp}B_{rp} + b_{pp}B_{pp}.
$$

Let

$$
(13)\qquad \{s_1, s_2, \ldots, s_r, s_p\} = \{B_{1p}, B_{2p}, \ldots, B_{rp}, B_{pp}\}.
$$

With this notation, and using the fact that $|B_p| = 0$, the relation (12) becomes

$$
(14)\qquad s_1b_{1p} + s_2b_{2p} + \cdots + s_rb_{rp} + s_pb_{pp} = 0.
$$

Thus the pth relation in the set (10) has been established.

Similarly, if i is any integer such that $1 \leq i \leq n$, and if B_i is defined by replacing the last row of B_p by $b_{1i}, b_{2i}, \ldots, b_{ri}, b_{pi}$, then B_i is singular since $|B_i|$ is zero; for if $1 \leq i \leq r$, then $|B_i|$ is zero because it has two rows alike; and if $r + 1 \leq i \leq n$, then $|B_i|$ is zero because it is an $(r + 1)$-rowed minor of a matrix of rank r. Also, for each value of $i (i = 1, 2, \ldots, n)$ the minors of the elements of the last row of B_i are equal respectively to

$$
(15)\qquad s_1 = B_{1p}, s_2 = B_{2p}, \ldots, s_r = B_{rp}, s_p = B_{pp}
$$

as defined above. Therefore for these same values of $s_1, s_2, \ldots, s_r, s_p$ it is true that

$$
(16)\qquad s_1b_{1i} + s_2b_{2i} + \cdots + s_rb_{ri} + s_pb_{pi} = 0
$$

for $i = 1, 2, \ldots, n$. But the relations (16) are precisely the conditions (10) for the linear dependence of the p vectors $\{b_{1i}\}, \{b_{2i}\}, \ldots, \{b_{ri}\}, \{b_{pi}\}$ of B. The scalar conditions (16) may be expressed in the matrix form

$$(17) \qquad \begin{bmatrix} b_{11} & b_{21} & \ldots & b_{r1} & b_{p1} \\ b_{12} & b_{22} & \ldots & b_{r2} & b_{p2} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ b_{1r} & b_{2r} & \ldots & b_{rr} & b_{pr} \\ b_{1p} & b_{2p} & \ldots & b_{rp} & b_{pp} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ b_{1n} & b_{2n} & \ldots & b_{rn} & b_{pn} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ \cdot \\ \cdot \\ \cdot \\ s_r \\ s_p \end{bmatrix} = O.$$

Clearly, if the condition (17) is satisfied, so also is

$$(18) \qquad B\gamma = O,$$

where

$$\gamma = \{s_1, s_2, \ldots, s_r, s_p, s_{p+1}, \ldots, s_m\} = \{s_1, s_2, \ldots, s_r, s_p, 0, \ldots 0\}.$$

That is, if rank $B = r < m$, then the m-dimensional vector with its first $(r + 1)$ elements as the cofactors of the elements in the $(r + 1)$th row of $|B_p|$ and zeros as the last $(n - r - 1)$ elements satisfies the equation $B\gamma = O$. Thus Theorem V is proved.

As an illustration of Theorem V, let $m = 4$, $n = 6$, $r = 2$. Then the relation corresponding to (18) is

$$(19) \qquad B\gamma = \begin{bmatrix} b_{11} & b_{21} & b_{31} & b_{41} \\ b_{12} & b_{22} & b_{32} & b_{42} \\ b_{13} & b_{23} & b_{33} & b_{43} \\ b_{14} & b_{24} & b_{34} & b_{44} \\ b_{15} & b_{25} & b_{35} & b_{45} \\ b_{16} & b_{26} & b_{36} & b_{46} \end{bmatrix} \begin{bmatrix} \begin{vmatrix} b_{21} & b_{31} \\ b_{22} & b_{32} \end{vmatrix} \\ -\begin{vmatrix} b_{11} & b_{31} \\ b_{12} & b_{32} \end{vmatrix} \\ \begin{vmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{vmatrix} \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

All of the elements in $B\gamma$ are zero, either because they are expansions of the determinant of the leading square submatrix of order $(r + 1)$, which is singular, or because they are expansions of $(2 + 1)$-rowed minors of B taken from the first 2 rows and some later row of B. All such minors are zero, since the rank of B is $r = 2$.

Clearly, the vector which we constructed as the solution of (18) in general and (19) in particular is not usually unique. For any column after the rth could have been moved into the position of the $(r + 1)$th column, thereby producing in general a different leading square submatrix of order $(r + 1)$, and consequently different cofactors as the elements of γ; thus in (19) the column $\{b_{4i}\}$, might just as well play the role of the elements of $\{b_{3i}\}$ in the vector γ.

From Theorems IV and V there follows

THEOREM VI. *The necessary and sufficient condition for the homogeneous equations*

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

in the m unknowns with coefficient matrix A of order n by m, $m \leq n$, to have a solution

$$\alpha = \{x_1, x_2, \ldots, x_m\}$$

different from the null vector is that the rank r of A satisfy the condition $r < m$.

In Theorem II we proved that if the columns of a square matrix A are linearly dependent, then $|A| = 0$. At that time we were not prepared to assert its converse, but we are now, and we do so in

THEOREM VII. *If A is a square matrix of order n and $|A| = 0$, then the columns of A are linearly dependent.*

Since $|A| = 0$, we know that $r < n$, r being the rank of A. So by Theorem VI the columns of A must be linearly dependent, and the theorem is proved.

THEOREM VIII. *If A is a square matrix of order n and $|A| = 0$, then the rows of A are linearly dependent.*

Theorem VIII follows since the rows of A are the columns of A', and from the fact that if $|A'| = 0$, so also is $|A| = 0$, since $|A| = |A'|$.

It is convenient to combine Theorems VII and VIII together in

THEOREM IX. *The vanishing of the determinant of a square matrix is a necessary and sufficient condition for the linear dependence of its rows and columns.*

We are now prepared to dispose of the very important

THEOREM X. *If the determinant rank of the rectangular matrix B is r, then the maximum number of linearly independent rows or columns in B is r; that is, the determinant rank, row rank, and column rank of a matrix are equal.*

Since B is a rectangular matrix of rank r, it has a nonsingular square submatrix A of order r. Consequently the columns (or rows) of B which contain A are linearly independent. On the other hand, any $(r + 1)$ columns (or rows) of B are linearly dependent; for if they were not they would contain some nonsingular square submatrix of order $r + 1$, which is contrary to the hypothesis that the rank of B be r. So the theorem is established.

We saw in Section 3–5 that any three 2-dimensional vectors are linearly dependent and that any four 3-dimensional vectors are linearly dependent; also in Ex. 3 of Section 4–6 we proved that any five 4-dimensional vectors are linearly dependent. We are now in a position to prove the correspondingly general theorem, namely,

THEOREM XI. *If m and n are positive integers such that $m > n$, then m n-dimensional vectors are linearly dependent.*

To prove Theorem XI, suppose the given m n-dimensional vectors to be column vectors of a matrix A, and adjoin $m - n$ rows with zero elements to make a larger matrix B of order m by m:

$$B = \begin{bmatrix} A \\ R \end{bmatrix},$$

R being the submatrix of B constituted of the $m - n$ rows with zero scalar elements. Then $|B| = 0$, and by Theorem IX the columns of B are linearly dependent. Therefore the columns of A are linearly dependent, and the theorem is established.

THEOREM XII. *Let $B = [A, \beta]$ be a matrix of order n by $(n + 1)$, with the nonsingular square submatrix A of order n. Then β, the $(n + 1)$th column of B, is either the null vector or is linearly dependent on the n columns of A in the manner $\beta = A\delta$.*

Since the rank of B is n, its $(n + 1)$ columns are linearly dependent by Theorem X. Therefore, by Theorem I there exists a non-null vector γ such that

(20) $$B\gamma = O, \qquad \text{or} \qquad [A, \beta]\gamma = O,$$

where

(21) $$\gamma = \{x_1, x_2, \ldots, x_n, -1\} = \{\delta, -1\}, \text{ say.}$$

Combining (20) and (21), we get

(22) $$[A, \beta]\begin{bmatrix} \delta \\ -1 \end{bmatrix} = O, \qquad \text{or} \qquad A\delta - \beta = O,$$

and finally

(23) $$A\delta = \beta.$$

Since A is nonsingular, equation (23) has a unique vector solution $\delta = A^{-1}\beta$, and the latter is the desired relation of linear dependence of the $(n+1)$ columns of B.

As a consequence of Theorem XII, we have

THEOREM XIII. *Any non-null n-dimensional column vector is expressible as a linear combination of the columns of a nonsingular square matrix A of order n.*

A similar argument gives us

THEOREM XIV. *Any non-null n-dimensional row vector is expressible as a linear combination of the rows of a nonsingular square matrix A of order n.*

Recalling the concept of a *basis* for a vector space, considered in Section 4–6, we may state as a consequence of the last two theorems,

THEOREM XV. *The vectors of a nonsingular square matrix of order n constitute a basis in terms of which an arbitrary vector of the same order and type (row or column) can be linearly expressed.*

EXERCISES

In each of the following exercises show that the rank r of the matrix A with the m given vectors as columns is less than m. Find r of the vectors on which each of the remaining $m - r$ vectors is linearly dependent. Represent each such dependence as an equation.

1. $\alpha_1 = \{1, 3, 2, -4\}$, $\alpha_2 = \{0, 2, 1, -1\}$, $\alpha_3 = \{-1, 1, 0, 2\}$.
2. $\alpha_1 = \{3, 2, 1, 0\}$, $\alpha_2 = \{-4, -3, 1, 0\}$, $\alpha_3 = \{2, 1, 1, 0\}$, $\alpha_4 = \{1, -1, 6, 0\}$.
3. $\alpha_1 = \{-1, 2, 5, 3\}$, $\alpha_2 = \{4, -8, -20, -12\}$.

9–6 Elementary transformations of a matrix. The concept of the rank of a matrix is a most important one, but if we were forced to find the (determinant) rank of a matrix directly by means of the definition of such rank, the task in most cases would be quite onerous. Fortunately the determination of the rank of a matrix A is facilitated by the following *elementary transformations* of a matrix:

(*i*) The interchange of two rows, or of two columns.

(*ii*) The addition to a row of a scalar multiple of another row, or the addition to a column of a scalar multiple of another column.

(*iii*) The multiplication of a row or a column by a nonzero scalar.

A matrix which is obtained from the unit matrix I by an elementary transformation is called an *elementary transformation matrix*. Corresponding to each of the elementary transformations of a given matrix, there is an elementary transformation matrix; for these we use the special distinguishing symbols explained below.

Type (*i*). *The interchange of two rows, or of two columns.* Let $E_{(ij)}$ be the matrix obtained by interchanging the ith and jth rows of the unit matrix I of order n. The effect of multiplying a given matrix A on the left by $E_{(ij)}$, giving $E_{(ij)}A$, is to *interchange the ith and jth rows of A*, and the effect of multiplying A on the right by $E_{(ij)}$, giving $AE_{(ij)}$, is to interchange the ith and jth columns of A. To illustrate, for $n = 3$,

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

Type (*ii*). *The addition to a row of a scalar multiple of another row, or the addition to a column of a scalar multiple of another column.* Let $H_{(ij)}$ be the matrix obtained from the unit matrix I of order n by the introduction of the scalar h in the ith row and the jth column, where $i \neq j$. The effect of multiplying a given matrix A on the left by $H(_{ij})$, giving $H_{(ij)}A$, is to add to the elements of the ith row of A, h times the elements of the jth row of A. The effect of multiplying A on the right by $H_{(ij)}$, yielding $AH_{(ij)}$, is *to add to the elements of the jth column of A, h times the elements of the ith column of A*. For example, when $n = 3$,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & h & 1 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} a_{11}, & a_{12}, & a_{13} \\ a_{21}, & a_{22}, & a_{23} \\ a_{31} + ha_{21}, & a_{32} + ha_{22}, & a_{33} + ha_{23} \end{bmatrix}.$$

Type (*iii*). *The multiplication of a row or a column by a scalar.* Let $K_{(ij)}$ stand for the matrix obtained from the identity matrix I by replacing the unit element in the ith row and ith column by k. The effect of multiplying A on the left by $K_{(ij)}$, giving $K_{(ij)}A$, is to multiply each element in the ith row of A by k; and the operation of multiplying A on the right by $K_{(ij)}$, giving $AK_{(ij)}$, *multiplies each element in the ith column of A by k*. To illustrate,

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

The matrices designated above by $E_{(ij)}$, $H_{(ij)}$, and $K_{(ij)}$, and defined as the results of elementary transformations on the unit matrix I, are called *elementary transformation matrices* of types (*i*), (*ii*), and (*iii*), respectively.

The various elementary transformations of a matrix A which we have just described are all of the form

(24) $PAQ = B, \qquad \text{where} \qquad |P| \neq 0 \qquad \text{and} \qquad |Q| \neq 0,$

with P and Q in some cases being the unit matrix I. In connection with (24) we also have

(25) $$|P| \cdot |A| \cdot |Q| = |B|.$$

The elementary rules for simplifying determinants may be related to examples of matrix multiplication just as (25) is related to (24).

9–7 Equivalent matrices. Two matrices A and B, each of order m by n, are said to be *equivalent* if and only if A can be changed to B by a finite number of elementary transformations of the kind specified in Section 9–6. This definition, together with the observation made above in connection with equation (24), leads directly to

THEOREM XVI. *An m by n matrix A is equivalent to a matrix B of the same order if and only if $B = PAQ$ for suitable nonsingular square matrices P and Q, of orders m and n respectively.*

In the above discussion of equivalence it is understood that the elements of P, A, Q, and therefore B, all belong to the same field. It is sometimes convenient to symbolize the relation of equivalence between A and B by $A \sim B$. We consider briefly some important characteristics of the equivalence relation.

(1) The property of equivalence is reflexive, for by taking P and Q each to be the unit matrix I, we get $IAI = A$; that is, A is equivalent to itself.

(2) The property of equivalence is symmetric: if A is equivalent to B, then B is equivalent to A. For if $B = PAQ$, then $A = P^{-1}BQ^{-1}$, since P and Q are nonsingular.

(3) The property of equivalence is transitive: if A is equivalent to B and B is equivalent to C, then A is equivalent to C. For if $B = PAQ$ and $C = RBS$, then $C = (RP)A(QS)$ with (RP) and (QS) nonsingular.

THEOREM XVII. *Equivalent matrices have the same rank.*

To prove this theorem, we reason as follows. Clearly, under any permutation of rows or columns of a matrix A the minors of A will receive at most a change of sign, but not of numerical value; so an elementary transformation of Type (i) will not nullify a nonvanishing minor of A or give a vanishing minor of A a nonzero value. A similar statement applies to the multiplication of a row or a column of A by a nonzero scalar, or an elementary operation of Type (iii). Therefore, elementary transformations of Types (i) and (iii) will leave the rank r of a matrix A unchanged. We now consider the effect of an elementary transformation of Type (ii) on the rank r of A. Let B be a matrix obtained from A by an elementary transformation of the second type. Any determinant $|K|$ of order $(r + 1)$ belonging to B either is a determinant of A [the value of a determinant is not changed if to the elements of any row (or column) are added a constant multiple of the corresponding elements of any other row (or column)], or it is of the form $|K| = |M| + k|N|$, where M and N are square $(r + 1)$ by $(r + 1)$ submatrices of A [if one of the rows (or columns) of a determinant consists of binomial elements, the determinant may be expressed as the sum of two determinants]. Since $|M|$ and $|N|$ are $(r + 1)$-rowed determinants of A, and A is of rank r, it follows that $|K| = 0$. Consequently the rank of A cannot be increased by an elementary transformation of Type (ii). The rank of B cannot be less than r; for if it were, an elementary transformation of Type (ii), which changes B into A, would entail an increase in rank; this we have just seen is impossible. Therefore the rank of B is r, and the theorem is proved.

As an illustrative exercise, verify that

$$A = \begin{bmatrix} 2 & 1 & -5 & -3 & 1 \\ 1 & 2 & 0 & 1 & 3 \\ 3 & 2 & -24 & -20 & -2 \\ 6 & 1 & -1 & 4 & 4 \end{bmatrix}$$

$$\sim \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 & 3 \\ 3 & 2 & -24 & -20 & -2 \\ 6 & 1 & -1 & 4 & 4 \end{bmatrix} r_1 \to 5r_1 - r_2 - r_3 - r_4$$

$$\sim \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 23 & 42 & -24 & -20 & 58 \\ 2 & -7 & -1 & 4 & -8 \end{bmatrix} \begin{cases} c_1 \to c_1 - c_4; \\ c_2 \to c_2 - 2c_4; \\ c_5 \to c_5 - 3c_4. \end{cases}$$

In the above notation, by $r_1 \to 5r_1 - r_2 - r_3 - r_4$ we mean that the first row of the given matrix is replaced by the indicated linear combination of the four rows, and similarly for the columns. Obviously, the last matrix has a nonzero third order determinant; so the rank of $A = 3$.

EXERCISES

1. Explain the nature of each of the following elementary transformation matrices; that is, explain the effect of multiplying a given matrix $A = [a_{ij}]$ on the left by each of the given matrices, and also on the right by the given matrix.

$$P = \begin{bmatrix} 1 & h & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}; \quad Q = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & k \\ 0 & 0 & 1 \end{bmatrix}; \quad R = \begin{bmatrix} 1 & 0 & r - hk \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

2. For the matrices of Ex. 1 show that

$$PQR = \begin{bmatrix} 1 & h & r \\ 0 & 1 & k \\ 0 & 0 & 1 \end{bmatrix}.$$

Prove in general that any matrix with units along the principal diagonal and all elements zero on one side of that diagonal can be expressed as a product of elementary transformation matrices.

With the aid of elementary transformations, find the rank r of each of the following matrices. If the rank r of the given matrix is less than m, m being the number of columns in the given matrix, find r of the column vectors on which each of the remaining $m - r$ vectors is linearly dependent. Exhibit each such dependence as an equation.

3. $$\begin{bmatrix} -1 & 2 & 3 & -5 \\ 3 & -4 & 5 & 2 \\ 5 & -6 & 13 & -1 \\ 0 & 2 & 14 & -13 \end{bmatrix}.$$

4. $$\begin{bmatrix} -5 & 1 & -5 & 5 & 1 \\ 2 & 1 & 0 & -1 & -1 \\ 5 & 3 & 1 & -1 & -1 \\ -1 & 1 & 0 & 3 & 2 \\ 3 & 1 & 2 & -1 & 0 \end{bmatrix}.$$ 5. $$\begin{bmatrix} 2 & 11 & 9 & 25 \\ 3 & 4 & 1 & 0 \\ 5 & 7 & 2 & 1 \\ 4 & 11 & 7 & 17 \\ 2 & 5 & 3 & 7 \end{bmatrix}.$$

9-8 System of homogeneous linear equations. We now consider in detail a number of aspects of the system of the m homogeneous linear equations in n unknowns,

$$\text{(26)} \qquad \begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = 0, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = 0, \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = 0; \end{cases}$$

or

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ x_n \end{bmatrix} = O,$$

or

$$A\delta = O, \tag{27}$$

where A is the coefficient matrix of order m by n (with m rows and n columns), and $\delta = \{x_1, x_2, \ldots, x_n\}$ is an indeterminate vector.

Note carefully that the roles of m and n here are interchanged with respect to the roles of these indices in Section 9–5; in particular, compare equations (26) of this section and (6) and (7) of Section 9–5. This apparent inconsistent use of the indicial symbolism can perhaps be justified by the following observations. In any consideration of an indeterminate vector, or of a set of such vectors, it is convenient and often desirable to represent the dimension of that vector or of that set of vectors by the letter n. In Section 9–5, as a natural consequence of the matrix interpretation (7) of the dependence relation (5) or (6), it was convenient to regard the column vectors of the related matrix as n-dimensional. But in (26) emphasis is being given to the indeterminate column vector $\delta = \{x_1, x_2, \ldots, x_n\}$, whose dimension we prefer to represent by n; this choice necessitates that n also be the number of columns in the coefficient matrix A, leaving m to denote the number of rows of A.

As a means of adapting the principal results of Section 9–5 to the present indicial symbolism, we restate Theorem VI in the form of

THEOREM XVIII. *A system of m homogeneous linear equations in n unknowns has a solution different from the zero vector if and only if the rank r of the coefficient matrix A is less than the number n of unknowns.*

It is evident that the equations (26) are satisfied by $\delta = \{0, 0, \ldots, 0\}$. This solution we call the *trivial solution*, and any solution different from it we call a *nontrivial solution*.

Suppose $m = n$, and A is a square matrix. For there to be a nontrivial solution δ such that $A\delta = O$ the columns of A must be linearly dependent. However, by Theorem II, if the columns of a square matrix A are linearly dependent, then $|A| = 0$. We have proved

THEOREM XIX. *A system of n homogeneous linear equations in n*

unknowns has a nontrivial solution if and only if the determinant of the coefficient matrix is zero.

In case there are more scalar variables than there are scalar equations in (26), $n > m$. Then r is at most equal to m, and so $r < n$. There follows

THEOREM XX. *A system of $m < n$ homogeneous equations in n unknowns always has a nontrivial solution.*

Our next project is to study the nontrivial solution of $m < n$ homogeneous equations and the nontrivial solution of $m = n$ homogeneous equations for which $r < n$. As we remarked in connection with Theorem V, such solution is not in general unique. Before investigating the arbitrariness of the solution we shall consider the reduction of a matrix to an equivalent form.

9–9 The reduction of a matrix to equivalent form. In equation (30) of Section 8–6 we had an excellent illustration of the reduction of a given matrix A to diagonal form. For the similarity transformation

(28) $$B = C^{-1}AC$$

on the matrix A is a special type of the equivalent transformation

(29) $$B = PAQ.$$

In Section 8–6 we found that a similarity equivalent transformation on a matrix A reduces A to the diagonal form. We now consider another equivalent transformation which reduces a given matrix A to a desirable form.

Let the m by n matrix A of equations (26) and (27) be of rank r. By elementary transformations we may bring a nonsingular square submatrix of order r in the upper left-hand corner. Therefore, there exist nonsingular square matrices P and Q of orders m and n such that

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

is of rank r. Since A_{11} is nonsingular, it has an inverse A_{11}^{-1}. This makes possible further reductions by means of multiplying partitioned matrices. For

(30) $$\begin{bmatrix} I & O \\ -A_{21}A_{11}^{-1} & I \end{bmatrix}\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ O & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix}.$$

The matrix

$$A_{22} - A_{21}A_{11}^{-1}A_{12}$$

is a null matrix. For if it had a nonzero element k, then the submatrix consisting of A_{11} and bordered by the row and column containing k would be nonsingular and of rank $(r + 1)$. But that is impossible, since the rank r of the matrix A is invariant under equivalent transformations. We have proved

THEOREM XXI. *A rectangular matrix A of rank r can be reduced by equivalent transformations to*

$$\begin{bmatrix} A_{11} & A_{12} \\ O & O \end{bmatrix},$$

where A_{11} is a nonsingular square matrix of rank r.

For subsequent reference it is desirable for us to note that the partitioned multiplication (30) is, as a consequence of the discussion immediately preceding Theorem XXI, an equivalent transformation of the form

$$(31) \quad P\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ O & O \end{bmatrix}, \quad \text{where} \quad P = \begin{bmatrix} I & O \\ -A_{21}A_{11}^{-1} & I \end{bmatrix}.$$

From (31) we get, since P is nonsingular,

$$(32) \qquad \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = P^{-1}\begin{bmatrix} A_{11} & A_{12} \\ O & O \end{bmatrix}.$$

Similarly, consider the equivalent transformation

$$\begin{bmatrix} I & O \\ -A_{21}A_{11}^{-1} & I \end{bmatrix}\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\begin{bmatrix} I & -A_{11}^{-1}A_{12} \\ O & I \end{bmatrix}$$
$$= \begin{bmatrix} A_{11} & O \\ O & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix} = \begin{bmatrix} A_{11} & O \\ O & O \end{bmatrix},$$

since the matrix $A_{22} - A_{21}A_{11}^{-1}A_{12}$ is a null matrix for the same reason given above. We then have

THEOREM XXII. *A rectangular matrix A of rank r can be reduced by equivalent transformations to the form*

$$\begin{bmatrix} A_{11} & O \\ O & O \end{bmatrix},$$

where A_{11} is a nonsingular square matrix of rank r.

Note further that

$$\begin{bmatrix} A_{11} & O \\ O & O \end{bmatrix}\begin{bmatrix} A_{11}^{-1} & O \\ O & I \end{bmatrix} = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix},$$

where the submatrix I_r is the unit matrix of order r. We have proved

THEOREM XXIII. *A rectangular matrix of rank r can be reduced by equivalent transformations to the form*

$$\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}$$

where I_r is the unit matrix of order r; that is, there exist nonsingular matrices P and Q such that for $A = [a]_m^m$,

$$PAQ = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix}.$$

This is called the canonical form of a matrix under equivalent transformations.

To illustrate the procedure, suppose that we want to find some matrices P and Q such that PAQ is diagonal, where

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

Let us interchange the first and second rows of A. By Section 9-6 this may be effected by premultiplying A by $E_{(12)} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Thus

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

In the latter matrix let us replace the first column by the first column minus twice the second column. This is equivalent to postmultiplication by an elementary transformation matrix of Type (*ii*), Section 9-6, the transformation matrix being $H_{(12)} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}$. Finally,

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

So a P and a Q such that PAQ is diagonal are

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad Q = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix}.$$

9–10 Solution of a system of homogeneous linear equations. Theorem XXI gives us a basis for discussing the arbitrariness of the solution of the homogeneous equations (26). We assume that the equations and the unknowns have been rearranged so that the leading square submatrix A_{11} is of order r and nonsingular. That is, we assume that we may write (27) in the form

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = O, \tag{33}$$

where δ has been partitioned into its first r and last $n - r$ elements. Combining (32) and (33), we get

$$P^{-1}\begin{bmatrix} A_{11} & A_{12} \\ O & O \end{bmatrix}\begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = O. \tag{34}$$

Premultiplying both sides of (34) by P we get

$$\begin{bmatrix} A_{11} & A_{12} \\ O & O \end{bmatrix}\begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = O. \tag{35}$$

Then

$$A_{11}\delta_1 + A_{12}\delta_2 = O, \quad \text{or} \quad A_{11}\delta_1 = -A_{12}\delta_2.$$

Consequently,

$$\delta_1 = -A_{11}^{-1}A_{12}\delta_2. \tag{36}$$

This form of the solution for δ_1 shows that the last $(n - r)$ elements of δ (the elements of δ_2) can be assigned arbitrarily, and when this is done the first r elements of δ (the elements of δ_1) are uniquely determined. Hence the number of linearly independent vectors of the type δ_2 cannot exceed $n - r$. We have proved

THEOREM XXIV. *If the coefficient matrix A of (26) is of rank $r < n$, n being the number of unknowns, and if the columns of A corresponding to a designated r unknowns constitute a matrix of rank r, then the remaining $n - r$ unknowns can be arbitrarily assigned. These assigned $n - r$ unknowns are parameters in terms of which the other r unknowns can be linearly and uniquely expressed.*

As an illustration, consider the equations

$$\begin{cases} x_1 + x_2 - \ \ x_3 + x_4 = 0, \\ x_1 - x_2 + 2x_3 - x_4 = 0, \\ \quad\ 3x_1 + \ x_2 + x_4 = 0; \end{cases} \tag{37}$$

or

$$\begin{bmatrix} 1 & 1 & -1 & 1 \\ 1 & -1 & 2 & -1 \\ 3 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = O.$$

Here rank A is 2, and $|A_{11}| = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2 \neq 0$. A_{11} is nonsingular and

$$A_{11}^{-1} = -\tfrac{1}{2}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}.$$

In this case the solution (36) becomes

$$(38) \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} x_3 \\ x_4 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} -1 & 0 \\ 3 & -2 \end{bmatrix}\begin{bmatrix} x_3 \\ x_4 \end{bmatrix};$$

or

$$x_1 = -\tfrac{1}{2}x_3, \quad x_2 = \tfrac{3}{2}x_3 - x_4.$$

Thus the three equations (37) in the four unknowns x_1, x_2, x_3, x_4 are solvable in terms of x_3 and x_4. From Theorem XXIV we know that there are two linearly independent solutions

$$\alpha_1 = \{x_1, x_2, x_3, x_4\} \qquad \text{and} \qquad \alpha_2 = \{x_1', x_2', x_3', x_4'\}$$

of (37), called a *fundamental set of solutions* of these equations, with the property that any other solution of them is a linear combination of the linearly independent solutions.

Take $x_3 = 1$, $x_4 = 0$, and we get from (38) as one solution of (37)

$$\alpha_1 = \{-\tfrac{1}{2}, \tfrac{3}{2}, 1, 0\}.$$

The choice of $x_3 = -2$, $x_4 = 1$, with the relations (38), gives as a second solution of (37)

$$\alpha_2 = \{1, -4, -2, 1\}.$$

Note that the choices of x_3 and x_4 were made so that α_1 and α_2 are linearly independent vectors. Therefore the vectors α_1 and α_2 just stipulated are two linearly independent solutions of equation (37). Any other solution δ of (37) is a linear combination of these vectors, as

$$\delta = s_1\alpha_1 + s_2\alpha_2,$$

where s_1 and s_2 are scalars to which values may be assigned arbitrarily.

Exercises

Find nonsingular matrices P and Q such that PAQ is diagonal for the matrix A in each of Exercises 1 and 2.

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

2. $A = \begin{bmatrix} 2 & -1 & -6 \\ -3 & 0 & 8 \\ 1 & 5 & 10 \end{bmatrix}$.

3. For the given matrix

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

verify that an equivalent transformation that will put A in diagonal form is $IAQ = I$, where $I = [1]_3^3$ and

$$Q = \begin{bmatrix} 3 & 1 & 2 \\ 2 & 1 & 2 \\ 6 & 2 & 5 \end{bmatrix}.$$

4. Show that if δ is a solution of equation (27), and if s is an arbitrary scalar, then $s\delta$ is also a solution of (27).

5. Show that if α_1 and α_2 are solutions of equation (27), and if s_1 and s_2 are arbitrary scalars, then $s_1\alpha_1 + s_2\alpha_2$ is also a solution of (27).

6. For each of the following sets of equations, find the rank r of the coefficient matrix. If $r < n$, find a set of $n - r$ linearly independent solutions like those in the illustrated exercise just given.

(*i*) $2x_1 - 7x_2 - 6x_3 = 0,$
$3x_1 + 5x_2 - 2x_3 = 0,$
$4x_1 - 2x_2 - 7x_3 = 0.$

(*ii*) $2x_1 - x_2 + x_3 - x_4 = 0,$
$x_1 + x_2 - x_3 + x_4 = 0,$
$4x_1 + x_2 - x_3 + 3x_4 = 0.$

(*iii*) $2x_1 + 3x_2 + 5x_3 = 0,$
$x_1 - x_2 + 2x_3 = 0,$
$x_1 + 4x_2 + 3x_3 = 0.$

(*iv*) $x_1 - 2x_2 + x_3 - x_4 = 0,$
$2x_1 - x_2 - 2x_3 + x_4 = 0,$
$-x_1 - 4x_2 + 7x_3 - 5x_4 = 0,$
$8x_1 - 7x_2 - 4x_3 + x_4 = 0.$

9–11 Solution of a system of nonhomogeneous linear equations. The system of m nonhomogeneous linear equations in n unknowns

(39)
$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdot \qquad \cdot \qquad \cdot \qquad \cdot \qquad \cdot \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

or

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

or

(40)
$$A\delta = \beta,$$

may be solved in a manner similar to the system of homogeneous linear equations considered in the preceding two sections. As there, we assume that the equations and the unknowns have been arranged so that the leading square submatrix A_{11} is of order r and nonsingular. Then (39) may be written in the form

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \tag{41}$$

where the vector δ has been partitioned into its first r and last $n - r$ elements, and β has been partitioned into its first r and last $m - r$ elements. From (31) and (41) there follows the relation

$$\begin{bmatrix} A_{11} & A_{12} \\ O & O \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = P \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}, \text{ say.} \tag{42}$$

Now the left member of equation (42) consists of a column of r nonzero elements followed by $m - r$ zeros. Consequently that equation can be solved for δ_1 if and only if γ_2 is a null vector. Referring to equation (31) we see that

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} = P \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} I & O \\ -A_{21}A_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \beta_1 \\ -A_{21}A_{11}^{-1}\beta_1 + \beta_2 \end{bmatrix}. \tag{43}$$

Note that β_1 is left unchanged upon premultiplication of $\{\beta_1, \beta_2\}$ by this matrix P. If $\gamma_2 = O$, equation (42) can be written

$$A_{11}\delta_1 + A_{12}\delta_2 = \beta_1, \tag{44}$$

or

$$\delta_1 = -A_{11}^{-1}A_{12}\delta_2 + A_{11}^{-1}\beta_1. \tag{45}$$

If $\gamma_2 = O$ and solution (45) is possible, equations (39) are said to be *consistent*; if $\gamma_2 \neq O$ and equations (39) have no solution, those equations are said to be *inconsistent*.

The condition for consistency, $\gamma_2 = O$, may be expressed equivalently using the concept of rank. If we compare the partitioned matrices

$$\begin{bmatrix} A_{11} & A_{12} \\ O & O \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} A_{11} & A_{12} & \gamma_1 \\ O & O & \gamma_2 \end{bmatrix}, \tag{46}$$

we see that they have the same rank if and only if $\gamma_2 = O$. Premultiplying each of the matrices (46) by P^{-1}, which operation does not alter rank, we have the condition that A and $[A, \beta]$ have the same rank. The matrix $[A, \beta]$ is called the *augmented matrix* of the system of equations (39). We have proved

THEOREM XXV. *The m nonhomogeneous linear equations in n unknowns (39) are consistent and solvable in terms of $n - r$ parameters if and only if the rank of the augmented matrix $[A, \beta]$ is equal to the rank r of A.*

To illustrate, consider the equations

$$(47)\qquad \begin{cases} x_1 + x_2 + x_3 + 2x_4 = 3, \\ 2x_1 - x_2 + 3x_3 + 6x_4 = 9, \\ 3x_1 \qquad\quad + 4x_3 + 8x_4 = 12, \\ x_1 - 2x_2 + 2x_3 + 4x_4 = 6; \end{cases}$$

or

$$\begin{bmatrix} 1 & 1 & 1 & 2 \\ 2 & -1 & 3 & 6 \\ 3 & 0 & 4 & 8 \\ 1 & -2 & 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3 \\ 9 \\ 12 \\ 6 \end{bmatrix}.$$

Here rank A = rank $[A, \beta]$ is 2, and A has a nonsingular 2 by 2 submatrix

$$A_{11} = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix} \quad \text{with the inverse} \quad A_{11}^{-1} = -\tfrac{1}{3} \begin{bmatrix} -1 & -1 \\ -2 & 1 \end{bmatrix}.$$

The equations (47) are consistent, and the general solution (45) in this case becomes

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} -1 & -1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} x_3 \\ x_4 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} -1 & -1 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 9 \end{bmatrix}$$

or

$$(48)\qquad x_1 = \tfrac{1}{3}(-4x_3 - 8x_4) + 4, \quad x_2 = \tfrac{1}{3}(x_3 + 2x_4) - 1.$$

Thus the four equations (47) in the four unknowns x_1, x_2, x_3, x_4 are solvable in terms of x_3, x_4. By Theorem XXV there are two linearly independent solutions,

$$\alpha_1 = \{x_1, x_2, x_3, x_4\} \qquad \text{and} \qquad \alpha_2 = \{x_1', x_2', x_3', x_4'\}$$

of (47), called a *fundamental set of solutions* of those equations, with the property that any other solution of them is a linear combination of the linearly independent solutions.

Take $x_3 = 0, x_4 = 1$, and we get from (48) as one solution of (47)

$$\alpha_1 = \{\tfrac{4}{3}, -\tfrac{1}{3}, 0, 1\}.$$

Take $x_3 = 1, x_4 = 0$, and we get from (48) as a second solution of (47)

$$\alpha_2 = \{\tfrac{8}{3}, -\tfrac{2}{3}, 1, 0\}.$$

It should be evident that the values of x_3 and x_4 were chosen so that α_1 and α_2 are linearly independent vectors. Therefore α_1 and α_2 are two linearly independent solutions of the equations (47). Any other solution δ of (47) is a linear combination of these two vectors, as

$$\delta = s_1\alpha_1 + s_2\alpha_2,$$

where s_1 and s_2 are arbitrary scalars.

Exercises

Determine whether the following systems of equations are consistent or inconsistent. If the equations are consistent, find $n - r$ linearly independent vector solutions as in the illustrated exercise just given.

1. $$\begin{aligned} 2x_1 - x_2 + 3x_3 &= 1, \\ 4x_1 - 2x_2 - x_3 &= -3, \\ 2x_1 - x_2 - 4x_3 &= -4, \\ 10x_1 - 5x_2 - 6x_3 &= -10. \end{aligned}$$

2. $$\begin{aligned} 2x_1 - 3x_2 + 4x_3 - x_4 - 3 &= 0, \\ x_1 + 2x_2 - x_3 + 2x_4 - 1 &= 0, \\ 3x_1 - x_2 + 2x_3 - 3x_4 - 1 &= 0, \\ 3x_1 - x_2 + x_3 - 7x_4 - 4 &= 0. \end{aligned}$$

9–12 Congruent matrices. A special instance of the equivalent transformation $B = RAQ$ is the *congruent transformation*

$$B = P'AP, \tag{49}$$

P' being the transpose of P. Such transformation plays a basic role in connection with symmetric bilinear forms and quadratic forms, which are considered in the next chapter. If for two given matrices A and B there exists a nonsingular matrix P so that (49) is satisfied, we say that A and B are *congruent matrices.*

Preparatory to giving a proof of Theorem XXVIII below, which is based upon congruent transformations of matrices, let us re-examine the elementary transformations of a matrix which were considered in Section 9–6. Recall that $E_{(ij)}$ denotes the matrix obtained from the unit matrix I of order n by interchanging the ith and jth rows without altering the remaining $n - 2$ rows. Thus from

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

we derive

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

by interchanging the second and third rows of A. Note that

$$E'_{(23)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = E_{(23)},$$

so for $n = 3$ the transformation

$$E_{(23)}AE_{(23)} = E'_{(23)}AE_{(23)}$$

is congruent. More generally, $E'_{(ij)} = E_{(ij)}$, and so

$$E_{(ij)}AE_{(ij)} = E'_{(ij)}AE_{(ij)}$$

is a congruent transformation. But the elementary transformation $E_{(ij)}AE_{(ij)}$ interchanges the ith and jth rows and the ith and jth columns of A. We have proved

THEOREM XXVI. *The elementary transformation*

(50) $$E_{(ij)}AE_{(ij)}$$

of a matrix A which results in the simultaneous interchange of the same rows and columns of A is a congruent transformation.

Recall that $H_{(ij)}$ signifies the matrix obtained from the unit matrix I of order n by the insertion of the scalar h in the ith row and jth column $(i \neq j)$ without changing the other elements of the unit matrix. Thus

$$H_{(32)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & h & 1 \end{bmatrix}.$$

Observe that

(51) $$H'_{(32)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & h \\ 0 & 0 & 1 \end{bmatrix} = H_{(23)}.$$

So

$$H_{(23)}AH_{(32)} = H'_{(32)}AH_{(32)}$$

is congruent. In general, $H'_{(ij)} = H_{(ji)}$. Consequently

$$H_{(ji)}AH_{(ij)} = H'_{(ij)}AH_{(ij)},$$

which effects the transformation of A symbolized by

$$\begin{cases} \text{row } j \to \text{row } j + h \text{ row } i \\ \text{col } j \to \text{col } j + h \text{ col } i, \end{cases}$$

is a congruent transformation. We have established

THEOREM XXVII. *The elementary transformation*

(52) $$H_{(ji)}AH_{(ij)}$$

of the matrix A which increases the elements of the jth row of A by h times the corresponding elements of the ith row, and at the same time increases the elements of the jth column of A by h times the corresponding elements of the ith column, is a congruent transformation.

We are now in a position to prove

THEOREM XXVIII. *A symmetric matrix $A = [a]_n^n$ of rank r can be reduced by a congruent transformation to a diagonal matrix of the same rank,*

$$P'AP = \begin{bmatrix} d_{11} & 0 & \dots & 0 & \dots & 0 \\ 0 & d_{22} & \dots & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & d_{rr} & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & 0 & \dots & 0 \end{bmatrix} \tag{52}$$

To prove this important theorem we reason as follows:

(*i*) If every a_{ij} is zero, the rank of A is zero, and the theorem is trivial.

(*ii*) If $a_{11} = 0$ but for some i, $a_{ii} \neq 0$, then a_{ii} can be brought to the leading position by an elementary transformation of the type (50). To illustrate,

$$E_{(12)} \begin{bmatrix} 0 & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} E_{(12)} = \begin{bmatrix} a_{21} & a_{22} & a_{23} \\ 0 & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} E_{(12)}$$

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

(*iii*) If all $a_{ii} = 0$ but some $a_{ij} \neq 0$ for $i < j$, an elementary transformation of the type (51),

$$\text{row } i \to \text{row } i + \text{row } j; \quad \text{col } i \to \text{col } i + \text{col } j,$$

places the element $2a_{ij}$ in the position a_{ii}. Then by (*ii*) above it can be brought to the leading position of a_{11}. For example, for $h = 1$ in (51), we have ($a_{23} = a_{32} \neq 0$)

$$H_{(23)}AH_{(32)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & a_{12} & a_{13} \\ a_{21} & 0 & a_{23} \\ a_{31} & a_{32} & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0, & a_{12}, & a_{13} \\ a_{21} + a_{31}, & a_{32}, & a_{23} \\ a_{31}, & a_{32}, & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} =$$

$$= \begin{bmatrix} 0, & a_{12} + a_{13}, & a_{13} \\ a_{21} + a_{31}, & a_{32} + a_{23}, & a_{23} \\ a_{31}, & a_{32}, & 0 \end{bmatrix}.$$

It should be realized that the transitive property previously established for equivalent transformations also holds for congruent transformations. For if $B = P'AP$ and $C = Q'BQ$, then $C = Q'P'APQ = R'AR$, where $R = PQ$. Therefore the net effect of the several congruent transformations made above on the given matrix A is expressible as some congruent transformation:

$$\text{(54)} \qquad P'AP = B = \begin{bmatrix} b_{11} & \beta' \\ \beta & B_{22} \end{bmatrix}, \quad \text{say.}$$

Here $\beta' = (b_{12}, b_{13}, \ldots, b_{1n})$. Since the given matrix A is symmetric, the matrix on the right of (54) is also symmetric. For

$$B' = (P'AP)' = P'A'P = P'AP.$$

Next consider the congruent transformation

$$\text{(55)} \quad \begin{bmatrix} 1 & O \\ -\beta b_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} b_{11} & \beta' \\ \beta & B_{22} \end{bmatrix} \begin{bmatrix} 1 & -b_{11}^{-1}\beta' \\ O & I \end{bmatrix}$$

$$= \begin{bmatrix} b_{11} & \beta' \\ O & -\beta b_{11}^{-1}\beta' + B_{22} \end{bmatrix} \begin{bmatrix} 1 & -b_{11}^{-1}\beta' \\ O & I \end{bmatrix} = \begin{bmatrix} b_{11} & O \\ O & B_{22} - \beta b_{11}^{-1}\beta' \end{bmatrix}.$$

Thus all elements of the first row and the first column are reduced to zero except the first. We proceed similarly with the submatrix of order $n - 1$, $B_{22} - \beta b_{11}^{-1}\beta'$, until r elements are isolated in the diagonal with $a_{ii} = 0 (i > r)$, or else $r = n$. All of these congruent transformations may be combined into one, indicated by (53).

The matrix on the right of equation (53) is called the *canonical form of a symmetric matrix under congruent transformations.* Since all of the transformations which we made on the elements of A to effect the reduction (53) were rational, that reduction is sometimes spoken of as the *rational reduction* of a symmetric matrix under congruent transformations.

To illustrate, let us find a matrix P such that $P'AP$ is diagonal, where $A = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}$. Since this is already in the form $B = \begin{bmatrix} b_{11} & \beta' \\ \beta & B_{22} \end{bmatrix}$ of (54) above, we see from (55) that

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

So $P = \begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & 1 \end{bmatrix}$.

Exercises

Find a nonsingular matrix P such that $P'AP$ is diagonal for each symmetric matrix A given below.

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

3. $A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & 2 & 0 \end{bmatrix}$. 4. $A = \begin{bmatrix} 0 & 2 & 3 \\ 2 & 4 & 0 \\ 3 & 0 & 1 \end{bmatrix}$. 5. $A = \begin{bmatrix} 2 & 1 & 8 \\ 1 & -3 & 9 \\ 8 & 9 & 2 \end{bmatrix}$.

6. Prove that if a symmetric matrix has rank 1, then the elements of its leading diagonal cannot all be zero.

7. Prove that all the elements of a symmetric matrix of rank 1 can be expressed in terms of the elements of its leading diagonal.

8. Prove that no skew symmetric matrix can have rank 1.

9. By considering the effect of interchanging rows and columns, prove that the determinant of a skew symmetric matrix of odd order is zero.

10. Prove that the rank of a skew symmetric matrix is always even.

CHAPTER 10

MATRICES AND ALGEBRAIC FORMS

10–1 Linear forms. Invariants. Let $A = [a_{ij}] = [a]_n^n$ be a matrix with elements in some field **F**, and let $\alpha = \{x_i\}$ be a vector with indeterminate scalar coordinates in the same field **F**. An expression such as

(1) $$a_{ij}x_j = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n,$$

for a fixed value of i, is called a *linear form* in the variables x_j. In this chapter we make frequent use of the summation convention, and the reader should review the explanation of that convention given in Section 5–10.

If we are dealing with only one or two linear forms, it is convenient to represent them by

(2) $$f = a_ix_i \quad \text{and} \quad g = b_ix_i.$$

From the rule for the addition of vectors we have a fundamental property of linear forms:

$$f + g = (a_i + b_i)x_i.$$

More often we are interested in *a system of n linear forms in n variables.* Such a system is represented by (1) if we let i take successively the values $1, 2, \ldots, n$. In detail such a system is

(3) $$\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n, \\ \cdot \qquad \cdot \qquad \cdots \qquad \cdot \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n. \end{array}$$

The matrix $A = [a_{ij}] = [a]_n^n$ is associated with the system (3) of n linear forms in n variables; we say A is the *matrix of the system.*

Suppose that the system (3) is subjected to the linear transformation

(4) $$x_i = t_{ij}\bar{x}_j$$

or

(5) $$[x]_1^n = [t]_n^n[\bar{x}]_1^n,$$

with matrix $T = [t]^n_n$. Change indices in (4) so that the transformation appears as

(6) $$x_j = t_{jk}\bar{x}_k.$$

Substituting from (6) in (1), we get

(7) $$a_{ij}t_{jk}\bar{x}_k = \bar{a}_{ik}\bar{x}_k, \quad \text{say.}$$

From (7) and the procedure for multiplying matrices, we see that

(8) $$\bar{a}_{ik} = a_{ij}t_{jk},$$

or

(9) $$[\bar{a}]^n_n = [a]^n_n[t]^n_n,$$

where $\bar{A} = [\bar{a}_{ik}] = [\bar{a}]^n_n$ is the matrix of the system of the n linear forms in the n variables $\bar{x}_i$. We have proved

THEOREM I. *A linear transformation with matrix T replaces a system of n linear forms in n variables with the matrix A by a system with the matrix AT.*

The determinant of the matrix of a system of linear forms is called the *eliminant* of the system. Thus $|A| = |a_{ij}|$ is the eliminant of the system (1). Taking the determinant of both sides of the relation (9), we get

(10) $$|\bar{A}| = |A| \cdot |T|.$$

The eliminant $|A|$ is one of an important class of mathematical objects called *invariants*.

A function of the coefficients of a form or of a system of forms in the x's is said to be a *relative invariant* if, whatever the matrix T of the transformation (5), the same function of the coefficients of the resulting forms in the $\bar{x}$'s is equal to the original function of the coefficients multiplied by a certain power of the determinant $|T|$. If the notation for the forms is that of (3) above, and $I(a_{11}, \ldots, a_{nn})$ represents a function of the coefficients of the forms, then $I(a_{11}, \ldots, a_{nn})$ is an invariant if

(11) $$I(\bar{a}_{11}, \ldots, \bar{a}_{nn}) = |T|^w I(a_{11}, \ldots, a_{nn}).$$

The power w of the determinant of T in (11) is called the *weight* of the invariant.

If either $|T| = 1$ or $w = 0$, then $|T|^w = 1$, and

(12) $$I(\bar{a}_{11}, \ldots, \bar{a}_{nn}) = I(a_{11}, \ldots, a_{nn}).$$

In such case we say that I is an *absolute invariant*. Thus an absolute invariant is an invariant under a transformation for which $|T| = 1$, or it is an invariant of weight zero.

From (10) and the definition of a relative invariant, we have

THEOREM II. *The eliminant of a system of n linear forms in n variables is a relative invariant of weight* 1.

From Theorem XIX of Section 9–8, we have

THEOREM III. *The vanishing of the eliminant of the n linear forms*

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n \quad (i = 1, 2, \ldots, n)$$

is the necessary and sufficient condition for the n equations

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = 0 \quad (i = 1, 2, \ldots, n)$$

to have a solution other than the trivial one $x_1 = \ldots = x_n = 0$.

The entities which we have just called *invariants* could be more descriptively termed *algebraic invariants*, to contrast them with arithmetic invariants, differential invariants, and integral invariants. Of the latter we use only arithmetic invariants. A number associated with a form or a system of forms in the x's which is unchanged when the variables are subjected to all the transformations of a specified set is an *arithmetic invariant*. Examples of *arithmetic invariants* are the following: the number of real points at which a straight line cuts a curve, and the degree of a curve with respect to linear transformations of its variables in its equation.

10–2 Bilinear forms.

An expression such as

$$f = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}x_iy_j, \tag{13}$$

which is a polynomial in the $m + n$ scalar variables

$$x_1, x_2, \ldots, x_m; \quad y_1, y_2, \ldots, y_n$$

with each of its terms of the first degree in the x's and also in the y's, is called a *bilinear form*.

For $m = 2$ and $n = 3$ the general bilinear form is

$$g = \sum_{i=1}^{2} \sum_{j=1}^{3} a_{ij}x_iy_j = a_{11}x_1y_1 + a_{12}x_1y_2 + a_{13}x_1y_3 + a_{21}x_2y_1 + a_{22}x_2y_2 + a_{23}x_2y_3, \tag{14}$$

or

$$g = (x_1, x_2)\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \alpha' A\beta. \tag{15}$$

Relative to (14) and (15) we say the given bilinear form g has the *matrix factors*

$$\alpha' = (x_1, x_2), \quad A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}, \quad \text{and} \quad \beta = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}.$$

Recall our agreement near the end of Section 5–6 that we would use *unprimed* small Greek letters as $\alpha, \beta, \gamma \ldots$ to represent *one-column matrices*, and relatedly to use primed small Greek letters $\alpha', \beta', \gamma', \ldots$ to represent *one-row matrices*. Analogous to the matric factorization (15), the general bilinear form (13) has the matric factorization

$$f = \alpha' A\beta = (x_1, x_2, \ldots, x_m)\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ y_n \end{bmatrix}. \tag{16}$$

Clearly, to every bilinear form like (13) there corresponds a matrix $A = [a]^m_n$, and conversely to every such matrix there corresponds a bilinear form; there is a one-to-one correspondence between such forms and matrices. $A = [a]^m_n$ is said to be the *matrix of the bilinear form* (13), and the rank of A is called the *rank of the form*. Suppose that the vectors

$$\alpha = \{x_1, x_2, \ldots, x_m\} \qquad \text{and} \qquad \beta = \{y_1, y_2, \ldots, y_n\}$$

of the bilinear form

$$f = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}x_iy_j = \alpha' A\beta \tag{17}$$

are subjected to the separate nonsingular linear transformations

$$x_i = p_{ij}\bar{x}_j \quad (i, j = 1, 2, \ldots, m) \tag{18}$$

or

$$\alpha = P\bar{\alpha}, \tag{19}$$

and

$$y_i = q_{ij}\bar{y}_j \quad (i, j = 1, 2, \ldots, n) \tag{20}$$

or

$$\beta = Q\bar{\beta}. \tag{21}$$

Note carefully that to effect the transformation (19) on (17) we have first to take the transpose of both sides of (19). Doing this, we get

$$\alpha' = \bar{\alpha}'P', \tag{22}$$

since the transpose of the product of two matrices is the product of their transposes in reverse order. Substituting from (21) and (22) in (17), we see that (17) is transformed into the bilinear form

$$\bar{f} = \bar{\alpha}'P'AQ\bar{\beta} = \bar{\alpha}'\bar{A}\bar{\beta}, \quad \text{say}. \tag{23}$$

We have proved

THEOREM IV. *If in the bilinear form* (17) *with matrix A we subject the x's to a linear transformation with matrix P and the y's to a linear transformation with matrix Q, we obtain a new bilinear form with matrix*

$$\bar{A} = P'AQ, \tag{24}$$

where P′ is the transpose of P.

Recall from Section 9–7 that if two matrices A and B satisfy the relation

$$B = PAQ, \tag{25}$$

then A and B are said to be *equivalent* matrices, and also that equivalent matrices have the same rank. These observations lead us to

THEOREM V. *The rank of the bilinear form* (17) *is an arithmetic invariant with respect to nonsingular linear transformations of the x's and y's.*

Two bilinear forms are said to be *equivalent forms* if their matrices are equivalent. So equivalent *bilinear forms have the same rank.* From Theorem XXIII of Chapter 9 and Theorem IV above there follows

THEOREM VI. *The bilinear form* (17) *of rank r can be reduced by nonsingular linear transformations of the variables as* (18) *and* (20) *to the normal form*

$$x_1y_1 + x_2y_2 + \cdots + x_ry_r. \tag{26}$$

We see that the reduction of the bilinear form (17) to its normal form (26) is essentially that of the reduction of the matrix $A = [a]_n^m$ of the bilinear form to its canonical form under equivalent transformations.

EXERCISES

Find the matrix A associated with each of the following bilinear forms. Reduce each bilinear form to the normal form (26) by finding nonsingular matrices P and Q such that PAQ is diagonal.

1. $2x_1y_1 + x_1y_2 + x_2y_1$.
2. $3x_1y_1 + 2x_1y_2 + 4x_2y_1 + x_2y_2$.
3. $2x_1y_1 - x_1y_2 - 6x_1y_3 - 3x_2y_1 + 3x_2y_3 + x_3y_1 + 5x_3y_2 + 19x_3y_3$.

10-3 Symmetric bilinear forms. The bilinear form

$$f = a_{ij}x_iy_j = \alpha' A\beta \tag{27}$$

in the two sets of n scalar variables

$$\alpha' = (x_1, x_2, \ldots, x_n) \qquad \text{and} \qquad \beta = \{y_1, y_2, \ldots, n\}$$

is called a *symmetric bilinear form* if the matrix $A = [a]_n^n$ of the bilinear form is symmetric. Observe that a symmetric bilinear form remains unchanged if the x's and the y's are interchanged. Since each index in (27) has the same range, we dispense with the summation symbols $\sum\sum$ which appear in the general bilinear form (13), and use the summation convention, it being understood that each repeated index takes the range 1, 2, . . . , n. If a particular value of n is under consideration, it will be specified.

Suppose that the x's and the y's of the symmetric bilinear form (26) are subjected to the same linear transformation:

$$x_i = c_{ij}\bar{x}_j \qquad \text{or} \qquad \alpha = C\bar{\alpha}, \tag{28}$$

and

$$y_i = c_{ij}\bar{y}_j \qquad \text{or} \qquad \beta = C\bar{\beta}. \tag{29}$$

Then from Theorem IV, or directly, we have

THEOREM VII. *If in a symmetric bilinear form $a_{ij}x_iy_j$ with matrix A we subject the x's and the y's to the same linear transformation with matrix C, we obtain a new symmetric bilinear form with matrix*

$$\bar{A} = C'AC. \tag{30}$$

If we take the transpose of both sides of the matric equation (30), we obtain

$$(\bar{A})' = (C'AC)' = C'A'(C')' = C'AC = \bar{A}, \tag{31}$$

since A is symmetric. We have proved

THEOREM VIII. *The symmetric bilinear form $a_{ij}x_iy_j$ remains symmetric if we subject the x's and the y's to the same linear transformation.*

Recall from Section 9–12 that if two matrices A and B satisfy the relation

(32) $$B = P'AP,$$

then A and B are congruent matrices, a special instance of equivalent matrices with the same rank. Therefore by Theorem XXVIII of Chapter 9, we have

THEOREM IX. *The symmetric bilinear form $a_{ij}x_iy_j$ of rank r can be reduced by cogredient nonsingular linear transformations to the normal form*

(33) $$a_1x_1y_1 + a_2x_2y_2 + \cdots + a_rx_ry_r.$$

10–4 Quadratic forms. If in a symmetric bilinear form $a_{ij}x_ix_j = \alpha'A\beta$ we set $\alpha = \beta$, we get a *quadratic form*

(34) $$f = a_{ij}x_ix_j = \alpha'A\alpha.$$

The matrix $A = [a]_n^n$ is the *matrix of the quadratic form*, and the determinant of A is called the *discriminant* of the quadratic form.

From Theorem VII above, or directly, we have

THEOREM X. *If in a quadratic form $a_{ij}x_ix_j$ with matrix A we subject the x's to the linear transformation with matrix C, we obtain a new quadratic form with matrix*

(35) $$\bar{A} = C'AC.$$

These follow the result:

THEOREM XI. *The rank of a quadratic form is an arithmetic invariant under a nonsingular linear transformation of its variables.*

Taking the determinant of both sides of (35), we get

$$|\bar{A}| = |C'| \cdot |A| \cdot |C| = |C|^2 \cdot |A|,$$

since $|C'| = |C|$. We have proved

THEOREM XII. *The discriminant of a quadratic form is a relative invariant of weight 2 with respect to a general nonsingular linear transformation of its variables.*

If the linear transformation of the variables

$$x_i = c_{ij}\bar{x}_j \qquad \text{or} \qquad \alpha = C\bar{\alpha}$$

is an orthogonal transformation, then $|C|^2 = 1$, and we have

THEOREM XIII. *The discriminant of a quadratic form is an absolute invariant under an orthogonal transformation of its variables.*

From Theorem X above and Theorem XXVIII of Chapter 9, we have

THEOREM XIV. *A quadratic form of rank r can be reduced by a nonsingular linear transformation to the normal form*

$$a_1x_1^2 + a_2x_2^2 + \cdots + a_rx_r^2. \tag{36}$$

It should be clear that the reduction of a quadratic form to its normal form (36) is essentially the reduction of the symmetric matrix $A = [a]_n^n$ of the quadratic form to its canonical form under congruent transformations.

In Section 9–7, we stated that when we deal with the equivalent transformation $PAQ = B$ we assume that all of the elements of P, A, Q, and therefore of B, belong to the same field **F**. The same understanding applies to the congruent transformation $P'AP = B$. Likewise, a similar understanding applies when we speak of the equivalence of forms. Thus we could more lengthily state the equivalence of two bilinear forms as follows: A bilinear form f in the variables $x_1, \ldots, x_m; y_1, \ldots, y_n$ with coefficients in a field **F** is equivalent to the bilinear form $\bar{f}$ if and only if f becomes $\bar{f}$ when the x's are subjected to a nonsingular linear transformation with matrix P and the y's are subjected to a nonsingular linear transformation with matrix Q, the elements of P and Q being in the same field **F**. In an analogous manner, we could say that two quadratic forms f and $\bar{f}$ are equivalent in a field **F** if and only if it is possible to pass from f to $\bar{f}$ by means of a nonsingular linear transformation on the x's, the coefficients of this transformation being in the field **F**. In this respect equivalence of quadratic (and symmetric bilinear) forms is more restrictive than the equivalence of the matrices. For two quadratic forms (or two symmetric bilinear forms) are equivalent if and only if their matrices are congruent.

In line with our general understanding that in a particular setting, problem, or theorem with which we are working, some given arbitrary field of scalar elements is assumed, we have not stated over and over "for a given field **F**." However, since we shall next consider a quadratic form in the particular complex and real fields, it seems desirable to restate Theorem XIV in the more emphatic form

THEOREM XIV. *A quadratic form in n variables and of rank r ($r \leq n$), with coefficients in a given field* **F**, *can be reduced by a nonsingular linear transformation with coefficients in* **F** *to the normal form*

$$a_1x_1^2 + a_2x_2^2 + \cdots + a_rx_r^2, \tag{37}$$

where the a's are nonzero elements in **F**.

In general, some of the coefficients in (37) are positive and some negative. When the underlying field is the field **C** of all complex numbers, we may let $z_i = (a_i)^{\frac{1}{2}}x_i$, and thereby transform (37) to

$$z_1^2 + z_2^2 + \cdots + z_r^2. \tag{38}$$

So we have

THEOREM XV. *A quadratic form of rank r with coefficients in the field* **C** *of complex numbers can be reduced by a nonsingular linear transformation to the normal form*

$$z_1^2 + z_2^2 + \cdots z_r^2.$$

EXERCISES

Find the symmetric matrix A associated with each of the following quadratic forms. Reduce each quadratic form to the normal form (37) by finding a nonsingular matrix P such that $P'AP$ is diagonal.

1. $2x^2 + xy$.
2. $2x^2 - 3y^2 + 2z^2 + 2xy + 11xz + 18yz$.

10–5 Real quadratic forms. If all the coefficients a_{ij} of the quadratic form $a_{ij}x_ix_j$ are in the field **R** of all real numbers, we call it a *real quadratic form*. Also, a nonsingular linear transformation is said to be *real* if all the elements of its matrix are in the field **R** of all real numbers. As we noted at the end of Section 9–12, only rational operations are involved in the reduction of a symmetric matrix to its canonical form under congruent transformations. Consequently, only rational operations are entailed in the reduction of a quadratic form of rank r to the normal form (37). Since rational operations performed on real numbers produce real numbers, we have

THEOREM XVI. *A real quadratic form in n variables and of rank r ($r \leq n$) can be reduced by a real nonsingular linear transformation to the normal form*

$$a_1x_1^2 + a_2x_2^2 + \cdots + a_rx_r^2, \tag{39}$$

where the a's are nonzero real constants.

An examination of the reduction of a symmetric matrix to its canonical form under congruent transformations (Section 9–12) shows that there are many different ways of effecting that reduction. For example, in step (*iii*) of that reduction there is considerable choice as to which nonzero a_{ij} is selected to become the first nonzero element b_{11}. Consequently the values of the a's in the normal form (39) may be different for different reductions. It is noteworthy that the signs of the coefficients, apart from the order in which they occur, do not depend on the particular reduction used. We give without proof

THEOREM XVII.* *If a real quadratic form of rank r is reduced to the normal form*

$$a_1x_1^2 + a_2x_2^2 + \cdots + a_rx_r^2, \tag{40}$$

then the numbers p of positive terms and r − p of negative terms will be the same no matter what real singular transformations are used.

This important theorem was discovered independently by Sylvester and Jacobi; it is commonly called Sylvester's law of inertia of quadratic forms. By Sylvester's law of inertia there is associated with a real quadratic form under real nonsingular linear transformations the arithmetic invariant p, in addition to arithmetic invariant r, the rank of the quadratic forms. Sometimes a third arithmetic invariant is used, namely the difference between the number of positive and the number of negative terms; this is called the *signature s* of the quadratic form:

$$s = p - (r - p) = 2p - r,$$

from which

$$p = \frac{r + s}{2}. \tag{41}$$

With a slight adaptation of the proof for Theorem XV above we can establish

THEOREM XVIII. *A real quadratic form of rank r and signature s can be reduced by a real nonsingular linear transformation to the normal form*

$$x_1^2 + x_2^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_r^2, \tag{42}$$

where $p = \frac{1}{2}(r + s)$.

* See H. W. Turnbull and A. C. Aitken, *An Introduction to the Theory of Canonical Matrices*, p. 89.

When the normal form (42) of a quadratic has no negative terms and has maximum rank n, so that $p = n$, the quadratic is said to be *positive definite*. When $p = 0$ and $r = n$, the canonical form has only negative terms and is said to be *negative definite*. When $p = r$, $r < n$, the quadratic form is called *non-negative definite of rank* r.

An important theorem in matrix theory * is

THEOREM XIX. *If A is a real symmetric matrix, there exists a real orthogonal matrix P such that $P'AP$ is a diagonal matrix whose diagonal elements are the characteristic roots of A:*

$$P'AP = [\lambda_i \delta_{ij}].$$

From this Theorem and Theorem X above there follows

THEOREM XX. *If $f = a_{ij}x_ix_j$ is a real quadratic form, there exists a real orthogonal transformation that reduces f to the canonical form*

$$\lambda_1 x_1^2 + \lambda_2 x_2^2 + \cdots + \lambda_n x_n^2,$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the characteristic roots of the symmetric matrix A of the quadratic form.

In Section 8–7 we have in fact made use of this theorem and the procedure based upon it. The illustrated examples in that section amply illustrate Theorem XX. It is noteworthy that there are any number of matrices which reduce a symmetric matrix A to diagonal form, and consequently a quadratic form to the algebraic sum of squares, but for a real symmetric matrix there is essentially only one real orthogonal matrix which effects that reduction, and that is the modal matrix of A.

---

* For proof see C. C. MacDuffee, *Vectors and Matrices*, p. 170.

CHAPTER 11

SOME APPLICATIONS OF MATRIX ALGEBRA

The purpose of this chapter is to indicate briefly some of the more elementary applications of those phases of matrix algebra considered in the preceding six chapters. Inherent in such an attempt as this is the problem of student unfamiliarity with technical concepts and terminology attendant to any particular application. We shall not endeavor to explain specialized technical terms in the brief space available, but we shall choose simple applications in which the number of technical terms is held to a minimum, and relate each application to one or more references from which additional information may be obtained. For more complete information on the references given see the Bibliography at the end of the book.

11–1 Finding the solution of n linear nonhomogeneous equations in n unknowns. We saw in Section 6–6 that the equations

(1) $$a_{ij}x_j = b_i$$

or

(2) $$A\delta = \beta,$$

where A is the nonsingular coefficient matrix of order n, $\delta = \{x_1, x_2, \ldots, x_n\}$, $\beta = \{b_1, b_2, \ldots, b_n\}$, have their vector solution δ given by

(3) $$\delta = A^{-1}\beta.$$

While the theoretical solution (3) is simply obtained, the numerical calculation of A^{-1} by direct elementary means for a large value of n is a problem of enormous magnitude. Fortunately, there is available a body of theory and procedures for reducing these apparent difficulties. The center of this special theory is the formula (15) for A^{-1} developed in Section 8–4, namely,*

* For interesting presentations of this theory see M. D. Bingham, "A New Method for Obtaining the Inverse Matrix," *Journal of the American Statistical Association*, **36** (1941), pp. 530–534, and H. Hotelling, "Some New Methods in Matrix Calculation," *Annals of Mathematical Statistics*, **14** (1943), pp. 1–34.

$$(4)\quad A^{-1} = -\frac{1}{(-1)^n p_n}[A^{n-1} - p_1 A^{n-2} + p_2 A^{n-3} + \cdots + (-1)^{n-1} p_{n-1} I].$$

To compute A^{-1} by (4) one has to know the values of the coefficients $p_1, p_2, \ldots, p_{n-1}, p_n$ in the characteristic function of A.

We have become acquainted with the trace $a_{ii} = a_{11} + a_{22} + \cdots + a_{nn}$ of the matrix $A = [a]_n^n$. Let $s_1, s_2, \ldots, s_n$ be scalars defined by

$$(5)\qquad s_1 = \text{tr}(A),\quad s_2 = \text{tr}(A^2),\quad \ldots,\quad s_n = \text{tr}(A^n);$$

that is, s_h is the trace of the hth power of the given matrix A. The scalars $p_1, p_2, \ldots, p_n$ can be computed successively by the recurrence formulas:

$$(6)\quad p_1 = s_1,\quad p_2 = \tfrac{1}{2}(p_1 s_1 - s_2),\quad p_3 = \tfrac{1}{3}(p_2 s_1 - p_1 s_2 + s_3),\quad \ldots,\quad p_n = \frac{1}{n}[p_{n-1} - p_{n-2}s_2 + \cdots + (-1)^{n-1}s_n].$$

We see then that the inverse matrix A^{-1} can be calculated by the following steps:

(*i*) Compute the first $n - 1$ powers $A, A^2, \ldots, A^{n-1}$ of A and the diagonal elements of A^n;

(*ii*) Calculate the scalars $s_1, s_2, \ldots, s_n$ by use of (5);

(*iii*) Determine $p_1, p_2, \ldots, p_n$ by use of (6);

(*iv*) Finally, calculate A^{-1} by substituting the values of the p's in (4).

In properly equipped computing centers punched card methods are used to calculate the powers of A, and the rest of the calculations for finding A^{-1} by use of (4) can be made with calculating machines.

If $|A| = 0$, but the n equations are still consistent with rank of $A = r$, their "indeterminate solution" is given by the formula

$$(7)\qquad \delta_1 = -A_{11}^{-1}A_{12}\delta_2 + A_{11}^{-1}\beta_1,$$

developed in Section 9–11. Again, the main work in obtaining the solution in a given numerical case rests primarily on the calculation of the inverse of the matrix A_{11}; that is, the determination of the inverse of the leading square submatrix of A whose rank is r. Hence the procedure outlined above for the calculation of the inverse of a matrix is applicable and useful.

The recent development of modern high-speed electronic computers, which are capable of multiplying large matrices in a few minutes and which make it humanly possible to calculate the inverse of a matrix of large order, has brought about a renewed interest in numerical procedures in connection with matrix algebra. For a discussion of such problems see the paper by John von Neumann and H. H. Goldstine, "Numerical inverting of matrices of high order," *Bulletin of the American Mathematical Society*, **53** (1947), pp. 1021–1099.

11–2 Finding the solution of m linear homogeneous equations in n unknowns. Section 9–10 gives an organized theoretical procedure for the solution of a system of m linear homogeneous equations in n unknowns. After the rank r of the coefficient matrix A is determined and the equations are rearranged so that the leading square submatrix, A_{11}, of A is of rank r = rank of A, the "indeterminate solution" of

(8) $$a_{ij}x_j = 0, \qquad \text{or} \qquad A\delta = O$$

is given by the formula

(9) $$\delta_1 = -A_{11}^{-1}A_{12}\delta_2,$$

which we developed in Section 9–10. Other than the relatively easy multiplication of matrices by vectors, the main work involved in this solution is the calculation of the inverse of the submatrix A_{11}. Clearly the procedure outlined in the preceding section may be helpful.

Systems of homogeneous linear equations appear rather infrequently in conventional college algebra, and consequently the student often has the feeling that such systems of equations have no place in applied mathematics. But that is not so; problems in mechanics and electric circuit theory often lead to such systems; H. Sohon in Chapter 3 (Dimensional Analysis) of his book *Engineering Mathematics* gives some interesting applications of matrix algebra to applied problems which lead to systems of linear homogeneous equations.

11–3 Determination of principal axes of inertia. Professor F. D. Murnaghan on page 340 of his book *Analytic Geometry*,* in speaking of the process of diagonalizing a square matrix A_2 of the second order, says, "The process by which A_2 is diagonalized, that is, reduced to a diagonal matrix, merely by changing the reference frame, is important not only for the present application but in many applications to engineering and physics. When you come to study principal axes

* Prentice-Hall Inc., 1946.

of inertia in engineering, and normal vibrations in physics, you will have to understand it thoroughly. It also occupies a dominant position in wave mechanics and quantum theory."

Let x_1, x_2, x_3 be the coordinates of any particle m of a given mass M relative to the OX_1, OX_2, OX_3 coordinate axes, and let k_1, k_2, k_3 be the direction cosines of the line L for which the moment of inertia of the mass M is I. In analytical mechanics it is shown that

$$(10) \qquad I = ak_1^2 + bk_2^2 + ck_3^2 - 2\,dk_2k_3 - 2ek_3k_1 - 2fk_1k_2,$$

where a, b, and c are the moments of inertia of M for the X_1, X_2, and X_3 axes, and d, e, and f are the products of inertia for the X_2X_3, X_3X_1, and X_1X_2 coordinate planes. Let an arbitrary length $OP = \rho$ be laid off on the line L. Then $x_1 = \rho k_1$, $x_2 = \rho k_2$, $x_3 = \rho k_3$. Substituting these values in (10), we get

$$(11) \qquad ax_1^2 + bx_2^2 + cx_3^2 - 2\,dx_2x_3 - 2ex_3x_1 - 2fx_1x_2 = \rho^2 I.$$

Equation (11) is the equation of a quadric surface, provided ρ is chosen for the different lines through the origin so that $\rho^2 I$ is constant. That is, if on every line through the origin a length $OP = d = h/\sqrt{I}$ is laid off, then the point P will lie on the quadric surface

$$(12) \qquad ax_1^2 + bx_2^2 + cx_3^2 - 2\,dx_2x_3 - 2ex_3x_1 - 2fx_1x_2 = h^2.$$

As moments of inertia are inherently positive quantities, the radius vectors of the surface (12) are all real, and that surface is an ellipsoid; it is called the *ellipsoid of inertia*, its principal axes are called the *principal axes of inertia*, and the moments of inertia for these axes are called the *principal moments of inertia*. It should be clear that the problem of determination of the principal axes of inertia is identical to the problem of diagonalizing the coefficient matrix of the quadric form (12), which procedure we considered in Section 8–7, and again in Chapter 10. In this connection see Frazer, Duncan, and Collar's *Elementary Matrices*, page 257.

11–4 Matrices in statistics. Matrices whose scalar elements are sums of products (or correlations) occur frequently in statistics, and workers in multivariate statistical analysis have made significant contributions to both the theory and the applications of matrix algebra. Many of these contributions are concerned with special methods for calculating the inverse of a matrix, and for finding the characteristic vectors of a matrix. The interested reader will find a considerable number of references as to the use of matrices in statistics in the articles by Professor Harold Hotelling, "Some New Methods in Matrix Calculation," *Annals of Mathematical Statistics*, **14** (March 1943), pp. 1–34; and "Practical Problems of Matrix Calculation," *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press (1949), pp. 275–293.

11–5 Matrix algebra in multiple factor analysis. The branch of psychometrics called multiple factor analysis is concerned mainly with three problems: (1) the determination of the smallest number of independent abilities that must be postulated in order to account for a given table of intercorrelations; (2) the determination of how much each independent ability is represented by each test; and (3) the setting up of regression equations by means of which an individual's amount of any primary ability can be estimated from scores on tests that depend upon that ability.

This theory of multiple factor analysis is mathematical in nature, and its chief mathematical tool is matrix algebra. Typical of the problems concerned with matrix algebra in such theory is the factorization of the symmetric matrix of correlations

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

as treated by Professor L. L. Thurstone in Chapters I and II of his book, *The Vectors of Mind.* Also, in multiple factor analysis considerable attention is given to the problem of diagonalizing matrices of correlations; for a treatment of this problem see Chapter XX of another book by Professor Thurstone, *Multiple Factor Analysis.* For additional references in this connection see the books on multiple factor analysis listed in the Bibliography.

11–6 Matrix and tensor calculus. Our work in this brief book has been limited to the algebra of matrices. Many important problems in engineering and physical science which are solved with the aid of matrix theory use not only the *algebra* of matrices but also the *calculus* of matrices. Basic in this connection is the book by Frazer, Duncan, and Collar, and the articles in the *Philosophical Magazine* by Duncan and Collar; all of these are listed in the Bibliography.

Closely related to matrix calculus is the subject of tensor calculus. A recent book which relates elementary matrix algebra as presented in the present book with the more sophisticated discipline of tensor analysis of networks is *Matrix Analysis of Electric Networks* by P. Le Corbeiller, Harvard University Press (1950).

Tensors of the second order play a particularly important role in numerous applications. Such a tensor is a matrix which obeys a specified law of transformation when the variable coordinates undergo a given transformation. In the matrix of a quadratic form we have an instance of that type of tensor. If we interpret the variables x_i of the quadratic form $a_{ij}x_ix_j$ (Section 10–4) as the coordinates of a reference system, then the congruent transformation induced on the matrix A of the quadratic form, when the variable coordinates are subjected to a linear transformation, is a tensor transformation. Hence the matrix A under the stated circumstances is a tensor. In general, a tensor of the second order is a matrix that is restricted with respect to a coordinate reference system. For a clear presentation of the relation of matrix and tensor algebra see H. Jeffreys and B. S. Jeffreys, *Methods of Mathematical Physics*, Cambridge University Press (1950), Chapters 3 and 4.

The algebra of matrices constitutes at the same time an algebra for tensors of the second order. For tensors of the second order, as for matrices in general, the properties of symmetry, skew symmetry, and diagonalization are of basic significance. Typical of work in applied science which uses tensors as the medium of expression, and which is based largely upon matrix algebra (especially the Cayley-Hamilton Theorem), is a paper by W. Prager, "Strain hardening under combined stresses," *Journal of Applied Physics*, **16** (1945), pp. 837–840.

APPENDIX

APPENDIX 1. *The determinant of the product of two square matrices is equal to the product of their determinants:* $|AB| = |A| \cdot |B|$.

In Exercise 8, page 63, Exercise 10, page 65, and Exercise 11, page 65, we have indicated how this theorem may be proved by the use of appropriate identities of vector algebra. Such proof may not appeal to some readers. Many other proofs appear in the literature on matrix algebra.

A proof commonly given makes use of Laplace's development of a determinant, considered in Section 9-3. Thus, for square matrices of the third order,

$$A = [a]_3^3 \qquad \text{and} \qquad B = [b]_3^3,$$

it follows from Laplace's development that the determinant

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & 0 & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & 0 & 0 \\ -1 & 0 & 0 & b_{11} & b_{12} & b_{13} \\ 0 & -1 & 0 & b_{21} & b_{22} & b_{23} \\ 0 & 0 & -1 & b_{31} & b_{32} & b_{33} \end{vmatrix}$$

is equal to $|A| \cdot |B|$, and also is equal to $|C|$ where $C = AB$. [See J. M. H. Olmstead, *Solid Analytic Geometry*, New York (1947), pp. 211–212.]

Another proof of this theorem makes use of elementary transformations of matrices (see G. Birkhoff and S. MacLane, *A Survey of Modern Algebra*, pp. 288–289), and still another uses the classical method of Weierstrass (see C. C. MacDuffee, *Vectors and Matrices*, p. 56).

APPENDIX 2. *If the columns of a square matrix A are linearly dependent, then* $|A| = 0$; *and conversely, if the determinant of a square matrix is zero, then the columns of that square matrix are linearly dependent.*

We use this theorem in Section 8–6, but do not prove it until we get to Section 9–5 (Theorem II and Theorem VII). However, the relative order of Section 8–6 and Section 9–5 is immaterial from a logical viewpoint. It is simply a matter of taste and convenience that the given order is followed.

BIBLIOGRAPHY

Vector and Matrix Algebra:

AITKEN, A. C., *Determinants and Matrices*, London, Oliver and Boyd, 1939.

ALBERT, A. A., *Introduction to Algebraic Theories*, Chicago, University of Chicago Press, 1941.

BIRKHOFF, G. AND MACLANE, S., *A Survey of Modern Algebra*, New York, Macmillan, 1941.

CULLIS, C. E., *Matrices and Determinoids*, Cambridge, Cambridge University Press, Vol. **I**, 1913; Vol. **II**, 1918; Vol. **III**, 1925.

DICKSON, L. E., *Modern Algebraic Theories*, Chicago, B. H. Sanborn, 1926.

DRESDEN, A., *Solid Analytical Geometry and Determinants*, New York, John Wiley and Sons, 1930.

FERRAR, W. L., *Algebra*, Oxford, Clarendon Press, 1941.

HALMOS, P. R., *Finite Dimensional Vector Spaces*, Princeton, Princeton University Press, 1942.

JOHNSTON, F. E., *Modern Algebra*, Washington, The Author, 1944.

LITTLEWOOD, D. E., *The Theory of Group Characters and Matrix Representations of Groups*, Oxford, Clarendon Press, 1940.

MACDUFFEE, C. C., *The Theory of Matrices*, New York, Chelsea, 1946.

——, *Vectors and Matrices*, Carus Mathematical Monograph No. 7, Menasha, Wisconsin, The Mathematical Association of America, 1943.

MURNAGHAN, F. D., *Analytic Geometry*, New York, Prentice-Hall, 1946.

TURNBULL, H. W. AND AITKEN, A. C., *An Introduction to the Theory of Canonical Matrices*, London, Blackie and Son, 1929.

TURNBULL, H. W., *The Theory of Determinants, Matrices, and Invariants*, London, Blackie and Son, 1929.

WEDDERBURN, J. H. M., *Lectures on Matrices*, New York, American Mathematical Society, 1934.

Applications of Matrix Algebra to Physics and Engineering:

BURINGTON, R. S., AND TORRANCE, C. C., *Higher Mathematics with Applications to Science and Engineering*, New York, McGraw-Hill, 1939; especially Chapter VI.

CORBEILLER, P. LE, *Matrix Analysis of Electric Networks*, Harvard University Press, 1950.

COWLEY, W. L., *Advanced Practical Mathematics*, London, Pitman and Sons, 1934; especially Chapter XI.

DUNCAN, W. J. AND COLLAR, A. R., "A Method for the Solution of Oscillation Problems by Matrices," *Philosophical Magazine*, Vol. **17**, 1934, pp. 865–909.

———, "Matrices Applied to the Motions of Damped Systems," *Philosophical Magazine*, Vol. **19**, 1935, pp. 197–219.

FRAZER, R. A, DUNCAN, W. J., AND COLLAR, A. R., *Elementary Matrices and Some Applications to Dynamics and Differential Equations*, New York, Macmillan, 1946.

GUILLEMIN, E. A., *The Mathematics of Circuit Analysis*, New York, Wiley, 1949; especially Chapters 1–4.

JEFFREYS, H. AND JEFFREYS, B. S., *Methods of Mathematical Physics*, Cambridge, Cambridge University Press, 1950; especially Chapters 3 and 4.

KARMAN, T. VON AND BIOT, M. A., *Mathematical Methods in Engineering*, New York, McGraw-Hill, 1940; especially Chapters V and VI.

MICHAL, A. D., *Matrix and Tensor Calculus with Applications to Mechanics, Elasticity, and Aeronautics*, New York, Wiley, 1946.

PIPES, L. A., *Applied Mathematics for Engineers and Physicists*, New York, McGraw-Hill, 1946.

SCHMEIDLER, W., *Vortrage uber Determinanten und Matrizen mit Anwendungen in Physik und Technik*, Berlin, Akademie-Verlag, 1949.

SOHON, H., *Engineering Mathematics*, New York, D. Van Nostrand Co., 1944.

ZURMÜHL, R., *Matrizen: Eine Darstellung fur Ingenieure*, Berlin, Springer-Verlag, 1950.

Applications of Matrix Algebra to Multiple Factor Analysis and Statistics:

HOTELLING, H., "Some New Methods in Matrix Calculation," *The Annals of Mathematical Statistics*, Vol. **14**, 1943, pp. 1–34.

———, "Practical Problems of Matrix Calculation," *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1949, pp. 275–293.

THOMSON, G. H., *The Factorial Analysis of Human Ability*, New York, Houghton Mifflin, 1939.

THURSTONE, L. L., *The Vectors of Mind*, Chicago, University of Chicago Press, 1935.

———, *Multiple Factor Analysis*, Chicago, University of Chicago Press, 1947.

INDEX

ANSWERS TO EXERCISES

Chapter 2

SECTION 2–1

1. Mag = 126^-.
   Direction = 11.5° approx. from the 80-lb force.
2. 25^- miles.
   25.1° approx. south of east.

SECTION 2–3

1. (5, 3); (10, 1); (3, 3).
4. (3, 7).
7. (a) (2, 4).
   (b) (2, 4); (4, 8).
   (c) No.

SECTION 2–4

1. Mag $\gamma = \sqrt{149}$.
2. Mag $\alpha = 5$.
3. $\alpha\beta = (2, 1)$.
4. (a) $(-2, 10)$.
   (b) $(-7, 16)$.
   (c) $(8, -2)$.
   (d) $(1, \frac{8}{3})$.

SECTION 2–5

1. (5, 3, 1); (3, 3, 9).
2. (5, 8, 2).
3. (2, 4, 2); others are (4, 8, 4); (6, 12, 6); etc.
5. $\alpha + \beta = (-2, 10, 3)$; $\alpha + 2\beta = (-7, 16, 5)$; $\beta - \gamma = (-10, 8, 1)$;
   $\frac{1}{3}(\alpha + \beta + \gamma) = (1, \frac{8}{3}, \frac{4}{3})$.
6. $\alpha + \beta = [(a_1 + b_1)\epsilon_1 + (a_2 + b_2)\epsilon_2 + (a_3 + b_3)\epsilon_3]$;
   $\alpha - \beta = [(a_1 - b_1)\epsilon_1 + (a_2 - b_2)\epsilon_2 + (a_3 - b_3)\epsilon_3]$.

SECTION 2–7

5. $\theta = \text{arc cos}\,\dfrac{1}{\sqrt{2}} = 45°$.
11. $|\gamma| = \sqrt{269}$;
    direction cos $= \left(\dfrac{6}{\sqrt{269}}, \dfrac{8}{\sqrt{269}}, \dfrac{13}{\sqrt{269}}\right)$.

Chapter 3

SECTION 3–3

2. For $n = 2$, $\dfrac{x_1 - a_1}{x_2 - a_2} = \dfrac{b_1 - a_1}{b_2 - a_2}$; for $n = 3$, $x_1 - a_1 = t(b_1 - a_1)$,
   $x_2 - a_2 = t(b_2 - a_2)$,
   $x_3 - a_3 = t(b_3 - a_3)$.
4. $\rho = (5, \frac{36}{5}, \frac{47}{5})$.

Section 3–5

1. (*i*) Dependent; $r = 3$, $s = 1$.
   (*ii*) Independent.
   (*iii*) Dependent; $r = 4$, $s = -1$.
   (*iv*) Independent.

2. (*i*) $k = -2$.
   (*ii*) For no value of k.

3. (*i*) Dependent; $r = -2$, $s = -3$.
   (*ii*) Dependent; $r = -2$, $s = -3$.
   (*iii*) Dependent; $r = 2$, $s = -1$.
   (*iv*) Independent.

4. $r = 1$, $s = 2$, $t = 3$.

Chapter 4

Section 4–6

1. (*i*) Independent.
   (*ii*) Dependent; $r = 4$, $s = 1$.

2. (*i*) Independent.
   (*ii*) Dependent; $r = 2$, $s = 1$.

3. Let the five four-dimensional vectors be denoted by α, β, γ, δ, ϵ and assume that α, β, γ, δ are linearly independent. Then we want to find scalars s_1, s_2, s_3, s_4 such that $\epsilon = s_1\alpha + s_2\beta + s_3\gamma + s_4\delta$. The desired results are

$$s_1 = \frac{|\epsilon\beta\gamma\delta|}{|\alpha\beta\gamma\delta|}, \quad s_2 = \frac{|\alpha\epsilon\gamma\delta|}{|\alpha\beta\gamma\delta|}, \quad s_3 = \frac{|\alpha\beta\epsilon\delta|}{|\alpha\beta\gamma\delta|}, \quad s_4 = \frac{|\alpha\beta\gamma\epsilon|}{|\alpha\beta\gamma\delta|}.$$

4. $s_1 = 1$, $s_2 = 2$, $s_3 = -1$, $s_4 = 3$.

Chapter 5

Section 5–2

2. $T + U = \begin{bmatrix} 7 & 4 & 6 \\ 8 & 9 & 12 \\ 14 & 16 & 17 \end{bmatrix}$.

Section 5–3

1. $2R - S = \begin{bmatrix} 3 & 13 \\ 1 & 5 \end{bmatrix}$.

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

Section 5–6

1. $SR = \begin{bmatrix} 22 & 24 \\ 22 & 69 \end{bmatrix}$, $|SR| = 990 = |RS|$.

2. $(QR)S = \begin{bmatrix} 40 & 2 \\ 155 & 82 \end{bmatrix}$, $Q(RS) = \begin{bmatrix} 40 & 2 \\ 155 & 82 \end{bmatrix}$.

Section 5–7

1. $TU = \begin{bmatrix} 35 & 34 & 39 \\ 86 & 76 & 90 \\ 137 & 118 & 141 \end{bmatrix}$. $|TU| = 0 = |UT|$; $|T| = 0$; $|U| = -64$.

$UT = \begin{bmatrix} 35 & 46 & 57 \\ 62 & 76 & 90 \\ 95 & 118 & 141 \end{bmatrix}$.

2. $AB = \begin{bmatrix} 1 & 7 \\ 28 & 14 \end{bmatrix}$; $\quad BA = \begin{bmatrix} 17 & -17 & 1 \\ -12 & -6 & 3 \\ 19 & -33 & 4 \end{bmatrix}$.

3. 2.

4. $\begin{bmatrix} 6 & -4 & 2 \\ 15 & -10 & 5 \\ 18 & -12 & 6 \end{bmatrix}$.

SECTION 5–10

1. $a_{11}x_1x_1 + a_{12}x_1x_2 + a_{13}x_1x_3 + a_{21}x_2x_1 + a_{22}x_2x_2 + a_{23}x_2x_3 + a_{31}x_3x_1 + a_{32}x_3x_2 + a_{33}x_3x_3$.

2. $a_{11}x_1 + a_{21}x_2 + a_{31}x_3 + a_{41}x_4$,
$a_{12}x_1 + a_{22}x_2 + a_{32}x_3 + a_{42}x_4$,
$a_{13}x_1 + a_{23}x_2 + a_{33}x_3 + a_{43}x_4$,
$a_{14}x_1 + a_{24}x_2 + a_{34}x_3 + a_{44}x_4$.

SECTION 5–13

1. $A^2 = \begin{bmatrix} 1 & 40 \\ -20 & 41 \end{bmatrix}$; $\quad A^2A = AA^2 = \begin{bmatrix} -77 & 284 \\ -142 & 207 \end{bmatrix}$.

3. $A^2 + AB + BA + B^2$.

4. $A^4 + AB^3 + A^2B^2 + A^3B + BA^3 + B^2A^2 + B^3A + B^4 + ABA^2 + A^2BA + B^2AB + BA^2B + BABA + ABAB + BAB^2 + AB^2A$.

Chapter 6

SECTION 6–4

1. $\tilde{A} = \begin{bmatrix} -2 & -4 \\ -1 & 3 \end{bmatrix}$, $\quad A^{-1} = -\frac{1}{10}\begin{bmatrix} -2 & -4 \\ -1 & 3 \end{bmatrix}$.

2. $\{x, y\} = \{5, -2\}$.

3. $\{x, y\} = \{4, 5\}$.

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

7. $\{x_1, x_2, x_3\} = \{1, 2, 3\}$.

8. $\{x_1, x_2, x_3\} = \{-2, 0, 6\}$.

SECTION 6–8

1. (*i*) $\left(-\frac{3}{4}, \frac{3}{4}, 0, \frac{1}{4}\right)$.

(*ii*) $(2, -2, 1, 3)$.

2. (*i*) $A^{-1} = \begin{bmatrix} -9 & 1 & -2 & 4 \\ -5 & 1 & -1 & 3 \\ 15 & -2 & 3 & -7 \\ -1 & 0 & 0 & 1 \end{bmatrix}$.

(*ii*) $|A| = 0$; so A is singular and has no inverse.

Chapter 8

SECTION 8–3

1. 1, −1. 2. 2, −2. 3. 10, −2, −3

6. (i) $A^{-1} = \frac{1}{10}\begin{bmatrix} 2 & 4 \\ 1 & -3 \end{bmatrix}$. (ii) $A^{-1} = -\frac{1}{18}\begin{bmatrix} -1 & 1 \\ 2 & -5 \end{bmatrix}$.

SECTION 8–4

1. Characteristic roots are 1, 3, −2. $A^{-1} = -\frac{1}{6}\begin{bmatrix} -4 & 7 & -5 \\ 2 & -5 & 1 \\ 2 & -8 & 4 \end{bmatrix}$.

2. $A^{-1} = \frac{1}{36}\begin{bmatrix} 8 & -4 & 1 \\ -4 & 8 & 1 \\ 4 & 4 & 11 \end{bmatrix}$. Characteristic roots are 3, 3, 12.

3. Characteristic roots are 2, 0, −1.
   A^{-1} does not exist, for $|A| = 0$.

SECTION 8–5

1. (i) $16\bar{x}^2 + 9\bar{y}^2 = 144$. (iii) $\bar{x}^2 - \bar{y}^2 = 24$.
   (ii) $-15\bar{x}^2 + 10\bar{y}^2 = 1$. (iv) $\bar{y}^2 = 6$.
2. $-\bar{x}^2 + 7\bar{y}^2 - 5\bar{z}^2 = -5$. 3. $\bar{x}^2 + 2\bar{y}^2 + 3\bar{z}^2 = 6$.

SECTION 8–6

1. The characteristic roots are 1, 3, −2; and $B = \begin{bmatrix} -1 & 1 & -11 \\ 1 & 1 & -1 \\ 1 & 1 & 14 \end{bmatrix}$.

2. The characteristic roots are 2, 0, −1.

SECTION 8–7

1. The characteristic roots are 9, 9, −9.
2. The characteristic roots are 2, −1, −1.

Chapter 9

SECTION 9–1

1. None. 2. None. 3. $r_3 = 2r_1 + r_2$.

SECTION 9–4

1. $r = 2$. 2. $r = 1$. 3. $r = 1$.

SECTION 9–5

1. $\alpha_3 = \{4, 2, 1\}$, rank $A = 2$.
2. $\alpha_3 = \{1, -4, 11, -19\}$, rank $A = 2$.

SECTION 9–6

1. Rank $A = 2$; $\alpha_3 = -\alpha_1 + 2\alpha_2$.
2. Rank $A = 3$; $\alpha_4 = \alpha_1 + 2\alpha_2 + 3\alpha_3$.
3. Rank $A = 1$; $\alpha_2 = -4\alpha_1$.

SECTION 9–7

3. Rank $A = 2$. 4. Rank $A = 3$. 5. Rank $A = 2$.

SECTION 9–9

6. (*i*) Since $r = n = 3$, there is no nontrivial solution.
(*ii*) Rank $A = 3$. For

$$A_{11} = \begin{bmatrix} 2 & -1 & -1 \\ 1 & 1 & 1 \\ 4 & 1 & 3 \end{bmatrix}, \quad \begin{bmatrix} x_1 \\ x_2 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} x_3.$$

SECTION 9–11

1. Rank A = rank $[A, \beta] = 2$.

$$\text{For } A_{11} = \begin{bmatrix} 2 & 3 \\ 4 & -1 \end{bmatrix}, \quad \begin{bmatrix} x_1 \\ x_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix} x_2 - \frac{1}{7} \begin{bmatrix} 4 \\ -5 \end{bmatrix}.$$

2. Rank A = rank $[A, \beta] = 3$.

$$\text{For } A_{11} = \begin{bmatrix} 2 & -3 & 4 \\ 1 & 2 & -1 \\ 3 & 1 & 2 \end{bmatrix}, \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{7} \begin{bmatrix} 16 \\ -29 \\ -28 \end{bmatrix} x_4 - \frac{1}{7} \begin{bmatrix} -9 \\ 1 \\ 0 \end{bmatrix}.$$

SECTION 10–2

1. $A = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}$. For the reduction see the illustration at the end of Section 9–9.

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

3. $A = \begin{bmatrix} 2 & -1 & -6 \\ -3 & 0 & 3 \\ 1 & 5 & 19 \end{bmatrix}$.

SECTION 10–4

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

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