INTRODUCTION TO VECTORS AND TENSORS

Linear and Multilinear Algebra

Volume 1

Ray M. Bowen
Mechanical Engineering
Texas A&M University
College Station, Texas

and

C.-C. Wang
Mathematical Sciences
Rice University
Houston, Texas

Copyright Ray M. Bowen and C.-C. Wang
(ISBN 0-306-37508-7 (v. 1))

PREFACE

To Volume 1

This work represents our effort to present the basic concepts of vector and tensor analysis. Volume I begins with a brief discussion of algebraic structures followed by a rather detailed discussion of the algebra of vectors and tensors. Volume II begins with a discussion of Euclidean Manifolds which leads to a development of the analytical and geometrical aspects of vector and tensor fields. a discussion of general differentiable manifolds. We have not included a discussion of general differentiable manifolds. However, we have included a chapter on vector and tensor fields defined on Hypersurfaces in a Euclidean Manifold.

In preparing this two volume work our intention is to present to Engineering and Science students a modern introduction to vectors and tensors. Traditional courses on applied mathematics have emphasized problem solving techniques rather than the systematic development of concepts. As a result, it is possible for such courses to become terminal mathematics courses rather than courses which equip the student to develop his or her understanding further.

As Engineering students our courses on vectors and tensors were taught in the traditional way. We learned to identify vectors and tensors by formal transformation rules rather than by their common mathematical structure. The subject seemed to consist of nothing but a collection of mathematical manipulations of long equations decorated by a multitude of subscripts and superscripts. Prior to our applying vector and tensor analysis to our research area of modern continuum mechanics, we almost had to relearn the subject. Therefore, one of our objectives in writing this book is to make available a modern introductory textbook suitable for the first in-depth exposure to vectors and tensors. Because of our interest in applications, it is our hope that this book will aid students in their efforts to use vectors and tensors in applied areas.

The presentation of the basic mathematical concepts is, we hope, as clear and brief as possible without being overly abstract. Since we have written an introductory text, no attempt has been made to include every possible topic. The topics we have included tend to reflect our personal bias. We make no claim that there are not other introductory topics which could have been included.

Basically the text was designed in order that each volume could be used in a one-semester course. We feel Volume I is suitable for an introductory linear algebra course of one semester. Given this course, or an equivalent, Volume II is suitable for a one semester course on vector and tensor analysis. Many exercises are included in each volume. However, it is likely that teachers will wish to generate additional exercises. Several times during the preparation of this book we taught a one semester course to students with a very limited background in linear algebra and no background in tensor analysis. Typically these students were majoring in Engineering or one of the Physical Sciences. However, we occasionally had students from the Social Sciences. For this one semester course, we covered the material in Chapters 0, 3, 4, 5, 7 and 8 from Volume I and selected topics from Chapters 9, 10, and 11 from Volume 2. As to level, our classes have contained juniors,

seniors and graduate students. These students seemed to experience no unusual difficulty with the material.

It is a pleasure to acknowledge our indebtedness to our students for their help and forbearance. Also, we wish to thank the U. S. National Science Foundation for its support during the preparation of this work. We especially wish to express our appreciation for the patience and understanding of our wives and children during the extended period this work was in preparation.

Houston, Texas *R.M.B.*
C.-C.W.

CONTENTS

Vol. 1 Linear and Multilinear Algebra

.

Vol. 2 Vector and Tensor Analysis

PART III. VECTOR AND TENSOR ANALYSIS

PART 1

BASIC MATHEMATICS

Selected Reading for Part I

BIRKHOFF, G., and S. MACLANE, *A Survey of Modern Algebra*, 2nd ed., Macmillan, New York, 1953.

FRAZER, R. A., W. J. DUNCAN, and A. R. COLLAR, *Elementary Matrices*, Cambridge University Press, Cambridge, 1938.

HALMOS, P. R., *Naïve Set Theory*, Van Nostrand, New York, 1960.

KELLEY, J. L., Introduction to Modern Algebra, Van Nostrand, New York 1965.

MOSTOW, G. D., J. H. SAMPSON, and J. P. MEYER, *Fundamental Structures of Algebra*, McGraw-Hill, New York, 1963.

VAN DER WAERDEN, B. L., *Modern Algebra*, rev. English ed., 2 vols., Ungar, New York, 1949, 1950.

Chapter 0

ELEMENTARY MATRIX THEORY

When we introduce the various types of structures essential to the study of vectors and tensors, it is convenient in many cases to illustrate these structures by examples involving matrices. It is for this reason we are including a very brief introduction to matrix theory here. We shall not make any effort toward rigor in this chapter. In Chapter V we shall return to the subject of matrices and augment, in a more careful fashion, the material presented here.

An M by N *matrix* A is a rectangular array of real or complex numbers A_{ij} arranged in M rows and N columns. A matrix is often written

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdot & \cdot & \cdot & A_{1N} \\ A_{21} & A_{22} & \cdot & \cdot & \cdot & A_{2N} \\ \cdot & & & & & \\ \cdot & & & & & \\ A_{M1} & A_{M2} & \cdot & \cdot & \cdot & A_{MN} \end{bmatrix} \tag{0.1}$$

and the numbers A_{ij} are called the *elements* or *components* of A. The matrix A is called a *real* matrix or a *complex* matrix according to whether the components of A are real numbers or complex numbers.

A matrix of M rows and N columns is said to be of *order* M by N or $M \times N$. It is customary to enclose the array with brackets, parentheses or double straight lines. We shall adopt the notation in (0.1). The location of the indices is sometimes modified to the forms A^{ij}, $A^{i}{}_{j}$, or $A_{i}{}^{j}$. Throughout this chapter the placement of the indices is unimportant and shall always be written as in (0.1). The elements $A_{i1}, A_{i2}, ..., A_{iN}$ are the elements of the i^{th} row of A, and the elements $A_{1k}, A_{2k}, ..., A_{Nk}$ are the elements of the k^{th} column. The convention is that the first index denotes the row and the second the column.

A row matrix is a $1 \times N$ matrix, e.g.,

$$\begin{bmatrix} A_{11} & A_{12} & \cdot & \cdot & \cdot & A_{1N} \end{bmatrix}$$

while a column matrix is an $M \times 1$ matrix, e.g.,

$$\begin{bmatrix} A_{11} \\ A_{21} \\ . \\ . \\ . \\ A_{M1} \end{bmatrix}$$

The matrix A is often written simply

$$A = \left[A_{ij} \right] \tag{0.2}$$

A *square* matrix is an $N \times N$ matrix. In a square matrix A, the elements $A_{11}, A_{22}, ..., A_{NN}$ are its diagonal elements. The sum of the diagonal elements of a square matrix A is called the *trace* and is written tr A. Two matrices A and B are said to be *equal* if they are identical. That is, A and B have the same number of rows and the same number of columns and

$$A_{ij} = B_{ij}, \qquad i = 1, ..., N, \qquad j = 1, ..., M$$

A matrix, every element of which is zero, is called the *zero* matrix and is written simply 0.

If $A = \left[A_{ij} \right]$ and $B = \left[B_{ij} \right]$ are two $M \times N$ matrices, their *sum* (*difference*) is an $M \times N$ matrix $A + B$ $(A - B)$ whose elements are $A_{ij} + B_{ij}$ $(A_{ij} - B_{ij})$. Thus

$$A \pm B = \left[A_{ij} \pm B_{ij} \right] \tag{0.3}$$

Two matrices of the same order are said to be *conformable* for addition and subtraction. Addition and subtraction are not defined for matrices which are not conformable. If λ is a number and A is a matrix, then λA is a matrix given by

$$\lambda A = \left[\lambda A_{ij} \right] = A\lambda \tag{0.4}$$

Therefore,

$$-A = (-1)A = \left[-A_{ij} \right] \tag{0.5}$$

These definitions of addition and subtraction and, multiplication by a number imply that

$$A + B = B + A \tag{0.6}$$

$$A+(B+C)=(A+B)+C \tag{0.7}$$

$$A+0=A \tag{0.8}$$

$$A-A=0 \tag{0.9}$$

$$\lambda(A+B)=\lambda A+\lambda B \tag{0.10}$$

$$(\lambda+\mu)A=\lambda A+\mu A \tag{0.11}$$

and

$$1A=A \tag{0.12}$$

where A, B and C are as assumed to be conformable.

If A is an $M\times N$ matrix and B is an $N\times K$ matrix, then the product of B by A is written AB and is an $M\times K$ matrix with elements $\sum_{j=1}^{N}A_{ij}B_{js}$, $i=1,...,M$, $s=1,...,K$. For example, if

$$A=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\\ A_{31} & A_{32}\end{bmatrix} \qquad \text{and} \qquad B=\begin{bmatrix}B_{11} & B_{12}\\ B_{21} & B_{22}\end{bmatrix}$$

then AB is a 3×2 matrix given by

$$AB=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\\ A_{31} & A_{32}\end{bmatrix}\begin{bmatrix}B_{11} & B_{12}\\ B_{21} & B_{22}\end{bmatrix}$$
$$=\begin{bmatrix}A_{11}B_{11}+A_{12}B_{21} & A_{11}B_{12}+A_{12}B_{22}\\ A_{21}B_{11}+A_{22}B_{21} & A_{21}B_{12}+A_{22}B_{22}\\ A_{31}B_{11}+A_{32}B_{21} & A_{32}B_{12}+A_{32}B_{22}\end{bmatrix}$$

The product AB is defined only when the number of columns of A is equal to the number of rows of B. If this is the case, A is said to be conformable to B for multiplication. If A is conformable to B, then B is not necessarily conformable to A. Even if BA is defined, it is not necessarily equal to AB. On the assumption that A, B ,and C are conformable for the indicated sums and products, it is possible to show that

$$A(B+C)=AB+AC \tag{0.13}$$

$$(A+B)C=AC+BC \tag{0.14}$$

and

$$A(BC) = (AB)C \tag{0.15}$$

However, $AB \neq BA$ in general, $AB = 0$ does not imply $A = 0$ or $B = 0$, and $AB = AC$ does not necessarily imply $B = C$.

The square matrix I defined by

$$I = \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & \cdot & \cdot & \cdot & 0 \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ 0 & 0 & \cdot & \cdot & \cdot & 1 \end{bmatrix} \tag{0.16}$$

is the *identity* matrix. The identity matrix is a special case of a diagonal matrix which has all of its elements zero except the diagonal ones. A square matrix A whose elements satisfy $A_{ij} = 0$, $i > j$, is called *an upper triangular matrix* , i.e.,

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} & \cdot & \cdot & \cdot & A_{1N} \\ 0 & A_{22} & A_{23} & \cdot & \cdot & \cdot & A_{2N} \\ 0 & 0 & A_{33} & \cdot & \cdot & \cdot & \\ \cdot & & & & & & \\ \cdot & & & & & & \\ \cdot & & & & & & \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & A_{NN} \end{bmatrix}$$

A lower triangular matrix can be defined in a similar fashion. A diagonal matrix is both an upper triangular matrix and a lower triangular matrix.

If A and B are square matrices of the same order such that $AB = BA = I$, then B is called the *inverse* of A and we write $B = A^{-1}$. Also, A is the inverse of B, i.e. $A = B^{-1}$. If A has an inverse it is said to be *nonsingular*. If A and B are square matrices of the same order with inverses A^{-1} and B^{-1} respectively, then

$$(AB)^{-1} = B^{-1}A^{-1} \tag{0.17}$$

Equation (0.17) follows because

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

and similarly

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

The matrix of order $N \times M$ obtained by interchanging the rows and columns of an $M \times N$ matrix A is called the *transpose* of A and is denoted by A^T. It is easily shown that

$$(A^T)^T = A \tag{0.18}$$

$$(\lambda A)^T = \lambda A^T \tag{0.19}$$

$$(A+B)^T = A^T + B^T \tag{0.20}$$

and

$$(AB)^T = B^T A^T \tag{0.21}$$

A square matrix A is *symmetric* if $A = A^T$ and *skew-symmetric* if $A = -A^T$. Therefore, for a symmetric matrix

$$A_{ij} = A_{ji} \tag{0.22}$$

and for a skew-symmetric matrix

$$A_{ij} = -A_{ji} \tag{0.23}$$

Equation (0.23) implies that the diagonal elements of a skew symmetric-matrix are all zero. Every square matrix A can be written uniquely as the sum of a symmetric matrix and a skew-symmetric matrix, namely

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T) \tag{0.24}$$

If A is a square matrix, its *determinant* is written $\det A$. The reader is assumed at this stage to have some familiarity with the computational properties of determinants. In particular, it should be known that

$$\det A = \det A^T \qquad \text{and} \qquad \det AB = (\det A)(\det B) \tag{0.25}$$

If A is an $N \times N$ square matrix, we can construct an $(N-1)\times(N-1)$ square matrix by removing the i^{th} row and the j^{th} column. The determinant of this matrix is denoted by M_{ij} and is the *minor* of A_{ij}. The *cofactor* of A_{ij} is defined by

$$\text{cof } A_{ij} = (-1)^{i+j} M_{ij} \tag{0.26}$$

For example, if

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \tag{0.27}$$

then

$$M_{11} = A_{22}, \qquad M_{12} = A_{21} \qquad M_{21} = A_{12} \qquad M_{22} = A_{11}$$
$$\text{cof } A_{11} = A_{22} \quad \text{cof } A_{12} = -A_{21}, \qquad \text{etc.}$$

The *adjoint* of an $N \times N$ matrix A, written $\text{adj}\, A$, is an $N \times N$ matrix given by

$$\text{adj}\, A = \begin{bmatrix} \text{cof } A_{11} & \text{cof } A_{21} & \cdot & \cdot & \cdot & \text{cof } A_{N1} \\ \text{cof } A_{12} & \text{cof } A_{22} & \cdot & \cdot & \cdot & \text{cof } A_{N2} \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ \text{cof } A_{1N} & \text{cof } A_{2N} & \cdot & \cdot & \cdot & \text{cof } A_{NN} \end{bmatrix} \tag{0.28}$$

The reader is cautioned that the designation "adjoint" is used in a different context in Section 18. However, no confusion should arise since the adjoint as defined in Section 18 will be designated by a different symbol. It is possible to show that

$$A(\text{adj}\, A) = (\det A) I = (\text{adj}\, A) A \tag{0.29}$$

We shall prove (0.29) in general later in Section 21; so we shall be content to establish it for $N = 2$ here. For $N = 2$

$$\text{adj}\, A = \begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{bmatrix}$$

Then

$$A(\text{adj}\,A) = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} A_{22} & -A_{21} \\ -A_{12} & A_{11} \end{bmatrix} = \begin{bmatrix} A_{11}A_{22} - A_{12}A_{21} & 0 \\ 0 & A_{11}A_{22} - A_{12}A_{21} \end{bmatrix}$$

Therefore

$$A(\text{adj}\,A) = (A_{11}A_{22} - A_{12}A_{21})I = (\det A)I$$

Likewise

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

If $\det A \neq 0$, then (0.29) shows that the inverse A^{-1} exists and is given by

$$A^{-1} = \frac{\text{adj}\,A}{\det A} \tag{0.30}$$

Of course, if A^{-1} exists, then $(\det A)(\det A^{-1}) = \det I = 1 \neq 0$. Thus nonsingular matrices are those with a nonzero determinant.

Matrix notation is convenient for manipulations of systems of linear algebraic equations. For example, if we have the set of equations

$$\begin{aligned} &A_{11}x_1 + A_{12}x_2 + A_{13}x_3 + \cdots + A_{1N}x_N = y_1 \\ &A_{21}x_1 + A_{22}x_2 + A_{23}x_3 + \cdots + A_{2N}x_N = y_2 \\ &\cdot \\ &\cdot \\ &\cdot \\ &A_{N1}x_1 + A_{N2}x_2 + A_{N3}x_3 + \cdots + A_{NN}x_N = y_2 \end{aligned}$$

then they can be written

$$\begin{bmatrix} A_{11} & A_{12} & \cdot & \cdot & \cdot & A_{1N} \\ A_{21} & A_{22} & & & & A_{2N} \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ A_{N1} & A_{N2} & \cdot & \cdot & \cdot & A_{NN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_N \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_N \end{bmatrix}$$

The above matrix equation can now be written in the compact notation

$$AX = Y \tag{0.31}$$

and if A is nonsingular the solution is

$$X = A^{-1}Y \tag{0.32}$$

For $N = 2$, we can immediately write

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{\det A}\begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \tag{0.33}$$

Exercises

0.1 Add the matrices

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} -5 \\ 6 \end{bmatrix}$$

Add the matrices

$$\begin{bmatrix} 2i & 3 & 7+2i \\ 5 & 4+3i & i \end{bmatrix} + \begin{bmatrix} 2 & 5i & -4 \\ -4i & 3+3i & -i \end{bmatrix}$$

0.2 Add

$$\begin{bmatrix} 1 \\ 2i \end{bmatrix} + 3\begin{bmatrix} -5i \\ 6 \end{bmatrix}$$

0.3 Multiply

$$\begin{bmatrix} 2i & 3 & 7+2i \\ 5 & 4+3i & i \end{bmatrix}\begin{bmatrix} 2i & 8 \\ 1 & 6i \\ 3i & 2 \end{bmatrix}$$

0.4 Show that the product of two upper (lower) triangular matrices is an upper lower triangular matrix. Further, if

$$A = \begin{bmatrix} A_{ij} \end{bmatrix}, \qquad B = \begin{bmatrix} B_{ij} \end{bmatrix}$$

are upper (lower) triangular matrices of order $N \times N$, then

$$(AB)_{ii} = (BA)_{ii} = A_{ii}B_{ii}$$

for all $i = 1, ..., N$. The off diagonal elements $(AB)_{ij}$ and $(BA)_{ij}$, $i \neq j$, generally are not equal, however.

0.5 What is the transpose of

$$\begin{bmatrix} 2i & 3 & 7+2i \\ 5 & 4+3i & i \end{bmatrix}$$

0.6 What is the inverse of

$$\begin{bmatrix} 5 & 3 & 1 \\ 0 & 7 & 2 \\ 1 & 4 & 0 \end{bmatrix}$$

0.7 Solve the equations

$$\begin{aligned} 5x + 3y &= 2 \\ x + 4y &= 9 \end{aligned}$$

by use of (0.33).

Chapter 1

SETS, RELATIONS, AND FUNCTIONS

The purpose of this chapter is to introduce an initial vocabulary and some basic concepts. Most of the readers are expected to be somewhat familiar with this material; so its treatment is brief.

Section 1. Sets and Set Algebra

The concept of a set is regarded here as primitive. It is a collection or family of things viewed as a simple entity. The things in the set are called *elements* or *members* of the set. They are said to be *contained in* or to *belong* to the set. Sets will generally be denoted by upper case script letters, $\mathscr{A},\mathscr{B},\mathscr{C},\mathscr{D},\ldots$, and elements by lower case letters $a,b,c,d,\ldots$. The sets of complex numbers, real numbers and integers will be denoted by $\mathscr{C}$, $\mathscr{R}$, and $\mathscr{I}$, respectively. The notation $a \in \mathscr{A}$ means that the element a is contained in the set $\mathscr{A}$; if a is not an element of $\mathscr{A}$, the notation $a \notin \mathscr{A}$ is employed. To denote the set whose elements are $a,b,c,$ and d, the notation $\{a,b,c,d\}$ is employed. In mathematics a set is not generally a collection of unrelated objects like a tree, a doorknob and a concept, but rather it is a collection which share some common property like the vineyards of France which share the common property of being in France or the real numbers which share the common property of being real. A set whose elements are determined by their possession of a certain property is denoted by $\{x|P(x)\}$, where x denotes a typical element and $P(x)$ is the property which determines x to be in the set.

If $\mathscr{A}$ and $\mathscr{B}$ are sets, $\mathscr{B}$ is said to be a *subset* of $\mathscr{A}$ if every element of $\mathscr{B}$ is also an element of $\mathscr{A}$. It is customary to indicate that $\mathscr{B}$ is a subset of $\mathscr{A}$ by the notation $\mathscr{B} \subset \mathscr{A}$, which may be read as "$\mathscr{B}$ is contained in $\mathscr{A}$," or $\mathscr{A} \supset \mathscr{B}$ which may be read as "$\mathscr{A}$ contains $\mathscr{B}$." For example, the set of integers $\mathscr{I}$ is a subset of the set of real numbers $\mathscr{R}$, $\mathscr{I} \subset \mathscr{R}$. Equality of two sets $\mathscr{A}$ and $\mathscr{B}$ is said to exist if $\mathscr{A}$ is a subset of $\mathscr{B}$ and $\mathscr{B}$ is a subset of $\mathscr{A}$; in equation form

$$\mathscr{A} = \mathscr{B} \Leftrightarrow \mathscr{A} \subset \mathscr{B} \text{ and } \mathscr{B} \subset \mathscr{A} \tag{1.1}$$

A nonempty subset $\mathscr{B}$ of $\mathscr{A}$ is called a *proper subset* of $\mathscr{A}$ if $\mathscr{B}$ is not equal to $\mathscr{A}$. The set of integers $\mathscr{I}$ is actually a proper subset of the real numbers. The *empty set* or *null set* is the set with no elements and is denoted by $\varnothing$. The *singleton* is the set containing a single element a and is denoted by $\{a\}$. A set whose elements are sets is often called a *class*.

Some operations on sets which yield other sets will now be introduced. The *union* of the sets $\mathscr{A}$ and $\mathscr{B}$ is the set of all elements that are either in the set $\mathscr{A}$ or in the set $\mathscr{B}$. The union of $\mathscr{A}$ and $\mathscr{B}$ is denoted by $\mathscr{A} \cup \mathscr{B}$

$$\mathscr{A} \cup \mathscr{B} \equiv \{a \mid a \in \mathscr{A} \text{ or } a \in \mathscr{B}\} \tag{1.2}$$

It is easy to show that the operation of forming a union is commutative,

$$\mathscr{A} \cup \mathscr{B} = \mathscr{B} \cup \mathscr{A} \tag{1.3}$$

and associative

$$\mathscr{A} \cup (\mathscr{B} \cup \mathscr{C}) = (\mathscr{A} \cup \mathscr{B}) \cup \mathscr{C} \tag{1.4}$$

The *intersection* of the sets $\mathscr{A}$ and $\mathscr{B}$ is the set of elements that are in both $\mathscr{A}$ and $\mathscr{B}$. The intersection is denoted by $\mathscr{A} \cap \mathscr{B}$ and is specified by

$$\mathscr{A} \cap \mathscr{B} = \{a \mid a \in \mathscr{A} \text{ and } a \in \mathscr{B}\} \tag{1.5}$$

Again, it can be shown that the operation of intersection is commutative

$$\mathscr{A} \cap \mathscr{B} = \mathscr{B} \cap \mathscr{A} \tag{1.6}$$

and associative

$$\mathscr{A} \cap (\mathscr{B} \cap \mathscr{C}) = (\mathscr{A} \cap \mathscr{B}) \cap \mathscr{C} \tag{1.7}$$

Two sets are said to be disjoint if they have no elements in common, i.e. if

$$\mathscr{A} \cap \mathscr{B} = \varnothing \tag{1.8}$$

The operations of union and intersection are related by the following distributive laws:

$$\begin{aligned} \mathscr{A} \cap (\mathscr{B} \cup \mathscr{C}) &= (\mathscr{A} \cap \mathscr{B}) \cup (\mathscr{A} \cap \mathscr{C}) \\ \mathscr{A} \cup (\mathscr{B} \cap \mathscr{C}) &= (\mathscr{A} \cup \mathscr{B}) \cap (\mathscr{A} \cup \mathscr{C}) \end{aligned} \tag{1.9}$$

The *complement* of the set $\mathscr{B}$ with respect to the set $\mathscr{A}$ is the set of all elements contained in $\mathscr{A}$ but not contained in $\mathscr{B}$. The complement of $\mathscr{B}$ with respect to the set $\mathscr{A}$ is denoted by $\mathscr{A} / \mathscr{B}$ and is specified by

$$\mathscr{A} / \mathscr{B} = \{a \mid a \in \mathscr{A} \text{ and } a \notin \mathscr{B}\} \tag{1.10}$$

It is easy to show that

$$\mathscr{A} /(\mathscr{A} / \mathscr{B}) = \mathscr{B} \quad \text{for} \quad \mathscr{B} \subset \mathscr{A} \tag{1.11}$$

and

$$\mathscr{A} / \mathscr{B} = \mathscr{A} \Leftrightarrow \mathscr{A} \cap \mathscr{B} = \varnothing \tag{1.12}$$

Exercises

1.1 List all possible subsets of the set $\{a, b, c, d\}$.

1.2 List two proper subsets of the set $\{a, b\}$.

1.3 Prove the following formulas:
- (a) $\mathscr{A} = \mathscr{A} \cup \mathscr{B} \Leftrightarrow \mathscr{B} \subset \mathscr{A}$.
- (b) $\mathscr{A} = \mathscr{A} \cup \varnothing$.
- (c) $\mathscr{A} = \mathscr{A} \cup \mathscr{A}$.

1.4 Show that $\mathscr{I} \cap \mathscr{R} = \mathscr{I}$.

1.5 Verify the distributive laws (1.9).

1.6 Verify the following formulas:
- (a) $\mathscr{A} \subset \mathscr{B} \Leftrightarrow \mathscr{A} \cap \mathscr{B} = \mathscr{A}$.
- (b) $\mathscr{A} \cap \varnothing = \varnothing$.
- (c) $\mathscr{A} \cap \mathscr{A} = \mathscr{A}$.

1.7 Give a proof of the commutative and associative properties of the intersection operation.

1.8 Let $\mathscr{A} = \{-1, -2, -3, -4\}$, $\mathscr{B} = \{-1, 0, 1, 2, 3, 7\}$, $\mathscr{C} = \{0\}$, and $\mathscr{D} = \{-7, -5, -3, -1, 1, 2, 3\}$. List the elements of $\mathscr{A} \cup \mathscr{B}, \mathscr{A} \cap \mathscr{B}, \mathscr{A} \cup \mathscr{C}, \mathscr{B} \cap \mathscr{C}, \mathscr{A} \cup \mathscr{B} \cup \mathscr{C}, \mathscr{A} / \mathscr{B}$, and $(\mathscr{D} / \mathscr{C}) \cup \mathscr{A}$.

Section 2. Ordered Pairs, Cartesian Products, and Relations

The idea of ordering is not involved in the definition of a set. For example, the set $\{a,b\}$ is equal to the set $\{b,a\}$. In many cases of interest it is important to order the elements of a set. To define an *ordered pair* (a,b) we single out a particular element of $\{a,b\}$, say a, and define (a,b) to be the class of sets; in equation form,

$$(a,b) \equiv \{\{a\},\{a,b\}\} \tag{2.1}$$

This ordered pair is then different from the ordered pair (b,a) which is defined by

$$(b,a) \equiv \{\{b\},\{b,a\}\} \tag{2.2}$$

It follows directly from the definition that

$$(a,b) = (c,d) \Leftrightarrow a = c \text{ and } b = d$$

Clearly, the definition of an ordered pair can be extended to an ordered N-tuple $(a_1, a_2, ..., a_N)$.

The *Cartesian Product* of two sets $\mathscr{A}$ and $\mathscr{B}$ is a set whose elements are ordered pairs (a,b) where a is an element of $\mathscr{A}$ and b is an element of $\mathscr{B}$. We denote the Cartesian product of $\mathscr{A}$ and $\mathscr{B}$ by $\mathscr{A} \times \mathscr{B}$.

$$\mathscr{A} \times \mathscr{B} = \{(a,b) \mid a \in \mathscr{A}, b \in \mathscr{B}\} \tag{2.3}$$

It is easy to see how a Cartesian product of N sets can be formed using the notion of an ordered N-tuple.

Any subset $\mathscr{K}$ of $\mathscr{A} \times \mathscr{B}$ defines a relation from the set $\mathscr{A}$ to the set $\mathscr{B}$. The notation $a\mathscr{K}b$ is employed if $(a,b) \in \mathscr{K}$; this notation is modified for negation as $a\not\mathscr{K}b$ if $(a,b) \notin \mathscr{K}$. As an example of a relation let $\mathscr{A}$ be the set of all Volkswagons in the United States and let $\mathscr{B}$ be the set of all Volkswagons in Germany. Let be $\mathscr{K} \subset \mathscr{A} \times \mathscr{B}$ be defined so that $a\mathscr{K}b$ if b is the same color as a.

If $\mathscr{A} = \mathscr{B}$, then $\mathscr{K}$ is a relation *on* $\mathscr{A}$. Such a relation is said to be *reflexive* if $a\mathscr{K}a$ for all $a \in \mathscr{A}$, it is said to be *symmetric* if $a\mathscr{K}b$ whenever $b\mathscr{K}a$ for any a and b in $\mathscr{A}$, and it is said to be transitive if $a\mathscr{K}b$ and $b\mathscr{K}c$ imply that $a\mathscr{K}c$ for any a, b and c in $\mathscr{A}$. A relation on a set $\mathscr{A}$ is said to be an *equivalence relation* if it is reflexive, symmetric and transitive. As an example of an equivalence relation $\mathscr{K}$ on the set of all real numbers $\mathscr{R}$ let $a\mathscr{K}b \Leftrightarrow a = b$ for all a, b. To verify that this is an equivalence relation, note that $a = a$ for each $a \in \mathscr{K}$ (reflexivity), $a = b \Rightarrow b = a$ for each $a \in \mathscr{K}$ (symmetry), and $a = b, b = c \Rightarrow a = c$ for a, b and c in $\mathscr{K}$ (transitivity).

A relation $\mathscr{K}$ on $\mathscr{A}$ is antisymmetric if $a\mathscr{K}b$ then $b\mathscr{K}a$ unless $b = a$. A relation on a set $\mathscr{A}$ is said to be a *partial ordering* if it is reflexive, antisymmetric and transitive. The equality relation is a trivial example of partial ordering. As another example of a partial ordering on $\mathscr{R}$, let the inequality $\leq$ be the relation. To verify that this is a partial ordering note that $a \leq a$ for every $a \in \mathscr{R}$ (reflexivity), $a \leq b$ and $b \leq a \Rightarrow a = b$ for all $a, b \in \mathscr{R}$ (antisymmetry), and $a \leq b$ and $b \leq c \Rightarrow a \leq c$ for all $a, b, c \in \mathscr{R}$ (transitivity). Of course inequality is not an equivalence relation since it is not symmetric.

Exercises

2.1 Define an ordered triple (a, b, c) by

$$(a, b, c) = ((a, b), c)$$

and show that $(a, b, c) = (d, e, f) \Leftrightarrow a = d, b = e$, and $c = f$.

2.2 Let $\mathscr{A}$ be a set and $\mathscr{K}$ an equivalence relation on $\mathscr{A}$. For each $a \in \mathscr{A}$ consider the set of all elements x that stand in the relation $\mathscr{K}$ to a; this set is denoted by $\{x \mid a\mathscr{K}x\}$, and is called an equivalence set.

(a) Show that the equivalence set of a contains a.

(b) Show that any two equivalence sets either coincide or are disjoint.

Note: (a) and (b) show that the equivalence sets form a partition of $\mathscr{A}$; $\mathscr{A}$ is the disjoint union of the equivalence sets.

2.3 On the set $\mathscr{R}$ of all real numbers does the strict inequality $<$ constitute a partial ordering on the set?

2.4 For ordered triples of real numbers define $(a_1, b_1, c_1) \leq (a_2, b_2, c_2)$ if $a_1 \leq a_2, b_1 \leq b_2, c_1 \leq c_2$. Does this relation define a partial ordering on $\mathscr{R} \times \mathscr{R} \times \mathscr{R}$?

Section 3. Functions

A relation f from $\mathscr{X}$ to $\mathscr{Y}$ is said to be a *function* (or a mapping) if $(x, y_1) \in f$ and $(x, y_2) \in f$ imply $y_1 = y_2$ for all $x \in \mathscr{X}$. A function is thus seen to be a particular kind of relation which has the property that for each $x \in \mathscr{X}$ there exists one and only one $y \in \mathscr{Y}$ such that $(x, y) \in f$. Thus a function f defines a *single-valued relation*. Since a function is just a particular relation, the notation of the last section could be used, but this is rarely done. A standard notation for a function from $\mathscr{X}$ to $\mathscr{Y}$ is

$$f : \mathscr{X} \to \mathscr{Y}$$

or

$$y = f(x)$$

which indicates the element $y \in \mathscr{Y}$ that f associates with $x \in \mathscr{X}$. The element y is called the *value* of f at x. The *domain* of the function f is the set $\mathscr{X}$. The *range* of f is the set of all y for which there exists an x such that $y = f(x)$. We denote the range of f by $f(\mathscr{X})$. When the domain of f is the set of real numbers $\mathscr{R}$, f is said to be a *function of a real variable.* When $f(\mathscr{X})$ is contained in $\mathscr{R}$, the function is said to be *real-valued*. *Functions of a complex variable* and *complex- valued* functions are defined analogously. A function of a real variable need not be real-valued, nor need a function of a complex variable be complex-valued.

If $\mathscr{X}_0$ is a subset of $\mathscr{X}$, $\mathscr{X}_0 \subset \mathscr{X}$, and $f : \mathscr{X} \to \mathscr{Y}$, the image of $\mathscr{X}_0$ under f is the set

$$f(\mathscr{X}_0) \equiv \{ f(x) \mid x \in \mathscr{X}_0 \} \tag{3.1}$$

and it is easy to prove that

$$f(\mathscr{X}_0) \subset f(\mathscr{X}) \tag{3.2}$$

Similarly, if $\mathscr{Y}_0$ is a subset of $\mathscr{Y}$, then the *preimage* of $\mathscr{Y}_0$ under f is the set

$$f^{-1}(\mathscr{Y}_0) \equiv \{ x \mid f(x) \in \mathscr{Y}_0 \} \tag{3.3}$$

and it is easy to prove that

$$f^{-1}(\mathscr{Y}_0) \subset \mathscr{X} \tag{3.4}$$

A function f is said be *into* $\mathscr{Y}$ if $f(\mathscr{X})$ is a proper subset of $\mathscr{Y}$, and it is said to be *onto* $\mathscr{Y}$ if $f(\mathscr{X}) = \mathscr{Y}$. A function that is onto is also called *surjective*. Stated in a slightly different way, a function is onto if for every element $y \in \mathscr{Y}$ there exists at least one element $x \in \mathscr{X}$ such that $y = f(x)$. A function f is said to be *one-to-one* (or *injective*) if for every $x_1, x_2 \in \mathscr{X}$

$$f(x_1) = f(x_2) \Rightarrow x_1 = x_2 \tag{3.5}$$

In other words, a function is one-to-one if for every element $y \in f(\mathscr{X})$ there exists only one element $x \in \mathscr{X}$ such that $y = f(x)$. A function f is said to form a *one-to-one correspondence* (or to be *bijective*) from $\mathscr{X}$ to $\mathscr{Y}$ if $f(\mathscr{X}) = \mathscr{Y}$, and f is one-to-one. Naturally, a function can be onto without being one-to-one or be one-to-one without being onto. The following examples illustrate these terminologies.

1. Let $f(x) = x^5$ be a particular real valued function of a real variable. This function is one-to-one because $x^5 = z^5$ implies $x = z$ when $x, z \in \mathscr{R}$ and it is onto since for every $y \in \mathscr{R}$ there is some x such that $y = x^5$.
2. Let $f(x) = x^3$ be a particular complex-valued function of a complex variable. This function is onto but it is not one-to-one because, for example,

$$f(1) = f(-\frac{1}{2} + i\frac{\sqrt{3}}{2}) = f(-\frac{1}{2} - i\frac{\sqrt{3}}{2}) = 1^3$$

   where $i = \sqrt{-1}$.
3. Let $f(x) = 2|x|$ be a particular real valued function of a real variable where $|x|$ denotes the absolute value of x. This function is into rather than onto and it is not one-to-one.

If $f : \mathscr{X} \to \mathscr{Y}$ is a one-to-one function, then there exists a function $f^{-1} : f(\mathscr{X}) \to \mathscr{X}$ which associates with each $y \in f(\mathscr{X})$ the unique $x \in \mathscr{X}$ such that $y = f(x)$. We call f^{-1} the *inverse* of f, written $x = f^{-1}(y)$. Unlike the preimage defined by (3.3), the definition of inverse functions is possible only for functions that are one-to-one. Clearly, we have $f^{-1}(f(\mathscr{X})) = \mathscr{X}$. The *composition* of two functions $f : \mathscr{X} \to \mathscr{Y}$ and $g :\to \mathscr{Y} \to \mathscr{Z}$ is a function $h : \mathscr{X} \to \mathscr{Z}$ defined by $h(x) = g(f(x))$ for all $x \in \mathscr{X}$. The function h is written $h = g \circ f$. The composition of any finite number of functions is easily defined in a fashion similar to that of a pair. The operation of composition of functions is not generally commutative,

$$g \circ f \neq f \circ g$$

Indeed, if $g \circ f$ is defined, $f \circ g$ may not be defined, and even if $g \circ f$ and $f \circ g$ are both defined, they may not be equal. The operation of composition is associative

$$h \circ (g \circ f) = (h \circ g) \circ f$$

if each of the indicated compositions is defined. The identity function $id : \mathscr{X} \to \mathscr{X}$ is defined as the function $id(x) = x$ for all $x \in \mathscr{X}$. Clearly if f is a one-to-one correspondence from $\mathscr{X}$ to $\mathscr{Y}$, then

$$f^{-1} \circ f = id_{\mathscr{X}}, \qquad f \circ f^{-1} = id_{\mathscr{Y}}$$

where $id_{\mathscr{X}}$ and $id_{\mathscr{Y}}$ denote the identity functions of $\mathscr{X}$ and $\mathscr{Y}$, respectively.

To close this discussion of functions we consider the special types of functions called sequences. A *finite sequence* of N terms is a function f whose domain is the set of the first N positive integers $\{1, 2, 3, ..., N\}$. The range of f is the set of N elements $\{f(1), f(2), ..., f(N)\}$, which is usually denoted by $\{f_1, f_2, ..., f_N\}$. The elements $f_1, f_2, ..., f_N$ of the range are called *terms* of the sequence. Similarly, an *infinite sequence* is a function defined on the positive integers $\mathscr{I}^+$. The range of an infinite sequence is usually denoted by $\{f_n\}$, which stands for the infinite set $\{f_1, f_2, f_3, ...\}$. A function g whose domain is $\mathscr{I}^+$ and whose range is contained in $\mathscr{I}^+$ is said to be *order preserving* if $m < n$ implies that $g(m) < g(n)$ for all $m, n \in \mathscr{I}^+$. If f is an infinite sequence and g is order preserving, then the composition $f \circ g$ is a *subsequence* of f. For example, let

$$f_n = 1/(n+1), \qquad g_n = 4$$

Then

$$f \circ g(n) = 1/(4^n + 1)$$

is a subsequence of f.

Exercises

3.1 Verify the results (3.2) and (3.4).

3.2 How may the domain and range of the sine function be chosen so that it is a one-to-one correspondence?

3.3 Which of the following conditions defines a function $y = f(x)$?

$$x^2 + y^2 = 1, \qquad xy = 1, \qquad x, y \in \mathscr{R}.$$

3.4 Give an example of a real valued function of a complex variable.

3.5 Can the implication of (3.5) be reversed? Why?

3.6 Under what conditions is the operation of composition of functions commutative?

3.7 Show that the arc sine function $\sin^{-1}$ really is not a function from $\mathscr{R}$ to $\mathscr{R}$. How can it be made into a function?

Chapter 2

GROUPS, RINGS, AND FIELDS

The mathematical concept of a group has been crystallized and abstracted from numerous situations both within the body of mathematics and without. Permutations of objects in a set is a group operation. The operations of multiplication and addition in the real number system are group operations. The study of groups is developed from the study of particular situations in which groups appear naturally, and it is instructive that the group itself be presented as an object for study. In the first three sections of this chapter certain of the basic properties of groups are introduced. In the last section the group concept is used in the definition of rings and fields.

Section 4. The Axioms for a Group

Central to the definition of a group is the idea of a binary operation. If $\mathcal{G}$ is a non-empty set, a binary operation on $\mathcal{G}$ is a function from $\mathcal{G} \times \mathcal{G}$ to $\mathcal{G}$. If $a, b \in \mathcal{G}$, the binary operation will be denoted by $*$ and its value by $a * b$. The important point to be understood about a binary operation on $\mathcal{G}$ is that $\mathcal{G}$ is closed with respect to $*$ in the sense that if $a, b \in \mathcal{G}$ then $a * b \in \mathcal{G}$ also. A binary operation $*$ on $\mathcal{G}$ is associative if

$$(a * b) * c = a * (b * c) \quad \text{for all } a, b, c \in \mathcal{G} \qquad \textbf{(4.1)}$$

Thus, parentheses are unnecessary in the combination of more than two elements by an associative operation. A semigroup is a pair $(\mathcal{G}, *)$ consisting of a nonempty set $\mathcal{G}$ with an associative binary operation $*$. A binary operation $*$ on $\mathcal{G}$ is commutative if

$$a * b = b * a \qquad \text{for all } a, b \in \mathcal{G} \qquad \textbf{(4.2)}$$

The multiplication of two real numbers and the addition of two real numbers are both examples of commutative binary operations. The multiplication of two N by N matrices ($N > 1$) is a noncommutative binary operation. A binary operation is often denoted by $a \circ b$, $a \cdot b$, or ab rather than by $a * b$; if the operation is commutative, the notation $a + b$ is sometimes used in place of $a * b$.

An element $e \in \mathcal{G}$ that satisfies the condition

$$e * a = a * e = a \qquad \text{for all } a \in \mathcal{G} \tag{4.3}$$

is called an identity element for the binary operation $*$ on the set $\mathcal{G}$. In the set of real numbers with the binary operation of multiplication, it is easy to see that the number 1 plays the role of identity element. In the set of real numbers with the binary operation of addition, 0 has the role of the identity element. Clearly $\mathcal{G}$ contains at most one identity element in e. For if e^1 is another identity element in $\mathcal{G}$, then $e^1 * a = a * e^1 = a$ for all $a \in G$ also. In particular, if we choose $a = e$, then $e^1 * e = e$. But from (4.3), we have also $e^1 * e = e^1$. Thus $e^1 = e$. In general, $\mathcal{G}$ need not have any identity element. But if there is an identity element, and if the binary operation is regarded as multiplicative, then the identity element is often called the unity element; on the other hand, if the binary operation is additive, then the identity element is called the zero element.

In a semigroup $\mathcal{G}$ containing an identity element e with respect to the binary operation $*$, an element a^{-1} is said to be an inverse of the element a if

$$a * a^{-1} = a^{-1} * a = e \tag{4.4}$$

In general, a need not have an inverse. But if an inverse a^{-1} of a exists, then it is unique, the proof being essentially the same as that of the uniqueness of the identity element. The identity element is its own inverse. In the set $\mathcal{R}/\{0\}$ with the binary operation of multiplication, the inverse of a number is the reciprocal of the number. In the set of real numbers with the binary operation of addition the inverse of a number is the negative of the number.

A group is a pair $(\mathcal{G}, *)$ consisting of an associative binary operation $*$ and a set $\mathcal{G}$ which contains the identity element and the inverses of all elements of $\mathcal{G}$ with respect to the binary operation $*$. This definition can be explicitly stated in equation form as follows.

Definition. A group is a pair $(\mathcal{G}, *)$ where $\mathcal{G}$ is a set and $*$ is a binary operation satisfying the following:

(a) $(a * b) * c = a * (b * c)$ for all $a, b, c \in \mathcal{G}$.

(b) There exists an element $e \in \mathcal{G}$ such that $a * e = e * a = a$ for all $a \in \mathcal{G}$.

(c) For every $a \in G$ there exists an element $a^{-1} \in G$ such that $a * a^{-1} = a^{-1} * a = e$.

If the binary operation of the group is commutative, the group is said to be a commutative (or Abelian) group.

The set $\mathscr{R}/\{0\}$ with the binary operation of multiplication forms a group, and the set $\mathscr{R}$ with the binary operation of addition forms another group. The set of positive integers with the binary operation of multiplication forms a semigroup with an identity element but does not form a group because the condition © above is not satisfied.

A notational convention customarily employed is to denote a group simply by $\mathscr{G}$ rather than by the pair $(\mathscr{G},*)$. This convention assumes that the particular $*$ to be employed is understood. We shall follow this convention here.

Exercises

4.1 Verify that the set $\mathscr{G} \in \{1, i, -1, -i\}$, where $i^2 = -1$, is a group with respect to the binary operation of multiplication.

4.2 Verify that the set $\mathscr{G}$ consisting of the four 2×2 matrices

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

constitutes a group with respect to the binary operation of matrix multiplication.

4.3 Verify that the set $\mathscr{G}$ consisting of the four 2×2 matrices

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

constitutes a group with respect to the binary operation of matrix multiplication.

4.4 Determine a subset of the set of all 3×3 matrices with the binary operation of matrix multiplication that will form a group.

4.5 If $\mathscr{A}$ is a set, show that the set $\mathscr{G}$, $\mathscr{G} = \{ f \mid f : \mathscr{A} \to \mathscr{A}, f$ is a one-to-one correspondence$\}$ is a one-to-one correspondence$\}$constitutes a group with respect to the binary operation of composition.

Section 5. Properties of a Group

In this section certain basic properties of groups are established. There is a great economy of effort in establishing these properties for a group in general because the properties will then be possessed by any system that can be shown to constitute a group, and there are many such systems of interest in this text. For example, any property established in this section will automatically hold for the group consisting of $\mathscr{R}/\{0\}$ with the binary operation of multiplication, for the group consisting of the real numbers with the binary operation of addition, and for groups involving vectors and tensors that will be introduced in later chapters.

The basic properties of a group $\mathscr{G}$ in general are:

1. The identity element $e \in \mathscr{G}$ is unique.

2. The inverse element a^{-1} of any element $a \in \mathscr{G}$ is unique. The proof of Property 1 was given in the preceding section. The proof of Property 2 follows by essentially the same argument.

If n is a positive integer, the powers of $a \in \mathscr{G}$ are defined as follows: *(i)* For $n = 1, a^1 = a$. *(ii)* For $n > 1, a^n = a^{n-1} * a$. *(iii)* $a^0 = e$. *(iv)* $a^{-n} = \left(a^{-1}\right)^n$.

3. If m, n, k are any integers, positive or negative, then for $a \in \mathscr{G}$

$$a^m * a^n = a^{m+n}, \quad (a^m)^n = a^{mn}, \quad \left(a^{m+n}\right)^k = a^{mk+nk}$$

In particular, when $m = n = -1$, we have:

4. $\left(a^{-1}\right)^{-1} = a$ for all $a \in \mathscr{G}$

5. $(a * b)^{-1} = b^{-1} * a^{-1}$ for all $a, b \in \mathscr{G}$

For square matrices the proof of this property is given by (0.17); the proof in general is exactly the same.

The following property is a useful algebraic rule for all groups; it gives the solutions x and y to the equations $x * a = b$ and $a * y = b$, respectively.

6. For any elements a, b in $\mathscr{G}$, the two equations $x * a = b$ and $a * y = b$ have the unique solutions $x = b * a^{-1}$ and $y = a^{-1} * b$.

Proof. Since the proof is the same for both equations, we will consider only the equation $x * a = b$. Clearly $x = b * a^{-1}$ satisfies the equation $x * a = b$,

$$x * a = \left(b * a^{-1}\right) * a = b * \left(a^{-1} * a\right) = b * e = b$$

Conversely, $x * a = b,$ implies

$$x = x * e = x * \left(a * a^{-1}\right) = \left(x * a\right) * a^{-1} = b * a^{-1}$$

which is unique. As a result we have the next two properties.

7. For any three elements a, b, c in $\mathcal{G}$, either $a * c = b * c$ or $c * a = c * b$ implies that $a = b$.

8. For any two elements a, b in the group $\mathcal{G}$, either $a * b = b$ or $b * a = b$ implies that a is the identity element.

A non-empty subset $\mathcal{G}'$ of $\mathcal{G}$ is a *subgroup* of $\mathcal{G}$ if $\mathcal{G}'$ is a group with respect to the binary operation of $\mathcal{G}$, i.e., $\mathcal{G}'$ is a subgroup of $\mathcal{G}$ if and only if *(i)* $e \in \mathcal{G}'$ *(ii)* $a \in \mathcal{G}' \Rightarrow a^{-1} \in \mathcal{G}'$ *(iii)* $a, b \in \mathcal{G}' \Rightarrow a * b \in \mathcal{G}'$.

9. Let $\mathcal{G}'$ be a nonempty subset of $\mathcal{G}$. Then $\mathcal{G}'$ is a subgroup if $a, b \in \mathcal{G}' \Rightarrow a * b^{-1} \in \mathcal{G}'$.

Proof. The proof consists in showing that the conditions *(i), (ii), (iii)* of a subgroup are satisfied. *(i)* Since $\mathcal{G}'$ is non-empty, it contains an element a, hence $a * a^{-1} = e \in \mathcal{G}'$. *(ii)* If $b \in \mathcal{G}'$ then $e * b^{-1} = b^{-1} \in \mathcal{G}'$. *(iii)* If $a, b \in \mathcal{G}'$, then $a * \left(b^{-1}\right)^{-1} = a * b \in \mathcal{G}'$.

If $\mathcal{G}$ is a group, then $\mathcal{G}$ itself is a subgroup of $\mathcal{G}$, and the group consisting only of the element e is also a subgroup of $\mathcal{G}$. A subgroup of $\mathcal{G}$ other than $\mathcal{G}$ itself and the group e is called a *proper subgroup* of $\mathcal{G}$.

10. The intersection of any two subgroups of a group remains a subgroup.

Exercises

5.1 Show that e is its own inverse.

5.2 Prove Property 3.

5.3 Prove Properties 7 and 8.

5.4 Show that $x = y$ in Property 6 if $\mathscr{G}$ is Abelian.

5.5 If $\mathscr{G}$ is a group and $a \in \mathscr{G}$ show that $a * a = a$ implies $a = e$.

5.6 Prove Property 10.

5.7 Show that if we look at the non-zero real numbers under multiplication, then (a) the rational numbers form a subgroup, (b) the positive real numbers form a subgroup, (c) the irrational numbers do not form a subgroup.

Section 6. Group Homomorphisms

A "group homomorphism" is a fairly overpowering phrase for those who have not heard it or seen it before. The word homomorphism means simply the same type of formation or structure. A group homomorphism has then to do with groups having the same type of formation or structure.

Specifically, if $\mathcal{G}$ and $\mathcal{H}$ are two groups with the binary operations $*$ and $\circ$, respectively, a function $f:\mathcal{G}\to\mathcal{H}$ is a *homomorphism* if

$$f(a*b)=f(a)\circ f(b) \qquad \text{for all } a,b\in\mathcal{G} \tag{6.1}$$

If a homomorphism exists between two groups, the groups are said to *be homomorphic*. A homomorphism $f:\mathcal{G}\to\mathcal{H}$ is an *isomorphism* if f is both one-to-one and onto. If an isomorphism exists between two groups, the groups are said to be *isomorphic*. Finally, a homomorphism $f:\mathcal{G}\to\mathcal{G}$ is called an *endomorphism* and an isomorphism $f:\mathcal{G}\to\mathcal{G}$ is called an *automorphism.* To illustrate these definitions, consider the following examples:

1. Let $\mathcal{G}$ be the group of all nonsingular, real, $N\times N$ matrices with the binary operation of matrix multiplication. Let $\mathcal{H}$ be the group $\mathcal{R}/\{0\}$ with the binary operation of scalar multiplication. The function that is the determinant of a matrix is then a homomorphism from $\mathcal{G}$ to $\mathcal{H}$.
2. Let $\mathcal{G}$ be any group and let $\mathcal{H}$ be the group whose only element is e. Let f be the function that assigns to every element of $\mathcal{G}$ the value e; then f is a (trivial) homomorphism.
3. The identity function $\mathrm{id}:\mathcal{G}\to\mathcal{G}$ is a (trivial) automorphism of any group.
4. Let $\mathcal{G}$ be the group of positive real numbers with the binary operation of multiplication and let $\mathcal{H}$ be the group of real numbers with the binary operation of addition. The log function is an isomorphism between $\mathcal{G}$ and $\mathcal{H}$.
5. Let $\mathcal{G}$ be the group $\mathcal{R}/\{0\}$ with the binary operation of multiplication. The function that takes the absolute value of a number is then an endomorphism of $\mathcal{G}$ into $\mathcal{G}$. The restriction of this function to the subgroup $\mathcal{H}$ of $\mathcal{H}$ consisting of all positive real numbers is a (trivial) automorphism of $\mathcal{H}$, however.

From these examples of homomorphisms and automorphisms one might note that a homomorphism maps identities into identities and inverses into inverses. The proof of this

observation is the content of Theorem 6.1 below. Theorem 6.2 shows that a homomorphism takes a subgroup into a subgroup and Theorem 6.3 proves a converse result.

Theorem 6.1. If $f:\mathcal{G}\to\mathcal{H}$ is a homomorphism, then $f(e)$ coincides with the identity element e_0 of $\mathcal{H}$ and

$$f(a^{-1}) = f(a)^{-1}$$

Proof. Using the definition (6.1) of a homomorphism on the identity $a = a*e$, one finds that

$$f(a) = f(a*e) = f(a)\circ f(e)$$

From Property 8 of a group it follows that $f(e)$ is the identity element e_0 of $\mathcal{H}$. Now, using this result and the definition of a homomorphism (6.1) applied to the equation $e = a*a^{-1}$, we find

$$e_0 = f(e) = f(a*a^{-1}) = f(a)\circ f(a^{-1})$$

It follows from Property 2 that $f(a^{-1})$ is the inverse of $f(a)$.

Theorem 6.2. If $f:\mathcal{G}\to\mathcal{H}$ is a homomorphism and if $\mathcal{G}'$ is a subgroup of $\mathcal{G}$, then $f(\mathcal{G}')$ is a subgroup of $\mathcal{H}$.

Proof. The proof will consist in showing that the set $f(\mathcal{G}')$ satisfies the conditions $(i),(ii),(iii)$ of a subgroup. (i) Since $e\in\mathcal{G}'$ it follows from Theorem 6.1 that the identity element e_0 of $\mathcal{H}$ is contained in $f(\mathcal{G}')$. (ii) For any $a\in\mathcal{G}'$, $f(a)^{-1}\in f(\mathcal{G}')$ by Theorem 6.1. (iii) For any $a,b\in\mathcal{G}'$, $f(a)\circ f(b)\in f(\mathcal{G}')$ since $f(a)\circ f(b) = f(a*b)\in f(\mathcal{G}')$.

As a corollary to Theorem 6.2 we see that $f(\mathcal{G})$ is itself a subgroup of $\mathcal{H}$.

Theorem 6.3. If $f:\mathcal{G}\to\mathcal{H}$ is a homomorphism and if $\mathcal{H}'$ is a subgroup of $\mathcal{H}$, then the preimage $f^{-1}(\mathcal{H}')$ is a subgroup of $\mathcal{G}$.

The kernel of a homomorphism $f:\mathcal{G}\to\mathcal{H}$ is the subgroup $f^{-1}(e_0)$ of $\mathcal{G}$. In other words, the kernel of f is the set of elements of $\mathcal{G}$ that are mapped by f to the identity element e_0 of $\mathcal{H}$.

The notation $K(f)$ will be used to denote the kernel of f. In Example 1 above, the kernel of f consists of $N \times N$ matrices with determinant equal to the real number 1, while in Example 2 it is the entire group $\mathscr{G}$. In Example 3 the kernel of the identity map is the identity element, of course.

Theorem 6.4. A homomorphism $f : \mathscr{G} \to \mathscr{H}$ is one-to-one if and only if $K(f) = \{e\}$.

Proof. This proof consists in showing that $f(a) = f(b)$ implies $a = b$ if and only if $K(f) = \{e\}$. If $f(a) = f(b)$, then $f(a) \circ f(b)^{-1} = e_0$, hence $f(a * b^{-1}) = e_{0,}$ and it follows that $a * b^{-1} \in K(f)$. Thus, if $K(f) = \{e\}$, then $a * b^{-1} = e$ or $a = b$. Conversely, now, we assume f is

one-to-one and since $K(f)$ is a subgroup of $\mathscr{G}$ by Theorem 6.3, it must contain e. If $K(f)$ contains any other element a such that $f(a) = e_{0,}$ we would have a contradiction since f is one-to-one; therefore $K(f) = \{e\}$.

Since an isomorphism is one-to-one and onto, it has an inverse which is a function from $\mathscr{H}$ onto $\mathscr{G}$. The next theorem shows that the inverse is also an isomorphism.

Theorem 6.5. If $f : \mathscr{G} \to \mathscr{H}$ is an isomorphism, then $f^{-1} : \mathscr{H} \to \mathscr{G}$ is an isomorphism.

Proof Since f is one-to-one and onto, it follows that f^{-1} is also one-to-one and onto (cf. Section 3). Let a_0 be the element of $\mathscr{H}$ such that $a = f^{-1}(a_0)$ for any $a \in \mathscr{G}$; then $a * b = f^{-1}(a_0) * f^{-1}(b_0)$. But $a * b$ is the inverse image of the element $a_0 \circ b_0 \in \mathscr{H}$ because $f(a * b) = f(a) \circ f(b) = a_0 * b_0$ since f is a homomorphism. Therefore $f^{-1}(a_0) * f^{-1}(b_0) = f^{-1}(a_0 \circ b_0)$, which shows that f^{-1} satisfies the definition (6.1) of a homomorphism.

Theorem 6.6. A homomorphism $f : \mathscr{G} \to \mathscr{H}$ is an isomorphism if it is onto and if its kernel contains only the identity element of $\mathscr{G}$.

The proof of this theorem is a trivial consequence of Theorem 6.4 and the definition of an isomorphism.

Exercises

6.1 If $f : \mathscr{G} \to \mathscr{H}$ and $g : \mathscr{H} \to \mathscr{M}$ are homomorphisms, then show that the composition of the mappings f and g is a homomorphism from $\mathscr{G}$ to $\mathscr{M}$.

6.2 Prove Theorems 6.3 and 6.6.

6.3 Show that the logarithm function is an isomorphism from the positive real numbers under multiplication to all the real numbers under addition. What is the inverse of this isomorphism?

6.4 Show that the function $f : f : (\mathscr{R}, +) \to (\mathscr{R}^+, \cdot)$ defined by $f(x) = x^2$ is not a homomorphism.

Section 7. Rings and Fields

Groups and semigroups are important building blocks in developing algebraic structures. In this section we shall consider sets that form a group with respect to one binary operation and a semigroup with respect to another. We shall also consider a set that is a group with respect to two different binary operations.

Definition. A *ring* is a triple $(\mathscr{D},+,\cdot)$ consisting of a set $\mathscr{D}$ and two binary operations $+$ and $\cdot$ such that

(a) $\mathscr{D}$ with the operation $+$ is an Abelian group.

(b) The operation $\cdot$ is associative.

(c) $\mathscr{D}$ contains an identity element, denoted by 1, with respect to the operation $\cdot$, i.e.,

$$1\cdot a = a\cdot 1 = a$$

for all $a \in \mathscr{D}$.

(d) The operations $+$ and $\cdot$ satisfy the distributive axioms

$$\begin{aligned} a\cdot(b+c) &= a\cdot b + a\cdot c \\ (b+c)\cdot a &= b\cdot a + c\cdot a \end{aligned} \qquad \textbf{(7.1)}$$

The operation + is called *addition* and the operation $\cdot$ is called *multiplication*. As was done with the notation for a group, the ring $(\mathscr{D},+,\cdot)$ will be written simply as $\mathscr{D}$. Axiom (a) requires that $\mathscr{D}$ contains the identity element for the $+$ operation. We shall follow the usual procedure and denote this element by 0. Thus, $a+0=0+a=a$. Axiom (a) also requires that each element $a\in\mathscr{D}$ have an additive inverse, which we will denote by $-a$, $a+(-a)=0$. The quantity $a+b$ is called the *sum* of a and b, and the *difference* between a and b is $a+(-b)$, which is usually written as $a-b$. Axiom (b) requires the set $\mathscr{D}$ and the multiplication operation to form a semigroup. If the multiplication operation is also commutative, the ring is called a *commutative ring*. Axiom (c) requires that $\mathscr{D}$ contain an identity element for multiplication; this element is called the *unity* element of the ring and is denoted by 1. The symbol 1 should not be confused with the real number one. The existence of the unity element is sometimes omitted in the ring axioms and the ring as we have defined it above is called the *ring with unity*. Axiom (d), the distributive axiom, is the only idea in the definition of a ring that has not appeared before. It provides a rule for the interchanging of the two binary operations.

The following familiar systems are all examples of rings.

1. The set $\mathscr{I}$ of integers with the ordinary addition and multiplication operations form a ring.
2. The set $\mathscr{R}_a$ of rational numbers with the usual addition and multiplication operations form a ring.
3. The set $\mathscr{R}$ of real numbers with the usual addition and multiplication operations form a ring.
4. The set $\mathscr{C}$ of complex numbers with the usual addition and multiplication operations form a ring.
5. The set of all *N* by *N* matrices form a ring with respect to the operations of matrix addition and matrix multiplication. The unity element is the unity matrix and the zero element is the zero matrix.
6. The set of all polynomials in one (or several) variable with real (or complex) coefficients form a ring with respect to the usual operations of addition and multiplication.

Many properties of rings can be deduced from similar results that hold for groups. For example, the zero element 0, the unity element 1, the negative element $-a$ of an element a are unique. Properties that are associated with the interconnection between the two binary operations are contained in the following theorems:

Theorem 7.1. For any element $a \in \mathscr{D}$, $a \cdot 0 = 0 \cdot a = 0$.

Proof. For any $b \in \mathscr{D}$ we have by Axiom (a) that $b + 0 = b$. Thus for any $a \in \mathscr{D}$ it follows that $(b + 0) \cdot a = b \cdot a$, and the Axiom (d) permits this equation to be recast in the form $b \cdot a + 0 \cdot a = b \cdot a$. From Property 8 for the additive group this equation implies that $0 \cdot a = 0$. It can be shown that $a \cdot 0 = 0$ by a similar argument.

Theorem 7.2. For all elements $a, b \in \mathscr{D}$

$$(-a) \cdot b = a \cdot (-b) = -(a \cdot b)$$

This theorem shows that there is no ambiguity in writing $-a \cdot b$ for $-(a \cdot b)$.

Many of the notions developed for groups can be extended to rings; for example, subrings and ring homomorphisms correspond to the notions of subgroups and group homomorphisms. The interested reader can consult the Selected Reading for a discussion of these ideas.

The set of integers is an example of an algebraic structure called an *integral domain.*

Definition. A ring $\mathscr{D}$ is an *integral domain* if it satisfies the following additional axioms:

(e) The operation $\cdot$ is commutative.

(f) If a, b, c are any elements of $\mathscr{D}$ with $c \neq 0,$ then

$$a \cdot c = b \cdot c \Rightarrow a = b \tag{7.2}$$

The cancellation law of multiplication introduced by Axiom (f) is logically equivalent to the assertion that a product of nonzero factors is nonzero. This is proved in the following theorem.

Theorem 7.3. A commutative ring $\mathscr{D}$ is an integral domain if and only if for all elements $a, b \in \mathscr{D}$, $a \cdot b \neq 0$ unless $a = 0$ or $b = 0.$

Proof. Assume first that $\mathscr{D}$ is an integral domain so the cancellation law holds. Suppose $a \cdot b = 0$ and $b \neq 0.$ Then we can write $a \cdot b = 0 \cdot b$ and the cancellation law implies $a = 0.$ Conversely, suppose that a product in a commutative ring $\mathscr{D}$ cannot be zero unless one of its factors is zero. Consider the expression $a \cdot c = b \cdot c,$ which can be rewritten as $a \cdot c - b \cdot c = 0$ or as $(a - b) \cdot c = 0.$ If $c \neq 0$ then by assumption $a - b = 0$ or $a = b.$ This proves that the cancellation law holds.

The sets of integers $\mathscr{I}$, rational numbers $\mathscr{R}_a$, real numbers $\mathscr{R}$, and complex numbers $\mathscr{C}$ are examples of integral domains as well as being examples of rings. Sets of square matrices, while forming rings with respect to the binary operations of matrix addition and multiplication, do not, however, form integral domains. We can show this in several ways; first by showing that matrix multiplication is not commutative and, second, we can find two nonzero matrices whose product is zero, for example,

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

The rational numbers, the real numbers, and the complex numbers are examples of an algebraic structure called a *field*.

Definition. A field $\mathscr{F}$ is an integral domain, containing more than one element, and such that any element $a \in \mathscr{F} / \{0\}$ has an inverse with respect to multiplication.

It is clear from this definition that the set $\mathscr{F}$ and the addition operation as well as the set $\mathscr{F}/\{0\}$ and the multiplication operation form Abelian groups. Hence the unity element 1, the zero element 0, the negative $(-a)$, as well as the reciprocal $(1/a), a \neq 0,$ are unique. The formulas of arithmetic can be developed as theorems following from the axioms of a field. It is not our purpose to do this here; however, it is important to convince oneself that it can be done.

The definitions of ring, integral domain, and field are each a step more restrictive than the other. It is trivial to notice that any set that is a field is automatically an integral domain and a ring. Similarly, any set that is an integral domain is automatically a ring.

The dependence of the algebraic structures introduced in this chapter is illustrated schematically in Figure 7.1. The dependence of the vector space, which is to be introduced in the next chapter, upon these algebraic structures is also indicated.

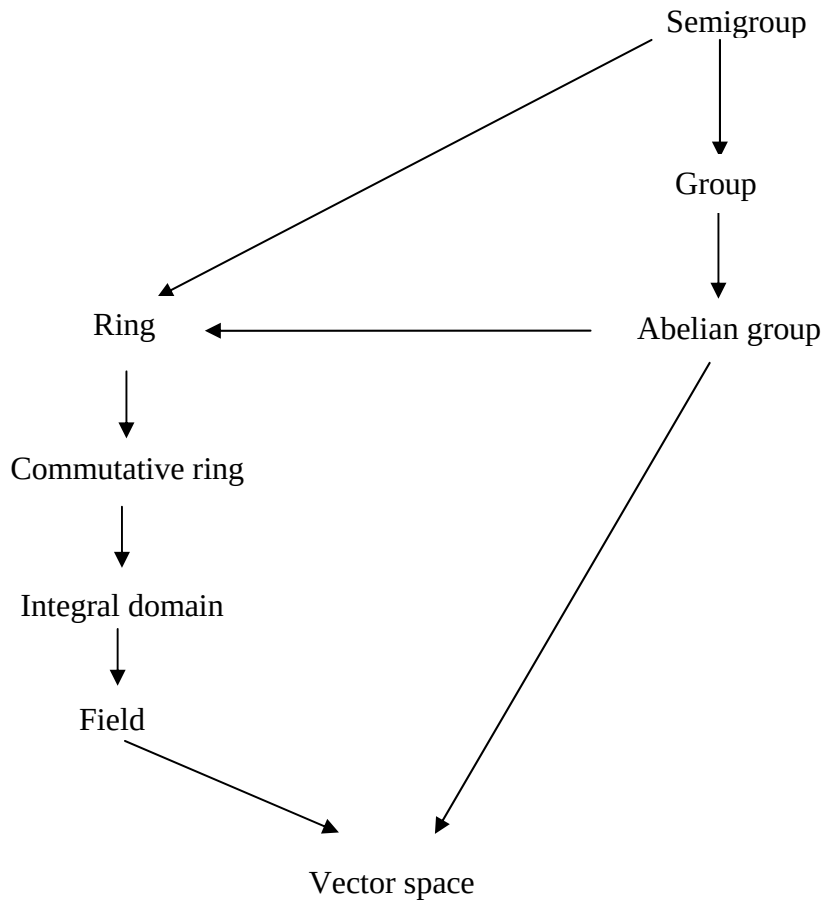

Figure 1. A schematic of algebraic structures.

Exercises

7.1 Verify that Examples 1 and 5 are rings.

7.2 Prove Theorem 7.2.

7.3 Prove that if a, b, c, d are elements of a ring $\mathscr{D}$, then

(a) $(-a)\cdot(-b) = a \cdot b.$

(b) $(-1)\cdot a = a \cdot (-1) = -a.$

(c) $(a+b)\cdot(c+d) = a\cdot c + a\cdot d + b\cdot c + b\cdot d.$

7.4 Among the Examples 1-6 of rings given in the text, which ones are actually integral domains, and which ones are fields?

7.5 Is the set of rational numbers a field?

7.6 Why does the set of integers not constitute a field?

7.7 If $\mathscr{F}$ is a field, why is $\mathscr{F} / \{0\}$ an Abelian group with respect to multiplication?

7.8 Show that for all rational numbers x, y, the set of elements of form $x + y\sqrt{2}$ constitutes a field.

7.9 For all rational numbers x, y, z, does the set of elements of the form $x + y\sqrt{2} + z\sqrt{3}$ form a field? If not, show how to enlarge it to a field.

PART 2

VECTOR AND TENSOR ALGEBRA

Selected Readings for Part II

FRAZER, R. A., W. J. DUNCAN, AND A. R. COLLAR, *Elementary Matrices*, Cambridge University Press, Cambridge, 1938.

GREUB, W. H., *Linear Algebra*, 3rd ed., Springer-Verlag, New York, 1967.

GREUB, W. H., *Multilinear Algebra*, Springer-Verlag, New York, 1967.

HALMOS, P. R., *Finite Dimensional Vector Spaces*, Van Nostrand, Princeton, New Jersey, 1958.

JACOBSON, N., *Lectures in Abstract Algebra*, Vol. II. Linear Algebra, Van Nostrand, New York, 1953.

MOSTOW, G. D., J. H. SAMPSON, AND J. P. MEYER, *Fundamental Structures of Algebra*, McGraw-Hill, New York, 1963.

NOMIZU, K., *Fundamentals of Linear Algebra*, McGraw-Hill, New York, 1966.

SHEPHARD, G. C., *Vector Spaces of Finite Dimensions*, Interscience, New York, 1966.

VEBLEN, O., *Invariants of Quadratic Differential Forms*, Cambridge University Press, Cambridge, 1962.

Chapter 3

VECTOR SPACES

Section 8. The Axioms for a Vector Space

Generally, one's first encounter with the concept of a vector is a geometrical one in the form of a directed line segment, that is to say, a straight line with an arrowhead. This type of vector, if it is properly defined, is a special example of the more general notion of a vector presented in this chapter. The concept of a vector put forward here is purely algebraic. The definition given for a vector is that it be a member of a set that satisfies certain algebraic rules.

A *vector space* is a triple $(\mathscr{V}, \mathscr{F}, f)$ consisting of (a) an additive Abelian group $\mathscr{V}$, (b) a field $\mathscr{F}$, and (c) a function $f : \mathscr{F} \times \mathscr{V} \to \mathscr{V}$ called *scalar multiplication* such that

$$\begin{aligned} f(\lambda, f(\mu, \mathbf{v})) &= f(\lambda\mu, \mathbf{v}) \\ f(\lambda + \mu, \mathbf{u}) &= f(\lambda, \mathbf{u}) + f(\mu, \mathbf{u}) \\ f(\lambda, \mathbf{u} + \mathbf{v}) &= f(\lambda, \mathbf{u}) + f(\lambda, \mathbf{v}) \\ f(1, \mathbf{v}) &= \mathbf{v} \end{aligned} \tag{8.1}$$

for all $\lambda, \mu \in \mathscr{F}$ and all $\mathbf{u}, \mathbf{v} \in \mathscr{V}$. A *vector* is an element of a vector space. The notation $(\mathscr{V}, \mathscr{F}, f)$ for a vector space will be shortened to simply $\mathscr{V}$. The first of (8.1) is usually called the *associative law* for *scalar multiplication*, while the second and third equations are *distributive laws*, the second for *scalar addition* and the third for *vector addition*.

It is also customary to use a simplified notation for the scalar multiplication function f. We shall write $f(\lambda, \mathbf{v}) = \lambda\mathbf{v}$ and also regard $\lambda\mathbf{v}$ and $\mathbf{v}\lambda$ to be identical. In this simplified notation we shall now list in detail the axioms of a vector space. In this definition the vector $\mathbf{u} + \mathbf{v}$ in $\mathscr{V}$ is called the *sum* of $\mathbf{u}$ and $\mathbf{v}$ and the *difference* of $\mathbf{u}$ and $\mathbf{v}$ is written $\mathbf{u} - \mathbf{v}$ and is defined by

$$\mathbf{u} - \mathbf{v} = \mathbf{u} + (-\mathbf{v}) \tag{8.2}$$

Definition. Let $\mathscr{V}$ be a set and $\mathscr{F}$ a field. $\mathscr{V}$ is a *vector space* if it satisfies the following rules:

(a) There exists a binary operation in $\mathscr{V}$ called addition and denoted by $+$ such that

(1) $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$ for all $\mathbf{u},\mathbf{v},\mathbf{w}\in\mathscr{V}$.

(2) $\mathbf{u}+\mathbf{v}=\mathbf{v}+\mathbf{u}$ for all $\mathbf{u},\mathbf{v}\in\mathscr{V}$.

(3) There exists an element $\mathbf{0}\in\mathscr{V}$ such that $\mathbf{u}+\mathbf{0}=\mathbf{u}$ for all $\mathbf{u}\in\mathscr{V}$.

(4) For every $\mathbf{u}\in\mathscr{V}$ there exists an element $-\mathbf{u}\in\mathscr{V}$. such that $\mathbf{u}+(-\mathbf{u})=\mathbf{0}$.

(b) There exists an operation called scalar multiplication in which every scalar $\lambda\in\mathscr{F}$ can be combined with every element $\mathbf{u}\in\mathscr{V}$ to give an element $\lambda\mathbf{u}\in\mathscr{V}$ such that

(1) $\lambda(\mu\mathbf{u})=(\lambda\mu)\mathbf{u}$

(2) $(\lambda+\mu)\mathbf{u}=\lambda\mathbf{u}+\mu\mathbf{u}$

(3) $\lambda(\mathbf{u}+\mathbf{v})=\lambda\mathbf{u}+\lambda\mathbf{v}$

(4) $1\mathbf{u}=\mathbf{u}$

for all $\lambda,\mu\in\mathscr{F}$ and all $\mathbf{u},\mathbf{v}\in\mathscr{V}$.

If the field $\mathscr{F}$ employed in a vector space is actually the field of real numbers $\mathscr{R}$, the space is called a *real vector space*. A *complex vector space* is similarly defined.

Except for a few minor examples, the material in this chapter and the next three employs complex vector spaces. Naturally the real vector space is a trivial special case. The reason for allowing the scalar field to be complex in these chapters is that the material on spectral decompositions in Chapter 6 has more usefulness in the complex case. After Chapter 6 we shall specialize the discussion to real vector spaces. Therefore, unless we provide some qualifying statement to the contrary, a *vector space* should be understood to be a complex vector space.

There are many and varied sets of objects that qualify as vector spaces. The following is a list of examples of vector spaces:

1. The vector space $\mathscr{C}^N$ is the set of all *N*-tuples of the form $\mathbf{u}=(\lambda_1,\lambda_2,...,\lambda_N)$, where *N* is a positive integer and $\lambda_1,\lambda_2,...,\lambda_N\in\mathscr{C}$. Since an *N*-tuple is an ordered set, if $\mathbf{v}=(\mu_1,\mu_2,...,\mu_N)$ is a second *N*-tuple, then $\mathbf{u}$ and $\mathbf{v}$ are equal if and only if

$$\mu_k=\lambda_k \quad \text{for all } k=1,2,...,N$$

The zero *N*-tuple is $\mathbf{0}=(0,0,...,0)$ and the negative of the *N*-tuple $\mathbf{u}$ is $-\mathbf{u}=(-\lambda_1,-\lambda_2,...,-\lambda_N)$. Addition and scalar multiplication of *N*-tuples are defined by the formulas

$$\mathbf{u}+\mathbf{v}=(\mu_1+\lambda_1,\mu_2+\lambda_2,...,\mu_N+\lambda_N)$$

and

$$\mu\mathbf{u} = \left(\mu\lambda_1, \mu\lambda_2, ..., \mu\lambda_N\right)$$

respectively. The notation $\mathscr{C}^N$ is used for this vector space because it can be considered to be an *Nth* Cartesian product of $\mathscr{C}$.

2. The set $\mathscr{V}$ of all $N \times M$ complex matrices is a vector space with respect to the usual operation of matrix addition and multiplication by a scalar.
3. Let $\mathscr{H}$ be a vector space whose vectors are actually functions defined on a set $\mathscr{A}$ with values in $\mathscr{C}$. Thus, if $\mathbf{h} \in \mathscr{H}$, $x \in \mathscr{A}$ then $\mathbf{h}(x) \in \mathscr{C}$ and $\mathbf{h}: \mathscr{A} \to \mathscr{C}$. If $\mathbf{k}$ is another vector of $\mathscr{H}$ then equality of vectors is defined by

$$\mathbf{h} = \mathbf{k} \Leftrightarrow \mathbf{h}(x) = \mathbf{k}(x) \text{ for all } x \in \mathscr{A}$$

The zero vector is the zero function whose value is zero for all x. Addition and scalar multiplication are defined by

$$(\mathbf{h} + \mathbf{k})(x) = \mathbf{h}(x) + \mathbf{k}(x)$$

and

$$(\lambda\mathbf{h})(x) = \lambda(\mathbf{h}(x))$$

respectively.

4. Let $\mathscr{P}$ denote the set of all polynomials *u* of degree *N* of the form

$$u = \lambda_0 + \lambda_1 x + \lambda_2 x^2 + \cdots + \lambda_N x^N$$

where $\lambda_0, \lambda_1, \lambda_2, ..., \lambda_N \in \mathscr{C}$. The set $\mathscr{P}$ forms a vector space over the complex numbers $\mathscr{C}$ if addition and scalar multiplication of polynomials are defined in the usual way.

5. The set of complex numbers $\mathscr{C}$, with the usual definitions of addition and multiplication by a real number, forms a real vector space.
6. The zero element $\mathbf{0}$ of any vector space forms a vector space by itself.

The operation of scalar multiplication is not a binary operation as the other operations we have considered have been. In order to develop some familiarity with the algebra of this operation consider the following three theorems.

Theorem 8.1. $\lambda\mathbf{u} = \mathbf{0} \Leftrightarrow \lambda = 0$ or $\mathbf{u} = \mathbf{0}$.

Proof. The proof of this theorem actually requires the proof of the following three assertions:

(a) $0\mathbf{u} = \mathbf{0}$, (b) $\lambda\mathbf{0} = \mathbf{0}$, (c) $\lambda\mathbf{u} = \mathbf{0} \Rightarrow \lambda = 0$ or $\mathbf{u} = \mathbf{0}$

To prove (a), take $\mu = 0$ in Axiom (b2) for a vector space; then $\lambda\mathbf{u} = \lambda\mathbf{u} + 0\mathbf{u}$. Therefore

$$\lambda\mathbf{u} - \lambda\mathbf{u} = \lambda\mathbf{u} - \lambda\mathbf{u} + 0\mathbf{u}$$

and by Axioms (a4) and (a3)

$$\mathbf{0} = \mathbf{0} + 0\mathbf{u} = 0\mathbf{u}$$

which proves (a).

To prove (b), set $\mathbf{v} = \mathbf{0}$ in Axiom (b3) for a vector space; then $\lambda\mathbf{u} = \lambda\mathbf{u} + \lambda\mathbf{0}$. Therefore

$$\lambda\mathbf{u} - \lambda\mathbf{u} = \lambda\mathbf{u} - \lambda\mathbf{u} + \lambda\mathbf{0}$$

and by Axiom (a4)

$$\mathbf{0} = \mathbf{0} + \lambda\mathbf{0} = \lambda\mathbf{0}$$

To prove (c), we assume $\lambda\mathbf{u} = \mathbf{0}$. If $\lambda = 0$, we know from (a) that the equation $\lambda\mathbf{u} = \mathbf{0}$ is satisfied. If $\lambda \neq 0$, then we show that $\mathbf{u}$ must be zero as follows:

$$\mathbf{u} = 1\mathbf{u} = \lambda\left(\frac{1}{\lambda}\right)\mathbf{u} = \frac{1}{\lambda}(\lambda\mathbf{u}) = \frac{1}{\lambda}(\mathbf{0}) = \mathbf{0}$$

Theorem 8.2. $(-\lambda)\mathbf{u} = -\lambda\mathbf{u}$ for all $\lambda \in \mathscr{C}, \mathbf{u} \in \mathscr{V}$.

Proof. Let $\mu = 0$ and replace λ by $-\lambda$ in Axiom (b2) for a vector space and this result follows directly.

Theorem 8.3. $-\lambda\mathbf{u} = \lambda(-\mathbf{u})$ for all $\lambda \in \mathscr{C}, \mathbf{u} \in \mathscr{V}$.

Finally, we note that the concepts of *length* and *angle* have not been introduced. They will be introduced in a later section. The reason for the delay in their introduction is to emphasize that certain results can be established without reference to these concepts.

Exercises

8.1 Show that at least one of the axioms for a vector space is redundant. In particular, one might show that Axiom (a2) can be deduced from the remaining axioms. [Hint: expand $(1+1)(\mathbf{u}+\mathbf{v})$ by the two different rules.]

8.2 Verify that the sets listed in Examples 1- 6 are actually vector spaces. In particular, list the zero vectors and the negative of a typical vector in each example.

8.3 Show that the axioms for a vector space still make sense if the field $\mathscr{F}$ is replaced by a ring. The resulting structure is called a *module over a ring*.

8.4 Show that the Axiom (b4) for a vector space can be replaced by the axiom

$$(b4)' \quad \lambda\mathbf{u} = \mathbf{0} \Leftrightarrow \lambda = 0 \text{ or } \mathbf{u} = \mathbf{0}$$

8.5 Let $\mathscr{V}$ and $\mathscr{U}$ be vector spaces. Show that the set $\mathscr{V} \times \mathscr{U}$ is a vector space with the definitions

$$(\mathbf{u},\mathbf{x}) + (\mathbf{v},\mathbf{y}) = (\mathbf{u}+\mathbf{v},\mathbf{x}+\mathbf{y})$$

and

$$\lambda(\mathbf{u},\mathbf{x}) = (\lambda\mathbf{u},\lambda\mathbf{x})$$

where $\mathbf{u},\mathbf{v} \in \mathscr{V}$; $\mathbf{x},\mathbf{y} \in \mathscr{U}$; and $\lambda \in \mathscr{F}$

8.6 Prove Theorem 8.3.

8.7 Let $\mathscr{V}$ be a vector space and consider the set $\mathscr{V} \times \mathscr{V}$. Define addition in $\mathscr{V} \times \mathscr{V}$ by

$$(\mathbf{u},\mathbf{v}) + (\mathbf{x},\mathbf{y}) = (\mathbf{u}+\mathbf{x},\mathbf{v}+\mathbf{y})$$

and multiplication by complex numbers by

$$(\lambda + i\mu)(\mathbf{u},\mathbf{v}) = (\lambda\mathbf{u} - \mu\mathbf{v}, \mu\mathbf{u} + \lambda\mathbf{v})$$

where $\lambda, \mu \in \mathscr{R}$. Show that $\mathscr{V} \times \mathscr{V}$ is a vector space over the field of complex numbers.

Section 9. Linear Independence, Dimension, and Basis

The concept of *linear independence* is introduced by first defining what is meant by *linear dependence* in a set of vectors and then defining a set of vectors that is not linearly dependent to be linearly independent. The general definition of linear dependence of a set of *N* vectors is an algebraic generalization and abstraction of the concepts of collinearity from elementary geometry.

Definition. A finite set of $N\,(N \geq 1)$ vectors $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_N\}$ in a vector space $\mathscr{V}$ is said to be *linearly dependent* if there exists a set of scalars $\{\lambda^1, \lambda^2, ..., \lambda^N\}$, not all zero, such that

$$\sum_{j=1}^{N} \lambda^j \mathbf{v}_j = \mathbf{0} \tag{9.1}$$

The essential content of this definition is that at least one of the vectors $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_N\}$ can be expressed as a linear combination of the other vectors. This means that if $\lambda^1 \neq 0,$ then $\mathbf{v}_1 = \sum_{i=2}^{N} \mu^j \mathbf{v}_j$,where $\mu^j = -\lambda^j / \lambda^1$ for $j = 2, 3, ..., N.$ As a numerical example, consider the two vectors

$$\mathbf{v}_1 = (1, 2, 3), \quad \mathbf{v}_2 = (3, 6, 9)$$

from $\mathscr{R}^3$. These vectors are linearly dependent since

$$\mathbf{v}_2 = 3\mathbf{v}_1$$

The proof of the following two theorems on linear dependence is quite simple.

Theorem 9.1. If the set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_N\}$ is linearly dependent, then every other finite set of vectors containing $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_N\}$ is linearly dependent.

Theorem 9.2. Every set of vectors containing the zero vector is linearly dependent.

A set of $N\,(N \geq 1)$ vectors that is not linearly dependent is said to be *linearly independent*. Equivalently, a set of $N\,(N \geq 1)$ vectors $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_N\}$ is linearly independent if (9.1) implies $\lambda^1 = \lambda^2 = ... = \lambda^N = 0$. As a numerical example, consider the two vectors $\mathbf{v}_1 = (1,2)$ and $\mathbf{v}_2 = (2,1)$ from $\mathscr{R}^2$. These two vectors are linearly independent because

$$\lambda \mathbf{v}_1 + \lambda \mathbf{v}_2 = \lambda(1,2) + \mu(2,1) = \mathbf{0} \Leftrightarrow \lambda + 2\mu = 0, 2\lambda + \mu = 0$$

and

$$\lambda + 2\mu = 0, \quad 2\lambda + \mu = 0 \Leftrightarrow \lambda = 0, \quad \mu = 0$$

Theorem 9.3. Every non empty subset of a linearly independent set is linearly independent.

A linearly independent set in a vector space is said to be *maximal* if it is not a proper subset of any other linearly independent set. A vector space that contains a (finite) maximal, linearly independent set is then said to be *finite dimensional*. Of course, if $\mathscr{V}$ is not finite dimensional, then it is called *infinite dimensional*. In this text we shall be concerned only with finite-dimensional vector spaces.

Theorem 9.4. Any two maximal, linearly independent sets of a finite-dimensional vector space must contain exactly the same number of vectors.

Proof. Let $\{\mathbf{v}_1, ..., \mathbf{v}_N\}$ and $\{\mathbf{u}_1, ..., \mathbf{u}_M\}$ be two maximal, linearly independent sets of $\mathscr{V}$. Then we must show that $N = M$. Suppose that $N \neq M$, say $N < M$. By the fact that $\{\mathbf{v}_1, ..., \mathbf{v}_N\}$ is maximal, the sets $\{\mathbf{v}_1, ..., \mathbf{v}_N, \mathbf{u}_1\}, ..., \{\mathbf{v}_1, ..., \mathbf{v}_N, \mathbf{u}_M\}$ are all linearly dependent. Hence there exist relations

$$\begin{aligned} \lambda_{11}\mathbf{v}_1 + \cdots + \lambda_{1N}\mathbf{v}_N + \mu_1 \mathbf{u}_1 &= \mathbf{0} \\ &\vdots \\ \lambda_{M1}\mathbf{v}_1 + \cdots + \lambda_{MN}\mathbf{v}_N + \mu_M \mathbf{u}_M &= \mathbf{0} \end{aligned} \tag{9.2}$$

where the coefficients of each equation are not all equal to zero. In fact, the coefficients $\mu_1, ..., \mu_M$ of the vectors $\mathbf{u}_1, ..., \mathbf{u}_M$ are all nonzero, for if $\mu_i = 0$ for any i, then

$$\lambda_{i1}\mathbf{v}_1 + \cdots + \lambda_{iN}\mathbf{v}_N = \mathbf{0}$$

for some nonzero $\lambda_{i1},...,\lambda_{iN}$ contradicting the assumption that $\{\mathbf{v}_1,...,\mathbf{v}_N\}$ is a linearly independent set. Hence we can solve (9.2) for the vectors $\{\mathbf{u}_1,....,\mathbf{u}_M\}$ in terms of the vectors $\{\mathbf{v}_1,...,\mathbf{v}_N\}$, obtaining

$$\begin{aligned} \mathbf{u}_1 &= \mu_{11}\mathbf{v}_1 + \cdots + \mu_{1N}\mathbf{v}_N \\ &\vdots \\ \mathbf{u}_M &= \mu_{M1}\mathbf{v}_1 + \cdots + \mu_{MN}\mathbf{v}_N \end{aligned} \tag{9.3}$$

where $\mu_{ij} = -\lambda_{ij} / \mu_i$ for $i = 1,...,M; j = 1,...,N$.

Now we claim that the first *N* equations in the above system can be inverted in such a way that the vectors $\{\mathbf{v}_1,...,\mathbf{v}_N\}$ are given by linear combinations of the vectors $\{\mathbf{u}_1,....,\mathbf{u}_N\}$. Indeed, inversion is possible if the coefficient matrix $\left[\mu_{ij}\right]$ for $i, j = 1,...,N$ is nonsingular. But this is clearly the case, for if that matrix were singular, there would be nontrivial solutions $\{\alpha_1,....,\alpha_N\}$ to the linear system

$$\sum_{i=1}^{N} \alpha_i \mu_{ij} = 0, \qquad j = 1,...,N \tag{9.4}$$

Then from (9.3) and (9.4) we would have

$$\sum_{i=1}^{N} \alpha_i \mathbf{u}_i = \sum_{j=1}^{N} \left(\sum_{i=1}^{N} \alpha_i \mu_{ij} \right) \mathbf{v}_j = \mathbf{0}$$

contradicting the assumption that set $\{\mathbf{u}_1,...,\mathbf{u}_N\}$,being a subset of the linearly independent set $\{\mathbf{u}_1,...,\mathbf{u}_M\}$, is linearly independent.

Now if the inversion of the first *N* equations of the system (9.3) gives

$$\begin{aligned} \mathbf{v}_1 &= \xi_{11}\mathbf{u}_1 + \cdots + \xi_{1N}\mathbf{u}_N \\ &\vdots \\ \mathbf{v}_N &= \xi_{N1}\mathbf{u}_1 + ... + \xi_{NN}\mathbf{u}_N \end{aligned} \tag{9.5}$$

where $\left[\xi_{ij}\right]$ is the inverse of $\left[\mu_{ij}\right]$ for $i, j = 1,...,N$,we can substitute (9.5) into the remaining $M - N$ equations in the system (9.3), obtaining

$$\begin{aligned}\mathbf{u}_{N+1} &= \sum_{j=1}^{N}\left(\sum_{i=1}^{N}\mu_{N+1\,j}\xi_{ji}\right)\mathbf{u}_i\\ &\vdots\\ \mathbf{u}_M &= \sum_{j=1}^{N}\left(\sum_{i=1}^{N}\mu_{Mj}\xi_{ji}\right)\mathbf{u}_i\end{aligned}$$

But these equations contradict the assumption that the set $\{\mathbf{u}_1,....,\mathbf{u}_M\}$ is linearly independent. Hence $M > N$ is impossible and the proof is complete.

An important corollary of the preceding theorem is the following.

Theorem 9.5. Let $\{\mathbf{u}_1,....,\mathbf{u}_N\}$ be a maximal linearly independent set in $\mathscr{V}$, and suppose that $\{\mathbf{v}_1,...,\mathbf{v}_N\}$ is given by (9.5). Then $\{\mathbf{v}_1,...,\mathbf{v}_N\}$ is also a maximal, linearly independent set if and only if the coefficient matrix $\left[\xi_{ij}\right]$ in (9.5) is nonsingular. In particular, if $\mathbf{v}_i = \mathbf{u}_i$ for $i = 1,...,k-1,\ k+1,...,N$ but $\mathbf{v}_k \neq \mathbf{u}_k$, then $\{\mathbf{u}_1,...,\mathbf{u}_{k-1},\mathbf{v}_k,\mathbf{u}_{k+1},...,\mathbf{u}_N\}$ is a maximal, linearly independent set if and only if the coefficient ξ_{kk} in the expansion of $\mathbf{u}_k$ in terms of $\{\mathbf{v}_1,...,\mathbf{v}_N\}$ in (9.5) is nonzero.

From the preceding theorems we see that the number N of vectors in a maximal, linearly independent set in a finite dimensional vector space $\mathscr{V}$ is an intrinsic property of $\mathscr{V}$. We shall call this number *N* the *dimension* of $\mathscr{V}$, written dim $\mathscr{V}$, namely $N = \dim\ \mathscr{V}$, and we shall call any maximal, linearly independent set of $\mathscr{V}$ a *basis* of that space. Theorem 9.5 characterizes all bases of $\mathscr{V}$ as soon as one basis is specified. A list of examples of bases for vector spaces follows:

1. The set of *N* vectors

$$\begin{gathered}(1,0,0,...,0)\\ (0,1,0,...,0)\\ (0,0,1,...,0)\\ \vdots\\ \underbrace{(0,0,0,...,1)}_{N}\end{gathered}$$

is linearly independent and constitutes a basis for C^N, called the *standard* basis.

2. If $\mathscr{M}^{2\times 2}$ denotes the vector space of all 2×2 matrices with elements from the complex numbers $\mathscr{C}$, then the four matrices

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

form a basis for $\mathscr{M}^{2\times 2}$ called the *standard* basis.

3. The elements 1 and $i = \sqrt{-1}$ form a basis for the vector space $\mathscr{C}$ of complex numbers over the field of real numbers.

The following two theorems concerning bases are fundamental.

Theorem 9.6. If $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_N\}$ is a basis for $\mathscr{V}$, then *every* vector in $\mathscr{V}$ has the representation

$$\mathbf{v} = \sum_{j=1}^{N} \xi^j \mathbf{e}_j \tag{9.6}$$

where $\{\xi^1, \xi^2, ..., \xi^N\}$ are elements of $\mathscr{C}$ which depend upon the vector $\mathbf{v} \in \mathscr{V}$.

The proof of this theorem is contained in the proof of Theorem 9.4.

Theorem 9.7. The N scalars $\{\xi^1, \xi^2, ..., \xi^N\}$ in (9.6) are unique.

Proof. As is customary in the proof of a uniqueness theorem, we assume a lack of uniqueness. Thus we say that $\mathbf{v}$ has two representations,

$$\mathbf{v} = \sum_{j=1}^{N} \xi^j \mathbf{e}_j, \quad \mathbf{v} = \sum_{j=1}^{N} \mu^j \mathbf{e}_j$$

Subtracting the two representations, we have

$$\sum_{j=1}^{N} (\xi^j - \mu^j) \mathbf{e}_j = \mathbf{0}$$

and the linear independence of the basis requires that

$$\xi^j = \mu^j, \quad j = 1,2,...,N$$

The coefficients $\{\xi^1, \xi^2, ..., \xi^N\}$ in the representation (9.6) are the *components* of $\mathbf{v}$ with respect to the basis $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_N\}$. The representation of vectors in terms of the elements of their basis is illustrated in the following examples.

1. The vector space $\mathscr{C}^3$ has a standard basis consisting of $\mathbf{e}_1, \mathbf{e}_2,$ and $\mathbf{e}_3$;

$$\mathbf{e}_1 = (1,0,0), \quad \mathbf{e}_2 = (0,1,0), \quad \mathbf{e}_3 = (0,0,1)$$

A vector $\mathbf{v} = (2+i, 7, 8+3i)$ can be written as

$$\mathbf{v} = (2+i)\mathbf{e}_1 + 7\mathbf{e}_2 + (8+3i)\mathbf{e}_3$$

2. The vector space of complex numbers $\mathscr{C}$ over the space of real numbers $\mathscr{R}$ has the basis $\{1, i\}$. Any complex number z can then be represented by

$$z = \mu + \lambda i$$

where $\mu, \lambda \in \mathscr{R}$.

3. The vector space $\mathscr{M}^{2\times 2}$ of all $2x2$ matrices with elements from the complex numbers $\mathscr{C}$ has the standard basis

$$\mathbf{e}_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{e}_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \mathbf{e}_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{e}_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Any $2x2$ matrix of the form

$$\mathbf{v} = \begin{bmatrix} \mu & \lambda \\ \nu & \xi \end{bmatrix}$$

where $\mu, \lambda, \nu, \xi \in \mathscr{C},$ can then be represented by

$$\mathbf{v} = \mu\mathbf{e}_{11} + \lambda\mathbf{e}_{12} + \nu\mathbf{e}_{21} + \xi\mathbf{e}_{22}$$

The basis for a vector space is not unique and the general rule of change of basis is given by Theorem 9.5. An important special case of that rule is made explicit in the following *exchange theorem.*

Theorem 9.8. If $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_N\}$ is a basis for $\mathscr{V}$ and if $\mathscr{B} = \{\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_k\}$ is a linearly independent set of $K (N \geq K)$ in $\mathscr{V}$, then it is possible to exchange a certain K of the original base vectors with $\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_k$ so that the new set is a basis for $\mathscr{V}$.

Proof. We select $\mathbf{b}_1$ from the set $\mathscr{B}$ and order the basis vectors such that the component ξ^1 of $\mathbf{b}_1$ is not zero in the representation

$$\mathbf{b}_1 = \sum_{j=1}^{N} \xi^j \mathbf{e}_j$$

By Theorem 9.5, the vectors $\{\mathbf{b}_1, \mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_N\}$ form a basis for $\mathscr{V}$. A second vector $\mathbf{b}_2$ is selected from $\mathscr{V}$ and we again order the basis vectors so that this time the component λ^2 is not zero in the formula

$$\mathbf{b}_2 = \lambda^1 \mathbf{b}_1 + \sum_{j=2}^{N} \lambda^j \mathbf{e}_j$$

Again, by Theorem 9.5 the vectors $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{e}_3, ..., \mathbf{e}_N\}$ form a basis for $\mathscr{V}$. The proof is completed by simply repeating the above construction $K - 2$ times.

We now know that when a basis for $\mathscr{V}$ is given, every vector in $\mathscr{V}$ has a representation in the form (9.6). Inverting this condition somewhat, we now want to consider a set of vectors $\mathscr{B} = \{\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_M\}$ of $\mathscr{V}$ with the property that every vector $\mathbf{v} \in \mathscr{V}$ can be written as

$$\mathbf{v} = \sum_{j=1}^{M} \lambda^j \mathbf{b}_j$$

Such a. set is called a *generating* set of $\mathscr{V}$ (or is said to *span* $\mathscr{V}$). In some sense a generating set is a counterpart of a linearly independent set. The following theorem is the counter part of Theorem 9.1.

Theorem 9.9. If $\mathscr{B} = \{\mathbf{b}_1,...,\mathbf{b}_M\}$ is a generating set of $\mathscr{V}$, then every other finite set of vectors containing $\mathscr{B}$ is also a generating set of $\mathscr{V}$.

In view of this theorem, we see that the counterpart of a maximal, linearly independent set is a minimal generating set, which is defined by the condition that a generating set $\{\mathbf{b}_1,...,\mathbf{b}_M\}$ is *minimal* if it contains no proper subset that is still a generating set. The following theorem shows the relation between a maximal, linearly independent set and a minimal generating set.

Theorem 9.10. Let $\mathscr{B} = \{\mathbf{b}_1,...,\mathbf{b}_M\}$ be a finite subset of a finite dimensional vector space $\mathscr{V}$. Then the following conditions are equivalent:

(*i*) $\mathscr{B}$ is a maximal linearly independent set.
(*ii*) $\mathscr{B}$ is a linearly independent generating set.
(*iii*) $\mathscr{B}$ is a minimal generating set.

Proof. We shall show that $(i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (i)$.

$(i) \Rightarrow (ii)$. This implication is a direct consequence of the representation (9.6).

$(ii) \Rightarrow (iii)$. This implication is obvious. For if $\mathscr{B}$ is a linearly independent generating set but not a minimal generating set, then we can remove at least one vector, say $\mathbf{b}_M$, and the remaining set is still a generating set. But this is impossible because $\mathbf{b}_M$, can then be expressed as a linear combination of $\{\mathbf{b}_1,...,\mathbf{b}_{M-1}\}$, contradicting the linear independence of $\mathscr{B}$.

$(iii) \Rightarrow (i)$. If $\mathscr{B}$ is a minimal generating set, then $\mathscr{B}$ must be linearly independent because otherwise one of the vectors of $\mathscr{B}$, say $\mathbf{b}_M$, can be written as a linear combination of the vectors $\{\mathbf{b}_1,...,\mathbf{b}_{M-1}\}$. It follows then that $\{\mathbf{b}_1,...,\mathbf{b}_{M-1}\}$ is still a generating set, contradicting the assumption that $\{\mathbf{b}_1,...,\mathbf{b}_M\}$ is minimal. Now a linearly independent generating set must be maximal, for otherwise there exists a vector $\mathbf{b} \in \mathscr{V}$ such that $\{\mathbf{b}_1,...,\mathbf{b}_M,\mathbf{b}\}$ is linearly independent. Then $\mathbf{b}$ cannot be expressed as a linear combination of $\{\mathbf{b}_1,...,\mathbf{b}_M\}$, thus contradicting the assumption that $\mathscr{B}$ is a generating set.

In view of this theorem, a basis $\mathscr{B}$ can be defined by any one of the three equivalent conditions (*i*), (*ii*), or (*iii*).

Exercises

9.1 In elementary plane geometry it is shown that two straight lines determine a plane if the straight lines satisfy a certain condition. What is the condition? Express the condition in vector notation.

9.2 Prove Theorems 9.1-9.3, and 9.9.

9.3 Let $\mathscr{M}^{3\times3}$ denote the vector space of al1 3×3 matrices with elements from the complex numbers $\mathscr{C}$. Determine a basis for $\mathscr{M}^{3\times3}$

9.4 Let $\mathscr{M}^{2\times2}$ denote the vector space of all 2×2 matrices with elements from the real numbers $\mathscr{R}$. Is either of the following sets a basis for $\mathscr{M}^{2\times2}$?

$$\left\{\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix} \quad \begin{bmatrix}0 & 6\\ 0 & 2\end{bmatrix}, \quad \begin{bmatrix}0 & 1\\ 3 & 0\end{bmatrix}, \quad \begin{bmatrix}0 & 1\\ 6 & 8\end{bmatrix}\right\}$$

$$\left\{\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}, \quad \begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}, \quad \begin{bmatrix}0 & 0\\ 0 & 1\end{bmatrix}, \quad \begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}\right\}$$

9.5 Are the complex numbers $2+4i$ and $6+2i$ linearly independent with respect to the field of real numbers $\mathscr{R}$? Are they linearly independent with respect to the field of complex numbers?

Section 10. Intersection, Sum and Direct Sum of Subspaces

In this section operations such as "summing" and "intersecting" vector spaces are discussed. We introduce first the important concept of a subspace of a vector space in analogy with a subgroup of a group. A non empty subset $\mathcal{U}$ of a vector space $\mathcal{V}$ is a *subspace* if:

(a) $\mathbf{u}, \mathbf{w} \in \mathcal{U} \Rightarrow \mathbf{u} + \mathbf{w} \in \mathcal{U}$ for all $\mathbf{u}, \mathbf{w} \in \mathcal{U}$.
(b) $\mathbf{u} \in \mathcal{U} \Rightarrow \lambda \mathbf{u} \in \mathcal{U}$ for all $\lambda \in \mathcal{C}$.

Conditions (a) and (b) in this definition can be replaced by the equivalent condition:

(a') $\mathbf{u}, \mathbf{w} \in \mathcal{U} \Rightarrow \lambda \mathbf{u} + \mu \mathbf{w} \in \mathcal{U}$ for all $\lambda \in \mathcal{C}$.

Examples of subspaces of vector spaces are given in the following list:

1. The subset of the vector space $\mathcal{C}^N$ of all *N*-tuples of the form $(0, \lambda_2, \lambda_3, ..., \lambda_N)$ is a subspace of $\mathcal{C}^N$.
2. Any vector space $\mathcal{V}$ is a subspace of itself.
3. The set consisting of the zero vector $\{\mathbf{0}\}$ is a subspace of $\mathcal{V}$.
4. The set of real numbers $\mathcal{R}$ can be viewed as a subspace of the real space of complex numbers $\mathcal{C}$.

The vector spaces $\{\mathbf{0}\}$ and $\mathcal{V}$ itself are considered to be *trivial* subspaces of the vector space $\mathcal{V}$. If $\mathcal{U}$ is not a trivial subspace, it is said to be a *proper subspace* of $\mathcal{V}$.

Several properties of subspaces that one would naturally expect to be true are developed in the following theorems.

Theorem 10.1. If $\mathcal{U}$ is a subspace of $\mathcal{V}$, then $\mathbf{0} \in \mathcal{U}$.

Proof. The proof of this theorem follows easily from (b) in the definition of a subspace above by setting $\lambda = 0$.

Theorem 10.2. If $\mathcal{U}$ is a subspace of $\mathcal{V}$, then $\dim \mathcal{U} \leq \dim \mathcal{V}$.

Proof. By Theorem 9.8 we know that any basis of $\mathcal{U}$ can be enlarged to a basis of $\mathcal{V}$; it follows that $\dim \mathcal{U} \leq \dim \mathcal{V}$.

Theorem 10.3. If $\mathcal{U}$ is a subspace of $\mathcal{V}$, then $\dim \mathcal{U} = \dim \mathcal{V}$ if and only if $\mathcal{U} = \mathcal{V}$.

Proof. If $\mathcal{U} = \mathcal{V}$, then clearly $\dim \mathcal{U} = \dim \mathcal{V}$. Conversely, if $\dim \mathcal{U} = \dim \mathcal{V}$, then a basis for $\mathcal{U}$ is also a basis for $\mathcal{V}$. Thus, any vector $\mathbf{v} \in \mathcal{V}$ is also in $\mathcal{U}$, and this implies $\mathcal{U} = \mathcal{V}$.

Operations that combine vector spaces to form other vector spaces are simple extensions of elementary operations defined on sets. If $\mathcal{U}$ and $\mathcal{W}$ are subspaces of a vector space $\mathcal{V}$, the *sum* of $\mathcal{U}$ and $\mathcal{W}$, written $\mathcal{U} + \mathcal{W}$, is the set

$$\mathcal{U} + \mathcal{W} = \{\mathbf{u} + \mathbf{w} \,|\, \mathbf{u} \in \mathcal{U}, \mathbf{w} \in \mathcal{W}\}$$

Similarly, if $\mathcal{U}$ and $\mathcal{W}$ are subspaces of a vector space $\mathcal{V}$, the *intersection* of $\mathcal{U}$ and $\mathcal{W}$, denoted by $\mathcal{U} \cap \mathcal{W}$, is the set

$$\mathcal{U} \cap \mathcal{W} = \{\mathbf{u} \,|\, \mathbf{u} \in \mathcal{U} \text{ and } \mathbf{u} \in \mathcal{W}\}$$

The *union* of two subspaces $\mathcal{U}$ *and* $\mathcal{W}$ *of* $\mathcal{V}$, denoted by $\mathcal{U} \cup \mathcal{W}$, is the set

$$\mathcal{U} \cup \mathcal{W} = \{\mathbf{u} \,|\, \mathbf{u} \in \mathcal{U} \text{ or } \mathbf{u} \in \mathcal{W}\}$$

Some properties of these two operations are stated in the following theorems.

Theorem 10.4. If $\mathcal{U}$ and $\mathcal{W}$ are subspaces of $\mathcal{V}$, then $\mathcal{U} + \mathcal{W}$ is a subspace of $\mathcal{V}$.

Theorem 10.5. If $\mathcal{U}$ and $\mathcal{W}$ are subspaces of $\mathcal{V}$, then $\mathcal{U}$ and $\mathcal{W}$ are also subspaces of $\mathcal{U} + \mathcal{W}$.

Theorem 10.6. If $\mathcal{U}$ and $\mathcal{W}$ are subspaces of $\mathcal{V}$, then the intersection $\mathcal{U} \cap \mathcal{W}$ is a subspace of .

Proof. Let $\mathbf{u}, \mathbf{w} \in \mathcal{U} \cap \mathcal{W}$. Then $\mathbf{u}, \mathbf{w} \in \mathcal{W}$ and $\mathbf{u}, \mathbf{w} \in \mathcal{U}$. Since $\mathcal{U}$ and $\mathcal{W}$ are subspaces, $\mathbf{u} + \mathbf{w} \in \mathcal{W}$ and $\mathbf{u} + \mathbf{w} \in \mathcal{U}$ which means that $\mathbf{u} + \mathbf{w} \in \mathcal{U} \cap \mathcal{W}$ also. Similarly, if $\mathbf{u} \in \mathcal{U} \cap \mathcal{W}$,then for all $\lambda \in \mathcal{R}$. $\lambda \mathbf{u} \in \mathcal{U} \cap \mathcal{W}$.

Theorem l0.7. If $\mathcal{U}$ and $\mathcal{W}$ are subspaces of $\mathcal{V}$, then the union $\mathcal{U} \cup \mathcal{W}$ is not generally a subspace of $\mathcal{V}$.

Proof. Let $\mathbf{u} \in \mathcal{U}, \mathbf{u} \notin \mathcal{W}$, and let $\mathbf{w} \in \mathcal{W}$ and $\mathbf{w} \notin \mathcal{U}$; then $\mathbf{u} + \mathbf{w} \notin \mathcal{U}$ and $\mathbf{u} + \mathbf{w} \notin \mathcal{W}$, which means that $\mathbf{u} + \mathbf{w} \notin \mathcal{U} \cup \mathcal{W}$.

Theorem 10.8. Let $\mathbf{v}$ be a vector in $\mathcal{U} + \mathcal{W}$, where $\mathcal{U}$ and $\mathcal{W}$ are subspaces of $\mathcal{V}$. The decomposition of $\mathbf{v} \in \mathcal{U} + \mathcal{W}$ into the form $\mathbf{v} = \mathbf{u} + \mathbf{w}$, where $\mathbf{u} \in \mathcal{U}$ and $\mathbf{w} \in \mathcal{W}$, is unique if and only if $\mathcal{U} \cap \mathcal{W} = \{\mathbf{0}\}$.

Proof. Suppose there are two ways of decomposing $\mathbf{v}$; for example, let there be a $\mathbf{u}$ and $\mathbf{u}'$ in $\mathcal{U}$ and a $\mathbf{w}$ and $\mathbf{w}'$ in $\mathcal{W}$ such that

$$\mathbf{v} = \mathbf{u} + \mathbf{w} \text{ and } \mathbf{v} = \mathbf{u}' + \mathbf{w}'$$

The decomposition of $\mathbf{v}$ is unique if it can be shown that the vector $\mathbf{b}$,

$$\mathbf{b} = \mathbf{u} - \mathbf{u}' = \mathbf{w}' - \mathbf{w}$$

vanishes. The vector $\mathbf{b}$ is contained in $\mathcal{U} \cap \mathcal{W}$ since $\mathbf{u}$ and $\mathbf{u}'$ are known to be in $\mathcal{U}$, and $\mathbf{w}$ and $\mathbf{w}'$ are known to be in $\mathcal{W}$. Therefore $\mathcal{U} \cap \mathcal{W} = \{\mathbf{0}\}$ implies uniqueness. Conversely, if we have uniqueness, then $\mathcal{U} \cap \mathcal{W} = \{\mathbf{0}\}$, for otherwise any nonzero vector $\mathbf{y} \in \mathcal{U} \cap \mathcal{W}$ has at least two decompositions $\mathbf{y} = \mathbf{y} + \mathbf{0} = \mathbf{0} + \mathbf{y}$.

The sum of two subspaces $\mathcal{U}$ and $\mathcal{W}$ is called the *direct sum* of $\mathcal{U}$ and $\mathcal{W}$ and denoted by $\mathcal{U} \oplus \mathcal{W}$ if $\mathcal{U} \cap \mathcal{W} = \{\mathbf{0}\}$. This definition is motivated by the result proved in Theorem 10.8. If $\mathcal{U} \oplus \mathcal{W} = \mathcal{V}$, then $\mathcal{U}$ is called the *direct complement* of $\mathcal{W}$ in $\mathcal{V}$. The operation of direct summing can be extended to a finite number of subspaces $\mathcal{V}_1, \mathcal{V}_2, ..., \mathcal{V}_Q$ *of* $\mathcal{V}$. The direct sum $\mathcal{V}_1 \oplus \mathcal{V}_2 \oplus \cdots \oplus \mathcal{V}_Q$ is required to satisfy the conditions

$$\mathcal{V}_R \cap \sum_{K=1}^{K=R-1} \mathcal{V}_K + \mathcal{V}_R \cap \sum_{K=R+1}^{Q} \mathcal{V}_K = \{\mathbf{0}\} \quad \textit{for } R = 1, 2, \ldots, Q$$

The concept of a direct sum of subspaces is an important tool for the study of certain concepts in geometry, as we shall see in later chapters. The following theorem shows that the dimension of a direct sum of subspaces is equal to the sum of the dimensions of the subspaces.

Theorem 10.9. If $\mathscr{U}$ and $\mathscr{W}$ are subspaces of $\mathscr{V}$, then

$$\dim \mathscr{U} \oplus \mathscr{W} = \dim \mathscr{U} + \dim \mathscr{W}$$

Proof. Let $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_R\}$ be a basis for $\mathscr{U}$ and $\{\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_Q\}$ be a basis for $\mathscr{W}$.. Then the set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_R, \mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_Q\}$ is linearly independent since $\mathscr{U} \cap \mathscr{W} = \{\mathbf{0}\}$. This set of vectors generates $\mathscr{U} \oplus \mathscr{W}$ since for any $\mathbf{v} \in \mathscr{U} \oplus \mathscr{W}$ we can write

$$\mathbf{v} = \mathbf{u} + \mathbf{w} = \sum_{j=1}^{R} \lambda^j \mathbf{u}_j + \sum_{j=1}^{Q} \mu^j \mathbf{w}_j$$

where $\mathbf{u} \in \mathscr{U}$ and $\mathbf{w} \in \mathscr{W}$. Therefore by Theorem 9.10, $\dim \mathscr{U} \oplus \mathscr{W} = R + Q$.

The result of this theorem can easily be generalized to

$$\dim\left(\mathscr{V}_1 \oplus \mathscr{V}_2 \oplus \cdots \oplus \mathscr{V}_Q\right) = \sum_{j=1}^{Q} \dim \mathscr{V}_i$$

The designation "direct sum" is sometimes used in a slightly different context. If $\mathscr{V}$ and $\mathscr{U}$ are vector spaces, not necessarily subspaces of a common vector space, the set $\mathscr{V} \times \mathscr{U}$ can be given the vector space structure by defining addition and scalar multiplication as in Exercise 8.5. The set $\mathscr{V} \times \mathscr{U}$ with this vector space structure is also called the direct sum and is written $\mathscr{V} \oplus \mathscr{U}$. This concept of direct sum is slightly more general since $\mathscr{V}$ and $\mathscr{U}$ need not initially be subspaces of a third vector space. However, after we have defined $\mathscr{U} \oplus \mathscr{V}$, then $\mathscr{U}$ and $\mathscr{V}$ can be viewed as subspaces of $\mathscr{U} \oplus \mathscr{V}$; further, in that sense, $\mathscr{U} \oplus \mathscr{V}$ is the direct sum of $\mathscr{U}$ and $\mathscr{V}$ in accordance with the original definition of the direct sum of subspaces.

Exercises

10.1 Under what conditions can a single vector constitute a subspace?

10.2 Prove Theorems 10.4 and 10.5.

10.3 If $\mathscr{U}$ and $\mathscr{W}$ are subspaces of $\mathscr{V}$, show that any subspace which contains $\mathscr{U} \cup \mathscr{W}$ also contains $\mathscr{U} + \mathscr{W}$.

10.4 If $\mathcal{V}$ and $\mathcal{U}$ are vector spaces and if $\mathcal{V} \oplus \mathcal{U}$ is their direct sum in the sense of the second definition given in this section, reprove Theorem 10.9.

10-5 Let $\mathcal{V} = \mathcal{R}^3$, and suppose that $\mathcal{U}$ is the subspace spanned by $\{(0,1,1)\}$ and $\mathcal{W}$ is the subspace spanned by $\{(1,0,1), (1,1,1)\}$. Show that $\mathcal{V} = \mathcal{U} \oplus \mathcal{W}$. Find another subspace $\bar{\mathcal{U}}$ such that $\mathcal{V} = \bar{\mathcal{U}} \oplus \mathcal{W}$.

Section 11. Factor Space

In Section 2 the concept of an equivalence relation on a set was introduced and in Exercise 2.2 an equivalence relation was used to partition a set into equivalence sets. In this section an equivalence relation is introduced and the vector space is partitioned into equivalence sets. The class of all equivalence sets is itself a vector space called the *factor space*.

If $\mathscr{U}$ is a subset of a vector space $\mathscr{V}$, then two vectors $\mathbf{w}$ and $\mathbf{v}$ in $\mathscr{V}$ are said to be *equivalent* with respect to $\mathscr{U}$, written $\mathbf{v} \sim \mathbf{w}$, if $\mathbf{w} - \mathbf{v}$ is a vector contained in $\mathscr{U}$. It is easy to see that this relation is an equivalence relation and it induces a partition of $\mathscr{V}$ into equivalence sets of vectors. If $\mathbf{v} \in \mathscr{V}$, then the *equivalence* set of $\mathbf{v}$, denoted by $\overline{\mathbf{v}}$,[1] is the set of all vectors of the form $\mathbf{v} + \mathbf{u}$, where $\mathbf{u}$ is any vector of $\mathscr{U}$,

$$\overline{\mathbf{v}} = \left\{ \mathbf{v} + \mathbf{u} \middle| \mathbf{u} \in \mathscr{U} \right\}$$

To illustrate the equivalence relation and its decomposition of a vector space into equivalence sets, we will consider the real vector space $\mathscr{R}^2$, which we can represent by the Euclidean plane. Let $\mathbf{u}$ be a fixed vector in $\mathscr{R}^2$, and define the subspace $\mathscr{U}$ of $\mathscr{R}^2$ by $\left\{ \lambda u \middle| \lambda \in \mathscr{R} \right\}$. This subspace consists of all vectors of the form $\lambda \mathbf{u}$, $\lambda \in \mathscr{R}$, which are all parallel to the same straight line. This subspace is illustrated in Figure 2. From the definition of equivalence, the vector $\mathbf{v}$ is seen to be equivalent to the vector $\mathbf{w}$ if the vector $\mathbf{v} - \mathbf{w}$ is parallel to the line representing $\mathscr{U}$. Therefore all vectors that differ from the vector $\mathbf{v}$ by a vector that is parallel to the line representing $\mathscr{U}$ are equivalent. The set of all vectors equivalent to the vector $\mathbf{v}$ is therefore the set of all vectors that terminate on the dashed line parallel to the line representing $\mathscr{U}$ in Figure 2. The equivalence set of $\mathbf{v}, \overline{\mathbf{v}}$, is the set of all such vectors. The following theorem is a special case of Exercise 2.2 and shows that an equivalence relation decomposes $\mathscr{V}$ into disjoint sets, that is to say, each vector is contained in one and only one equivalence set.

[1] Only in this section do we use a bar over a letter to denote an equivalence set. In other sections, unless otherwise specified, a bar denotes the complex conjugate.

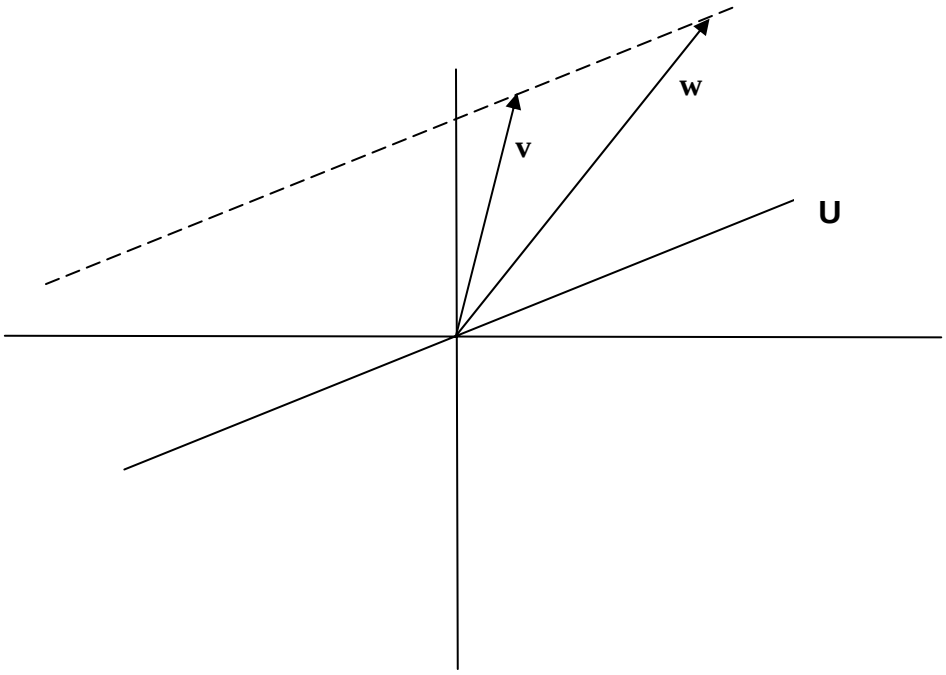

Figure 2

Theorem 11.l. If $\bar{\mathbf{v}} \neq \bar{\mathbf{w}}$, then $\bar{\mathbf{v}} \cap \bar{\mathbf{w}} = \varnothing$.

Proof. Assume that $\bar{\mathbf{v}} \neq \bar{\mathbf{w}}$ but that there exists a vector $\mathbf{x}$ in $\bar{\mathbf{v}} \cap \bar{\mathbf{w}}$. Then $\mathbf{x} \in \bar{\mathbf{v}}$, $\mathbf{x} \sim \mathbf{v}$, and $\mathbf{x} \in \bar{\mathbf{w}}$, $\mathbf{x} \sim \mathbf{w}$. By the transitive property of the equivalence relation we have $\mathbf{v} \sim \mathbf{w}$, which implies $\bar{\mathbf{v}} = \bar{\mathbf{w}}$ and which is a contradiction. Therefore $\bar{\mathbf{v}} \cap \bar{\mathbf{w}}$ contains no vector unless $\bar{\mathbf{v}} = \bar{\mathbf{w}}$, in which case $\bar{\mathbf{v}} \cap \bar{\mathbf{w}} = \bar{\mathbf{v}}$.

We shall now develop the structure of the factor space. The *factor class* of $\mathscr{U}$, denoted by $\mathscr{V}/\mathscr{U}$ is the class of all equivalence sets in $\mathscr{V}$ formed by using a subspace $\mathscr{U}$ of $\mathscr{V}$. The factor class is sometimes called a *quotient class*. Addition and scalar multiplication of equivalence sets are denoted by

$$\bar{\mathbf{v}} + \bar{\mathbf{w}} = \overline{\mathbf{v} + \mathbf{w}}$$

and

$$\lambda\bar{\mathbf{v}} = \overline{\lambda\mathbf{v}}$$

respectively. It is easy to verify that the addition and multiplication operations defined above depend only on the equivalence sets and not on any particular vectors used in representing the sets. The following theorem is easy to prove.

Theorem 11.2. The factor set $\mathscr{V}/\mathscr{U}$ forms a vector space, called a *factor space*, with respect to the operations of addition and scalar multiplication of equivalence classes defined above.

The factor space is also called a *quotient space*. The subspace $\mathscr{U}$ in $\mathscr{V}/\mathscr{U}$ plays the role of the zero vector of the factor space. In the trivial case when $\mathscr{U} = \mathscr{V}$, there is only one equivalence set and it plays the role of the zero vector. On the other extreme when $\mathscr{U} = \{\mathbf{0}\}$,then each equivalence set is a single vector and $\mathscr{V}/\{\mathbf{0}\} = \mathscr{V}$.

Exercises

11.1 Show that the relation between two vectors $\mathbf{v}$ and $\mathbf{w} \in \mathscr{V}$ that makes them equivalent is, in fact, an equivalence relation in the sense of Section 2.

11.2 Give a geometrical interpretation of the process of addition and scalar multiplication of equivalence sets in $\mathscr{R}^2$ in Figure 2.

11.3 Show that $\mathscr{V}$ is equal to the union of all the equivalence sets in $\mathscr{V}$. Thus $\mathscr{V}/\mathscr{U}$ is a class of nonempty, disjoint sets whose union is the entire space $\mathscr{V}$.

11.4 Prove Theorem 11.2.

11.5 Show that $\dim(\mathscr{V}/\mathscr{U}) = \dim\mathscr{V} - \dim\mathscr{U}$.

Section 12. Inner Product Spaces

There is no concept of length or magnitude in the definition of a vector space we have been employing. The reason for the delay in the introduction of this concept is that it is not needed in many of the results of interest. To emphasize this lack of dependence on the concept of magnitude, we have delayed its introduction to this point. Our intended applications of the theory, however, do employ it extensively.

We define the concept of length through the concept of an *inner product*. An inner product on a complex vector space $\mathscr{V}$ is a function $f : \mathscr{V} \times \mathscr{V} \to \mathscr{C}$ with the following properties:

(1) $f(\mathbf{u},\mathbf{v}) = \overline{f(\mathbf{v},\mathbf{u})}$;

(2) $\lambda f(\mathbf{u},\mathbf{v}) = f(\lambda\mathbf{u},\mathbf{v})$;

(3) $f(\mathbf{u}+\mathbf{w},\mathbf{v}) = f(\mathbf{u},\mathbf{v}) + f(\mathbf{w},\mathbf{v})$;

(4) $f(\mathbf{u},\mathbf{u}) \geq 0$ and $f(\mathbf{u},\mathbf{u}) = 0 \Leftrightarrow \mathbf{u} = \mathbf{0}$;

for all $\mathbf{u},\mathbf{v},\mathbf{w} \in \mathscr{V}$ and $\lambda \in \mathscr{C}$. In Property 1 the bar denotes the complex conjugate. Properties 2 and 3 require that f be *linear* in its first argument; i.e., $f(\lambda\mathbf{u}+\mu\mathbf{v},\mathbf{w}) = \lambda f(\mathbf{u},\mathbf{w}) + \mu f(\mathbf{v},\mathbf{w})$ for all $\mathbf{u},\mathbf{v},\mathbf{w} \in \mathscr{V}$ and all $\lambda,\mu \in \mathscr{C}$. Property 1 and the linearity implied by Properties 2 and 3 insure that f is conjugate linear in its second argument; i.e., $f(\mathbf{u},\lambda\mathbf{v}+\mu\mathbf{w}) = \bar{\lambda} f(\mathbf{u},\mathbf{v}) + \bar{\mu} f(\mathbf{u},\mathbf{w})$ for all $\mathbf{u},\mathbf{v},\mathbf{w} \in \mathscr{V}$ and all $\lambda,\mu \in \mathscr{C}$. Since Property 1 ensures that $f(\mathbf{u},\mathbf{u})$ is real, Property 4 is meaningful, and it requires that f be positive definite. There are many notations for the inner product. We shall employ the notation of the "dot product" and write

$$f(\mathbf{u},\mathbf{v}) = \mathbf{u}\cdot\mathbf{v}$$

An *inner product space* is simply a vector space with an inner product. To emphasize the importance of this idea and to focus simultaneously all its details, we restate the definition as follows.

Definition. A *complex inner product space*, or simply an *inner product space*, is a set $\mathscr{V}$ and a field $\mathscr{C}$ such that:

(a) There exists a binary operation in $\mathscr{V}$ called addition and denoted by $+$ such that:

(1) $(\mathbf{u}+\mathbf{v})+\mathbf{w} = \mathbf{u}+(\mathbf{v}+\mathbf{w})$ for all $\mathbf{u},\mathbf{v},\mathbf{w} \in \mathscr{V}$.

(2) $\mathbf{u}+\mathbf{v} = \mathbf{v}+\mathbf{u}$ for all $\mathbf{u},\mathbf{v} \in \mathscr{V}$.

(3) There exists an element $\mathbf{0} \in \mathscr{V}$ such that $\mathbf{u} + \mathbf{0} = \mathbf{u}$ for all $\mathbf{u} \in \mathscr{V}$.

(4) For every $\mathbf{u} \in \mathscr{V}$ there exists an element $-\mathbf{u} \in \mathscr{V}$ such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$.

(b) There exists an operation called *scalar multiplication* in which every scalar $\lambda \in \mathscr{C}$ can be combined with every element $\mathbf{u} \in \mathscr{V}$ to give an element $\lambda\mathbf{u} \in \mathscr{V}$ such that:

(l) $\lambda(\mu\mathbf{u}) = (\lambda\mu)\mathbf{u}$;

(2) $(\lambda + \mu)\mathbf{u} = \lambda\mathbf{u} + \mu\mathbf{u}$;

(3) $\lambda(\mathbf{u} + \mathbf{v}) = \lambda\mathbf{u} + \lambda\mathbf{v}$;

(4) $1\mathbf{u} = \mathbf{u}$;
for all $\lambda, \mathbf{u}, \in \mathscr{C}$ and all $\mathbf{u}, \mathbf{v} \in \mathscr{V}$;

(c) There exists an operation called *inner product* by which any ordered pair of vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathscr{V}$ determines an element of $\mathscr{C}$ denoted by $\mathbf{u} \cdot \mathbf{v}$ such that

(1) $\mathbf{u} \cdot \mathbf{v} = \overline{\mathbf{v} \cdot \mathbf{u}}$;

(2) $\lambda\mathbf{u} \cdot \mathbf{v} = (\lambda\mathbf{u}) \cdot \mathbf{v}$;

(3) $(\mathbf{u} + \mathbf{w}) \cdot \mathbf{v} = \mathbf{u} \cdot \mathbf{v} + \mathbf{w} \cdot \mathbf{v}$;

(4) $\mathbf{u} \cdot \mathbf{u} \geq \mathbf{0}$ and $\mathbf{u} \cdot \mathbf{u} = \mathbf{0} \Leftrightarrow \mathbf{u} = 0$;
for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathscr{V}$ and $\lambda \in \mathscr{C}$.

A *real* inner product space is defined similarly. The vector space $\mathscr{C}^N$ becomes an inner product space if, for any two vectors $\mathbf{u}, \mathbf{v} \in \mathscr{C}^N$. where $\mathbf{u} = (\lambda_1, \lambda_2, ..., \lambda_N)$ and $\mathbf{v} = (\mu_1, \mu_2, ..., \lambda_N)$, we define the inner product of $\mathbf{u}$ and $\mathbf{v}$ by

$$\mathbf{u} \cdot \mathbf{v} = \sum_{j=1}^{N} \lambda_j \bar{\mu}_j$$

The *length* of a vector is an operation, denoted by $\| \ \|$, that assigns to each nonzero vector $\mathbf{v} \in \mathscr{V}$ a positive real number by the following rule:

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} \tag{12.1}$$

Of course the length of the zero vector is zero. The definition represented by (12.1) is for an inner product space of N dimensions in general and therefore generalizes the concept of "length" or "magnitude" from elementary Euclidean plane geometry to *N*-dimensional spaces. Before continuing this process of algebraically generalizing geometric notions, it is necessary to pause and prove two inequalities that will aid in the generalization process.

Theorem 12.1. The *Schwarz inequality*,

$$|\mathbf{u}\cdot\mathbf{v}| \le \|\mathbf{u}\|\|\mathbf{v}\| \tag{12.2}$$

is valid for any two vectors $\mathbf{u},\mathbf{v}$ in an inner product space.

Proof. The Schwarz inequality is easily seen to be trivially true when either $\mathbf{u}$ or $\mathbf{v}$ is the $\mathbf{0}$ vector, so we shall assume that neither $\mathbf{u}$ nor $\mathbf{v}$ is zero. Construct the vector $(\mathbf{u}\cdot\mathbf{u})\mathbf{v}-(\mathbf{v}\cdot\mathbf{u})\mathbf{u}$ and employ Property (c4), which requires that every vector have a nonnegative length, hence

$$\left(\|\mathbf{u}\|^2\|\mathbf{v}\|^2-(\mathbf{u}\cdot\mathbf{v})\overline{(\mathbf{u}\cdot\mathbf{v})}\right)\|\mathbf{u}\|^2 \ge 0$$

Since $\mathbf{u}$ must not be zero, it follows that

$$\|\mathbf{u}\|^2\|\mathbf{v}\|^2 \ge (\mathbf{u}\cdot\mathbf{v})\left(\overline{\mathbf{u}\cdot\mathbf{v}}\right) = |\mathbf{u}\cdot\mathbf{v}|^2$$

and the positive square root of this equation is Schwarz's inequality.

Theorem 12.2. The triangle inequality

$$\|\mathbf{u}+\mathbf{v}\| \le \|\mathbf{u}\|+\|\mathbf{v}\| \tag{12.3}$$

is valid for any two vectors $\mathbf{u},\mathbf{v}$ in an inner product space.

Proof: The squared length of $\mathbf{u}+\mathbf{v}$ can be written in the form

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

where Re signifies the real part. By use of the Schwarz inequality this can be rewritten as

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

Taking the positive square root of this equation, we obtain the triangular inequality.

For a *real* inner product space the concept of angle is defined as follows. The *angle* between two vectors $\mathbf{u}$ and $\mathbf{v}$, denoted by θ, is defined by

$$\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} \tag{12.4}$$

This definition of a real-valued angle is meaningful because the Schwarz inequality in this case shows that the quantity on the right-hand side of (12.4) must have a value lying between 1 and -1, i.e.,

$$-1 \le \frac{\mathbf{u}\cdot\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} \le +1$$

Returning to complex inner product spaces in general, we say that two vectors $\mathbf{u}$ and $\mathbf{v}$ are *orthogonal* if $\mathbf{u}\cdot\mathbf{v}$ or $\mathbf{v}\cdot\mathbf{u}$ is zero. Clearly, this definition is consistent with the real case, since orthogonality then means $|\theta| = \pi/2$.

The inner product space is a very substantial algebraic structure and parts of the structure can be given slightly different interpretations. In particular it can be shown that the length $\|\mathbf{v}\|$ of a vector $\mathbf{v} \in \mathscr{V}$ is a particular example of a mathematical concept known as a norm, and thus an inner product space is a *normal space*. A *norm* on $\mathscr{V}$ is a real-valued function defined on $\mathscr{V}$ whose value is denoted by $\|\mathbf{v}\|$ and which satisfies the following axioms:

(1) $\|\mathbf{v}\| \ge 0$ and $\|\mathbf{v}\| = 0 \Leftrightarrow \mathbf{v} = \mathbf{0}$;

(2) $\|\lambda\mathbf{v}\| = |\lambda|\|\mathbf{v}\|$;

(3) $\|\mathbf{u}\| + \|\mathbf{v}\| \ge \|\mathbf{u}+\mathbf{v}\|$;

for all $\mathbf{u}, \mathbf{v} \in \mathscr{V}$ and all $\lambda \in \mathscr{C}$. In defining the norm of $\mathbf{v}$, we have employed the same notation as that for the length of $\mathbf{v}$ because we will show that the length defined by an inner product is a norm, but the converse need not be true.

Theorem 12.3. The operation of determining the length $\|\mathbf{v}\|$ of $\mathbf{v} \in \mathscr{V}$ is a norm on $\mathscr{V}$.

Proof. The proof follows easily from the definition of length. Properties (c1), (c2), and (c4) of an inner product imply Axioms l and 2 of a norm, and the triangle inequality is proved by Theorem 12.2.

We will now show that the inner product space is also a *metric space*. A *metric space* is a nonempty set $\mathscr{M}$ equipped with a positive real-valued function $\mathscr{M} \times \mathscr{M} \to \mathscr{R}$,called the *distance* function that satisfies the following axioms:

(1) $d(\mathbf{u},\mathbf{v}) \geq 0$ and $d(\mathbf{u},\mathbf{v}) = 0 \Leftrightarrow \mathbf{u} = \mathbf{v}$;

(2) $d(\mathbf{u},\mathbf{v}) = d(\mathbf{v},\mathbf{u})$;

(3) $d(\mathbf{u},\mathbf{w}) \leq d(\mathbf{u},\mathbf{v}) + d(\mathbf{v},\mathbf{w})$;

for all $\mathbf{u},\mathbf{v},\mathbf{w} \in \mathscr{M}$. In other words, a metric space is a set $\mathscr{M}$ of objects and a positive-definite, symmetric distance function d satisfying the triangle inequality. The boldface notation for the elements of $\mathscr{M}$ is simply to save the introduction of yet another notation.

Theorem 12.4. An inner product space $\mathscr{V}$ is a metric space with the distance function given by

$$d(\mathbf{u},\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\| \tag{12.5}$$

Proof. Let $\mathbf{w} = \mathbf{u} - \mathbf{v}$; then from the requirement that $\|\mathbf{w}\| \geq 0$ and $\|\mathbf{w}\| = 0 \Leftrightarrow \mathbf{w} = \mathbf{0}$ it follows that

$$\|\mathbf{u}-\mathbf{v}\| \geq 0 \text{ and } \|\mathbf{u}-\mathbf{v}\| = 0 \Leftrightarrow \mathbf{u} = \mathbf{v}$$

Similarly, from the requirement that $\|\mathbf{w}\| = \|-\mathbf{w}\|$, it follows that

$$\|\mathbf{u}-\mathbf{v}\| = \|\mathbf{v}-\mathbf{u}\|$$

Finally, let $\mathbf{u}$ be replaced by $\mathbf{u}-\mathbf{v}$ and $\mathbf{v}$ by $\mathbf{v}-\mathbf{w}$ in the triangle inequality (12.3); then

$$\|\mathbf{u}-\mathbf{w}\| \leq \|\mathbf{u}-\mathbf{v}\| + \|\mathbf{v}-\mathbf{w}\|$$

which is the third and last requirement for a distance function in a metric space.

The inner product as well as the expressions (12.1) for the length of a vector can be expressed in terms of any basis and the components of the vectors relative to that basis. Let $\{\mathbf{e}_1,\mathbf{e}_2,...,\mathbf{e}_N\}$ be a basis of $\mathscr{V}$ and denote the inner product of any two base vectors by e_{jk},

$$e_{jk} \equiv \mathbf{e}_j \cdot \mathbf{e}_k = \overline{e}_{kj} \tag{12.6}$$

Thus, if the vectors $\mathbf{u}$ and $\mathbf{v}$ have the representations

$$\mathbf{u} = \sum_{j=1}^{N} \lambda^j \mathbf{e}_j, \qquad \mathbf{v} = \sum_{k=1}^{N} \mu^k \mathbf{e}_k \tag{12.7}$$

relative to the basis $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_N\}$ then the inner product of $\mathbf{u}$ and $\mathbf{v}$ is given by

$$\mathbf{u} \cdot \mathbf{v} = \sum_{j=1}^{N} \lambda^j \mathbf{e}_j \cdot \sum_{k=1}^{N} \mu^k \mathbf{e}_k = \sum_{j=1}^{N} \sum_{k=1}^{N} \lambda^j \overline{\mu}^k e_{jk} \tag{12.8}$$

Equation (12.8) is the component expression for the inner product.

From the definition (12.1) for the length of a vector $\mathbf{v}$, and from (12.6) and $(12.7)_2$, we can write

$$\|\mathbf{v}\| = \left(\sum_{j=1}^{N} \sum_{k=1}^{N} \mathbf{e}_{jk} \mu^j \overline{\mu}^k \right)^{1/2} \tag{12.9}$$

This equation gives the component expression for the length of a vector. For a real inner product space, it easily follows that

$$\cos\theta = \frac{\displaystyle\sum_{j=1}^{N} \sum_{k=1}^{N} e_{jk} \mu^j \lambda^k}{\left(\displaystyle\sum_{p=1}^{N} \sum_{r=1}^{N} e_{pr} \mu^p \mu^r \right)^{1/2} \left(\displaystyle\sum_{l=1}^{N} \sum_{s=1}^{N} e_{sl} \lambda^s \lambda^l \right)^{1/2}} \tag{12.10}$$

In formulas (12.8)-(12.10) notice that we have changed the particular indices that indicate summation so that no more than two indices occur in any summand. The reason for changing these dummy indices of summation can be made apparent by failing to change them. For example, in order to write (12.9) in its present form, we can write the component expressions for $\mathbf{v}$ in two equivalent ways,

$$\mathbf{v} = \sum_{j=1}^{N} \mu^j \mathbf{e}_j, \qquad \mathbf{v} = \sum_{k=1}^{N} \mu^k \mathbf{e}_k$$

and then take the dot product. If we had used the same indices to represent summation in both cases, then that index would occur four times in (12.9) and all the cross terms in the inner product would be left out.

Exercises

12.1 Derive the formula

$$2\mathbf{u}\cdot\mathbf{v} = \|\mathbf{u}+\mathbf{v}\|^2 + i\|\mathbf{u}+i\mathbf{v}\|^2 - (1+i)\|\mathbf{u}\|^2 - (1+i)\|\mathbf{v}\|^2$$

which expresses the inner product of two vectors in terms of the norm. This formula is known as the *polar identity*.

12.2 Show that the norm $\|\ \|$ induced by an inner product according to the definition (12.1) must satisfy the following *parallelogram law*;

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

for all vectors $\mathbf{u}$ and $\mathbf{v}$. Prove by counterexample that a norm in general need not be an induced norm of any inner product.

12.3 Use the definition of angle given in this section and the properties of the inner product to derive the law of cosines.

12.4 If $\mathscr{V}$ and $\mathscr{U}$ are inner produce spaces, show that the equation

$$f\big((\mathbf{v},\mathbf{w}),(\mathbf{u},\mathbf{b})\big) = \mathbf{v}\cdot\mathbf{u} + \mathbf{w}\cdot\mathbf{b}$$

where $\mathbf{v},\mathbf{u}\in\mathscr{V}$ and $\mathbf{w},\mathbf{b}\in\mathscr{U}$, defines an inner product on $\mathscr{V}\oplus\mathscr{U}$.

12.5 Show that the only vector in $\mathscr{V}$ that is orthogonal to every other vector in $\mathscr{V}$ is the zero vector.

12.6 Show that the Schwarz and triangle inequalities become equalities if and only if the vectors concerned are linearly dependent.

12.7 Show that $\|\mathbf{u}_1 - \mathbf{u}_2\| \geq \|\mathbf{u}_1\| - \|\mathbf{u}_2\|$ for all $\mathbf{u}_1, \mathbf{u}_2$ in an inner product space $\mathscr{V}$.

12.8 Prove by direct calculation that $\sum_{j=1}^{N}\sum_{k=1}^{N} e_{jk}\mu^j\overline{\mu}^k$ in (12.9) is real.

12.9 If $\mathscr{U}$ is a subspace of an inner product space $\mathscr{V}$, prove that $\mathscr{U}$ is also an inner product space.

12.10 Prove that the $N\times N$ matrix $\left[e_{jk}\right]$ defined by (12.6) is nonsingular.

Section 13. Orthonormal Bases and Orthogonal Complements

Experience with analytic geometry tells us that it is generally much easier to use bases consisting of vectors which are orthogonal and of unit length rather than arbitrarily selected vectors. Vectors with a magnitude of 1 are called *unit vectors* or *normalized vectors*. A set of vectors in an inner product space $\mathscr{V}$ is said to be an *orthogonal set* if all the vectors in the set are mutually orthogonal, and it is said to be an *orthonormal set* if the set is orthogonal and if all the vectors are unit vectors. In equation form, an orthonormal set $\{\mathbf{i}_1,...,\mathbf{i}_M\}$ satisfies the conditions

$$\mathbf{i}_j \cdot \mathbf{i}_k = \delta_{jk} \equiv \begin{cases} 1 & \text{if} \quad j = k \\ 0 & \text{if} \quad j \neq k \end{cases} \tag{13.1}$$

The symbol δ_{jk} introduced in the equation above is called the *Kronecker* delta.

Theorem 13.1. An orthonormal set is linearly independent.

Proof. Assume that the orthonormal set $\{\mathbf{i}_1,\mathbf{i}_2,...,\mathbf{i}_M\}$ is linearly dependent, that is to say there exists a set of scalars $\{\lambda^1,\lambda^2,\ldots,\lambda^M\}$, not all zero, such that

$$\sum_{j=1}^{M} \lambda^j \mathbf{i}_j = \mathbf{0}$$

The inner product of this sum with the unit vector $\mathbf{i}_k$ gives the expression

$$\sum_{j=1}^{M} \lambda^j \delta_{jk} = \lambda^1 \delta_{1k} + \cdots + \lambda^M \delta_{Mk} = \lambda^k = 0$$

for $k = 1,2,...,M$. A contradiction has therefore been achieved and the theorem is proved.

As a corollary to this theorem, it is easy to see that an orthonormal set in an inner product space $\mathscr{V}$ can have no more than $N = \dim \mathscr{V}$ elements. An orthonormal set is said to be *complete* in an inner product space $\mathscr{V}$ if it is not a proper subset of another orthonormal set in the same space. Therefore every orthonormal set with $N = \dim \mathscr{V}$ elements is complete and hence maximal in the sense of a linearly independent set. It is possible to prove the converse: Every complete orthonormal set in an inner product space $\mathscr{V}$ has $N = \dim \mathscr{V}$ elements. The same result is put in a slightly different fashion in the following theorem.

Theorem 13.2. A complete orthonormal set is a basis for $\mathscr{V}$; such a basis is called an *orthonormal basis*.

Any set of linearly independent vectors can be used to construct an orthonormal set, and likewise, any basis can be used to construct an orthonormal basis. The process by which this is done is called the *Gram-Schmidt orthogonalization process* and it is developed in the proof of the following theorem.

Theorem 13.3. Given a basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N\}$ of an inner product space $\mathscr{V}$, then there exists an orthonormal basis $\{\mathbf{i}_1, \ldots, \mathbf{i}_N\}$ such that $\{\mathbf{e}_1 \ldots, \mathbf{e}_k\}$ and $\{\mathbf{i}_1, \ldots, \mathbf{i}_k\}$ generate the same subspace $\mathscr{U}_k$ of $\mathscr{V}$, for each $k = 1, \ldots, N$.

Proof. The construction proceeds in two steps; first a set of orthogonal vectors is constructed, then this set is normalized. Let $\{\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_N\}$ denote a set of orthogonal, but not unit, vectors. This set is constructed from $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_N\}$ as follows: Let $\mathbf{d}_1 = \mathbf{e}_1$ and put

$$\mathbf{d}_2 = \mathbf{e}_2 + \xi \mathbf{d}_1$$

The scalar ξ will be selected so that $\mathbf{d}_2$ is orthogonal to $\mathbf{d}_1$; orthogonality of $\mathbf{d}_1$ and d_2 requires that their inner product be zero; hence

$$d_2 \cdot d_1 = 0 = e_2 \cdot d_1 + \xi d_1 \cdot d_1$$

implies

$$\xi = -\frac{e_2 \cdot d_1}{d_1 \cdot d_1}$$

where $d_1 \cdot d_1 \neq 0$ since $d_1 \neq 0$. The vector d_2 is not zero, because e_1 and e_2 are linearly independent. The vector d_3 is defined by

$$d_3 = e_3 + \xi^2 d_2 + \xi^1 d_1$$

The scalars ξ^2 and ξ^1 are determined by the requirement that d_3 be orthogonal to both d_1 and d_2; thus

$$d_3 \cdot d_1 = e_3 \cdot d_1 + \xi^1 d_1 \cdot d_1 = 0$$

$$d_3 \cdot d_2 = e_3 \cdot d_2 + \xi^2 d_2 \cdot d_2 = 0$$

and, as a result,

$$\xi^1 = -\frac{e_1 \cdot d_1}{d_1 \cdot d_1}, \quad \xi^2 = -\frac{e_3 \cdot d_2}{d_2 \cdot d_2}$$

The linear independence of $e_1, e_2,$ and e_3 requires that d_3 be nonzero. It is easy to see that this scheme can be repeated until a set of *N* orthogonal vectors $\{\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_N\}$ has been obtained. The orthonormal set is then obtained by defining

$$\mathbf{i}_k = \mathbf{d}_k / \|\mathbf{d}_k\|, \quad k = 1, 2, \ldots, N$$

It is easy to see that $\{\mathbf{e}_1 \ldots, \mathbf{e}_k\}$, $\{\mathbf{d}_1, \ldots, \mathbf{d}_k\}$, and, hence, $\{\mathbf{i}_1, \ldots, \mathbf{i}_k\}$ generate the same subspace for each *k*.

The concept of mutually orthogonal vectors can be generalized to mutually orthogonal subspaces. In particular, if $\mathscr{U}$ is a subspace of an inner product space, then the *orthogonal complement* of $\mathscr{U}$ is a subset of $\mathscr{V}$, denoted by $\mathscr{U}^\perp$, such that

$$\mathscr{U}^\perp = \{\mathbf{v} \,|\, \mathbf{v} \cdot \mathbf{u} = 0 \text{ for all } \mathbf{u} \in \mathscr{U}\} \tag{13.2}$$

The properties of orthogonal complements are developed in the following theorem.

Theorem 13.4. If $\mathscr{U}$ is a subspace of $\mathscr{V}$, then (a) $\mathscr{U}^\perp$ is a subspace of $\mathscr{V}$ and (b) $\mathscr{V} = \mathscr{U} \oplus \mathscr{U}^\perp$.

Proof. (a) If $\mathbf{u}_1$ and $\mathbf{u}_2$ are in $\mathscr{U}^\perp$, then

$$\mathbf{u}_1 \cdot \mathbf{v} = \mathbf{u}_2 \cdot \mathbf{v} = 0$$

for all $\mathbf{v} \in \mathscr{U}$. Therefore for any $\lambda_1, \lambda_2 \in \mathscr{C}$,

$$\left(\lambda_1\mathbf{u}_1 + \lambda_2\mathbf{u}_2\right)\cdot\mathbf{v} = \lambda_1\mathbf{u}_1\cdot\mathbf{v} + \lambda_2\mathbf{u}_2\cdot\mathbf{v} = 0$$

Thus $\lambda_1\mathbf{u}_1 + \lambda_2\mathbf{u}_2 \in \mathscr{U}^\perp$.

(b) Consider the vector space $\mathscr{U} + \mathscr{U}^\perp$. Let $\mathbf{v} \in \mathscr{U} \cap \mathscr{U}^\perp$. Then by (13.2), $\mathbf{v}\cdot\mathbf{v} = 0$, which implies $\mathbf{v} = \mathbf{0}$. Thus $\mathscr{U} + \mathscr{U}^\perp = \mathscr{U} \oplus \mathscr{U}^\perp$. To establish that $\mathscr{V} = \mathscr{U} \oplus \mathscr{U}^\perp$, let $\{\mathbf{i}_1,\ldots,\mathbf{i}_R\}$ be an orthonormal basis for $\mathscr{U}$. Then consider the decomposition

$$\mathbf{v} = \left\{\mathbf{v} - \sum_{q=1}^{R}\left(\mathbf{v}\cdot\mathbf{i}_q\right)\mathbf{i}_q\right\} + \sum_{q=1}^{R}\left(\mathbf{v}\cdot\mathbf{i}_q\right)\mathbf{i}_q$$

The term in the brackets is orthogonal to each $\mathbf{i}_q$ and it thus belongs to $\mathscr{U}^\perp$. Also, the second term is in $\mathscr{U}$ Therefore $\mathscr{V} = \mathscr{U} + \mathscr{U}^\perp = \mathscr{U} \oplus \mathscr{U}^\perp$, and, the proof is complete.

As a result of the last theorem, any vector $\mathbf{v} \in \mathscr{V}$ can be written uniquely in the form

$$\mathbf{v} = \mathbf{u} + \mathbf{w} \tag{13.3}$$

where $\mathbf{u} \in \mathscr{U}$ and $\mathbf{w} \in \mathscr{U}^\perp$. If $\mathbf{v}$ is a vector with the decomposition indicated in (13.3), then

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

It follows from this equation that

$$\|\mathbf{v}\|^2 \geq \|\mathbf{u}\|^2, \qquad \|\mathbf{v}\|^2 \geq \|\mathbf{w}\|^2$$

The equation is the well-known *Pythagorean theorem*, and the inequalities are special cases of an inequality known as *Bessel's inequality*.

Exercises

13.1 Show that with respect to an orthonormal basis the formulas (12.8)-(12.10) can be written as

$$\mathbf{u}\cdot\mathbf{v} = \sum_{j=1}^{N}\lambda^j\overline{\mu}^j \tag{13.4}$$

$$\|\mathbf{v}\|^2 = \sum_{j=1}^{N} \mu^j \bar{\mu}^j \tag{13.5}$$

and, for a real vector space,

$$\cos\theta = \frac{\sum_{j=1}^{N} \lambda^j \mu^j}{\left(\sum_{p=1}^{N} \mu^p \mu^p\right)^{1/2} \left(\sum_{l=1}^{N} \mu^l \mu^l\right)^{1/2}} \tag{13.6}$$

13.2 Prove Theorem 13.2.

13.3 If $\mathscr{U}$ is a subspace of the inner product space $\mathscr{V}$, show that

$$\left(\mathscr{U}^{\perp}\right)^{\perp} = \mathscr{U}$$

13.4 If $\mathscr{V}$ is an inner product space, show that

$$\mathscr{V}^{\perp} = \{\mathbf{0}\}$$

Compare this with Exercise 12.5.

13.5 Given the basis

$$\mathbf{e}_1 = (1,1,1), \quad \mathbf{e}_2 = (0,1,1), \quad \mathbf{e}_3 = (0,0,1)$$

for $\mathscr{R}^3$, construct an orthonormal basis according to the Gram-Schmidt orthogonalization process.

13.6 If $\mathscr{U}$ is a subspace of an inner product space $\mathscr{V}$, show that

$$\dim \mathscr{U}^{\perp} = \dim \mathscr{V} - \dim \mathscr{U}$$

13.7 Given two subspaces $\mathscr{V}_1$ and $\mathscr{V}_2$ of $\mathscr{V}$, show that

$$\left(\mathscr{V}_1 + \mathscr{V}_2\right)^{\perp} = \mathscr{V}_1^{\perp} \cap \mathscr{V}_2^{\perp}$$

and

$$\left(\mathscr{V}_1 \cap \mathscr{V}_2\right)^{\perp} = \mathscr{V}_1^{\perp} + \mathscr{V}_2^{\perp}$$

13.8 In the proof of Theorem 13.4 we used the fact that if a vector is orthogonal to each vector $\mathbf{i}_q$ of a basis $\{\mathbf{i}_1,\ldots,\mathbf{i}_R\}$ of $\mathscr{U}$, then that vector belongs to $\mathscr{U}^\perp$. Give a proof of this fact.

Section 14. Reciprocal Basis and Change of Basis

In this section we define a special basis, called the *reciprocal basis*, associated with each basis of the inner product space $\mathscr{V}$. Components of vectors relative to both the basis and the reciprocal basis are defined and formulas for the change of basis are developed.

The set of N vectors $\{\mathbf{e}^1, \mathbf{e}^2, ..., \mathbf{e}^N\}$ is said to be the *reciprocal basis* relative to the basis $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_N\}$ of an inner product space $\mathscr{V}$ if

$$\mathbf{e}^k \cdot \mathbf{e}_s = \delta_s^k, \qquad k, s = 1, 2, \ldots, N \tag{14.1}$$

where the symbol δ_s^k is the *Kronecker delta* defined by

$$\delta_s^k = \begin{cases} 1, & k = s \\ 0, & k \neq s \end{cases} \tag{14.2}$$

Thus each vector of the reciprocal basis is orthogonal to $N-1$ vectors of the basis and when its inner product is taken with the *Nth* vector of the basis, the inner product has the value one. The following two theorems show that the reciprocal basis just defined exists uniquely and is actually a basis for $\mathscr{V}$.

Theorem 14.1. The reciprocal basis relative to a given basis exists and is unique.

Proof. Existence. We prove only existence of the vector $\mathbf{e}^1$; existence of the remaining vectors can be proved similarly. Let $\mathscr{U}$ be the subspace generated by the vectors $\mathbf{e}_2, ..., \mathbf{e}_N$, and suppose that $\mathscr{U}^\perp$ is the orthogonal complement of $\mathscr{U}$. Then $\dim \mathscr{U} = N-1$, and from Theorem 10.9, $\dim \mathscr{U}^\perp = 1$. Hence we can choose a nonzero vector $\mathbf{w} \in \mathscr{U}^\perp$. Since $\mathbf{e}_1 \notin \mathscr{U}$, $\mathbf{e}_1$ and w are not orthogonal,

$$\mathbf{e}_1 \cdot \mathbf{w} \neq 0$$

we can simply define

$$\mathbf{e}^1 \equiv \frac{1}{\mathbf{e}_1 \cdot \mathbf{w}} \mathbf{w}$$

Then $\mathbf{e}^1$ obeys (14.1) for $k = 1$.

Uniqueness. Assume that there are two reciprocal bases, $\{\mathbf{e}^1, \mathbf{e}^2, ..., \mathbf{e}^N\}$ and $\{\mathbf{d}^1, \mathbf{d}^2, ..., \mathbf{d}^N\}$ relative to the basis $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_N\}$ of $\mathscr{V}$. Then from (14.1)

$$\mathbf{e}^k \cdot \mathbf{e}_s = \delta_s^k \text{ and } \mathbf{d}^k \cdot \mathbf{e}_s = \delta_s^k \text{ for } k, s = 1, 2, \ldots, N$$

Subtracting these two equations, we obtain

$$\left(\mathbf{e}^k - \mathbf{d}^k\right) \cdot \mathbf{e}_s = 0, \qquad k, s = 1, 2, \ldots, N \tag{14.3}$$

Thus the vector $\mathbf{e}^k - \mathbf{d}^k$ must be orthogonal to $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_N\}$ Since the basis generates $\mathscr{V}$, (14.3) is equivalent to

$$\left(\mathbf{e}^k - \mathbf{d}^k\right) \cdot \mathbf{v} = 0 \qquad \text{for all} \quad \mathbf{v} \in \mathscr{V} \tag{14.4}$$

In particular, we can choose $\mathbf{v}$ in (14.4) to be equal to $\mathbf{e}^k - \mathbf{d}^k$, and it follows then from the definition of an inner product space that

$$\mathbf{e}^k = \mathbf{d}^k, \qquad k = 1, 2, \ldots, N \tag{14.5}$$

Therefore the reciprocal basis relative to $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_N\}$ is unique.

The logic in passing from (14.4) to (14.5) has appeared before; cf. Exercise 13.8.

Theorem 14.2. The reciprocal basis $\{\mathbf{e}^1, \mathbf{e}^2, ..., \mathbf{e}^N\}$ with respect to the basis $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_N\}$ of the inner product space $\mathscr{V}$ is itself a basis for $\mathscr{V}$.

Proof. Consider the linear relation

$$\sum_{q=1}^{N} \lambda_q \mathbf{e}^q = \mathbf{0}$$

If we compute the inner product of this equation with $\mathbf{e}_k$, $k = 1, 2, ..., N$, then

$$\sum_{q=1}^{N} \lambda_q \mathbf{e}^q \cdot \mathbf{e}_k = \sum_{q=1}^{N} \lambda_q \delta_k^q = \lambda_k = 0$$

Thus the reciprocal basis is linearly independent. But since the reciprocal basis contains the same number of vectors as that of a basis, it is itself a basis for $\mathscr{V}$.

Since $\mathbf{e}^k$, $k = 1, 2, ..., N$, is in $\mathscr{V}$, we can always write

$$\mathbf{e}^k = \sum_{q=1}^{N} e^{kq} \mathbf{e}_q \tag{14.6}$$

where, by (14.1) and (12.6),

$$\mathbf{e}^k \cdot \mathbf{e}_s = \delta_s^k = \sum_{q=1}^{N} e^{kq} e_{qs} \tag{14.7}$$

and

$$\mathbf{e}^k \cdot \mathbf{e}^j = \sum_{q=1}^{N} e^{kq} \delta_q^j = e^{kj} = \overline{e^{jk}} \tag{14.8}$$

From a matrix viewpoint, (14.7) shows that the N^2 quantities e^{kq} $(k, q = 1, 2, \ldots, N)$ are the elements of the inverse of the matrix whose elements are e_{qs}. In particular, the matrices $\left[e^{kq} \right]$ and $\left[e_{qs} \right]$ are nonsingular. This remark is a proof of Theorem 14.2 also by using Theorem 9.5. It is possible to establish from (14.6) and (14.7) that

$$\mathbf{e}_s = \sum_{k=1}^{N} e_{sk} \mathbf{e}^k \tag{14.9}$$

To illustrate the construction of a reciprocal basis by an algebraic method, consider the real vector space $\mathscr{R}^2$, which we shall represent by the Euclidean plane. Let a basis be given for $\mathscr{R}^2$ which consists of two vectors 45° degrees apart, the first one, $\mathbf{e}_1$, two units long and the second one, $\mathbf{e}_2$, one unit long. These two vectors are illustrated in Figure 3.

To construct the reciprocal basis, we first note that from the given information and equation (12.6) we can write

$$e_{11} = 4, \quad e_{12} = e_{21} = \sqrt{2}, \quad e_{22} = 1 \tag{14.10}$$

Writing equations $(14.7)_2$ out explicitly for the case $N = 2$, we have

$$\begin{aligned} &e^{11}e_{11} + e^{12}e_{21} = 1, \qquad e^{21}e_{11} + e^{22}e_{21} = 0 \\ &e^{11}e_{12} + e^{12}e_{22} = 0, \qquad e^{21}e_{12} + e^{22}e_{22} = 1 \end{aligned} \tag{14.11}$$

Substituting (14.10) into (14.11), we find that

$$e^{11} = \frac{1}{2}, \qquad e^{12} = e^{21} = -1/\sqrt{2}, \qquad e^{22} = 2 \tag{14.12}$$

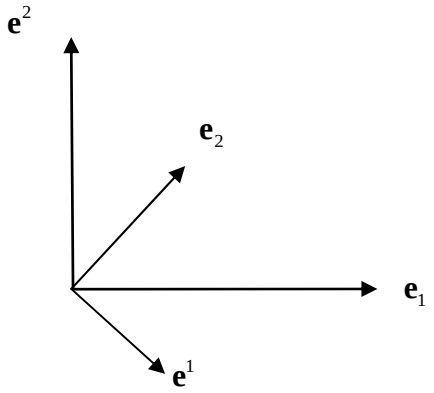

Figure 3. A basis and reciprocal basis for $\mathbf{R}^2$

When the results (14.2) are put into the special case of (14.6) for $N = 2$, we obtain the explicit expressions for the reciprocal basis $\{\mathbf{e}^1, \mathbf{e}^2\}$,

$$\mathbf{e}^1 = \frac{1}{2}\mathbf{e}_1 - \frac{1}{\sqrt{2}}\mathbf{e}_2, \quad \mathbf{e}^2 = -\frac{1}{\sqrt{2}}\mathbf{e}_1 + 2\mathbf{e}_2 \tag{14.13}$$

The reciprocal basis is illustrated in Figure 3 also.

Henceforth we shall use the same kernel letter to denote the components of a vector relative to a basis as we use for the vector itself. Thus, a vector $\mathbf{v}$ has components v^k, $k = 1, 2, ..., N$ relative to the basis $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_N\}$ and components v_k, $k = 1, 2, ..., N$, relative to its reciprocal basis,

$$\mathbf{v} = \sum_{k=1}^{N} v^k \mathbf{e}_k, \qquad \mathbf{v} = \sum_{k=1}^{N} v_k \mathbf{e}^k \tag{14.14}$$

The components $v^1, v^2, ..., v^N$ with respect to the basis $\{\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_N\}$ are often called the *contravariant components* of $\mathbf{v}$, while the components $v_1, v_2, ..., v_N$ with respect to the reciprocal basis $\{\mathbf{e}^1, \mathbf{e}^2, ..., \mathbf{e}^N\}$ are called *covariant components*. The names covariant and contravariant are somewhat arbitrary since the basis and reciprocal basis are both bases and we have no particular procedure to choose one over the other. The following theorem illustrates further the same remark.

Theorem 14. 3. If $\{\mathbf{e}^1, ..., \mathbf{e}^N\}$ is the reciprocal basis of $\{\mathbf{e}_1, ..., \mathbf{e}_N\}$ then $\{\mathbf{e}_1, ..., \mathbf{e}_N\}$ is also the reciprocal basis of $\{\mathbf{e}^1, ..., \mathbf{e}^N\}$.

For this reason we simply say that the bases $\{\mathbf{e}_1, ..., \mathbf{e}_N\}$ and $\{\mathbf{e}^1, ..., \mathbf{e}^N\}$ are (mutually) reciprocal. The contravariant and covariant components of v are related to one another .by the formulas

$$\begin{aligned} v^k &= \mathbf{v} \cdot \mathbf{e}^k = \sum_{q=1}^{N} e^{qk} v_q \\ v_k &= \mathbf{v} \cdot \mathbf{e}_k = \sum_{q=1}^{N} e_{qk} v^q \end{aligned} \tag{14.15}$$

where equations (12.6) and (14.8) have been employed. More generally, if $\mathbf{u}$ has contravariant components u^i and covariant components u_i relative to $\{\mathbf{e}_1, ..., \mathbf{e}_N\}$ and $\{\mathbf{e}^1, ..., \mathbf{e}^N\}$, respectively, then the inner product of u and v can be computed by the formulas

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{N} u_i \overline{v}^i = \sum_{i=1}^{N} u^i \overline{v}_i = \sum_{i=1}^{N} \sum_{j=1}^{N} e^{ij} u_i \overline{v}_j = \sum_{i=1}^{N} \sum_{j=1}^{N} e_{ij} u^i \overline{v}^j \tag{14.16}$$

which generalize the formulas (13.4) and (12.8).

As an example of the covariant and contravariant components of a vector, consider a vector $\mathbf{v}$ in $\mathscr{R}^2$ which has the representation

$$\mathbf{v} = \frac{3}{2}\mathbf{e}_1 + 2\mathbf{e}_2$$

relative to the basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ illustrated in Figure 3. The contravariant components of $\mathbf{v}$ are then $(3/2, 2)$; to compute the covariant components, the formula $(14.16)_2$ is written out for the case $N = 2$,

$$v_1 = e_{11}v^1 + e_{21}v^2, \quad v_2 = e_{12}v^1 + e_{22}v^2$$

Then from (14.10) and the values of the contravariant components the covariant components are given by

$$v_1 = 6 + 2\sqrt{2}, \qquad v_2 = \left(3/\sqrt{2}\right) + 2$$

hence

$$\mathbf{v} = \left(6 + 2\sqrt{2}\right)\mathbf{e}^1 + \left(3/\sqrt{2} + 2\right)\mathbf{e}^2$$

The contravariant components of the vector $\mathbf{v}$ are illustrated in Figure 4. To develop a geometric feeling for covariant and contravariant components, it is helpful for one to perform a vector decomposition of this type for oneself.

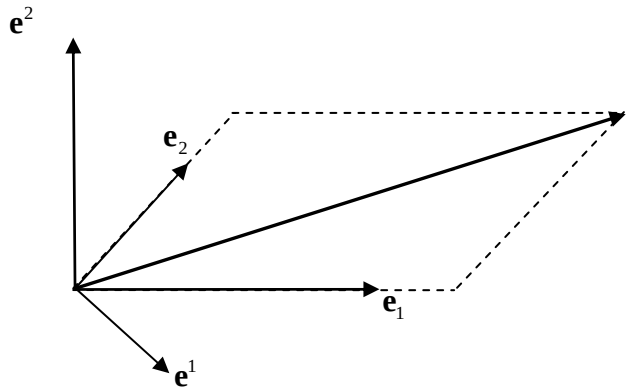

Figure 4. The covariant and contravariant components of a vector in $\mathbf{R}^2$.

A substantial part of the usefulness of orthonormal bases is that each orthonormal basis is self-reciprocal; hence contravariant and covariant components of vectors coincide. The self-

reciprocity of orthonormal bases follows by comparing the condition (13.1) for an orthonormal basis with the condition (14.1) for a reciprocal basis. In orthonormal systems indices are written only as subscripts because there is no need to distinguish between covariant and contravariant components.

Formulas for transferring from one basis to another basis in an inner product space can be developed for both base vectors and components. In these formulas one basis is the set of linearly independent vectors $\{\mathbf{e}_1,...,\mathbf{e}_N\}$, while the second basis is the set $\{\hat{\mathbf{e}}_1,...,\hat{\mathbf{e}}_N\}$. From the fact that both bases generate $\mathscr{V}$, we can write

$$\hat{\mathbf{e}}_k = \sum_{s=1}^{N} T_k^s \mathbf{e}_s, \quad k = 1, 2, ..., N \tag{14.17}$$

and

$$\mathbf{e}_q = \sum_{k=1}^{N} \hat{T}_q^k \hat{\mathbf{e}}_k, \quad q = 1, 2, ..., N \tag{14.18}$$

where $\hat{T}_q^k$ and T_k^s are both sets of N^2 scalars that are related to one another. From Theorem 9.5 the $N \times N$ matrices $\left[T_k^s\right]$ and $\left[\hat{T}_q^k\right]$ must be nonsingular.

Substituting (14.17) into (14.18) and replacing $\mathbf{e}_q$ by $\sum_{s=1}^{N} \delta_q^s \mathbf{e}_s$, we obtain

$$\mathbf{e}_q = \sum_{k=1}^{N} \sum_{s=1}^{N} \hat{T}_q^k T_k^s \mathbf{e}_s = \sum_{s=1}^{N} \delta_q^s \mathbf{e}_s$$

which can be rewritten as

$$\sum_{s=1}^{N} \left(\sum_{k=1}^{N} \hat{T}_q^k T_k^s - \delta_q^s \right) \mathbf{e}_s = \mathbf{0} \tag{14.19}$$

The linear independence of the basis $\{\mathbf{e}_s\}$ requires that

$$\sum_{k=1}^{N} \hat{T}_q^k T_k^s = \delta_q^s \tag{14.20}$$

A similar argument yields

$$\sum_{k=1}^{N} \hat{T}_k^q T_s^k = \delta_s^q \tag{14.21}$$

In matrix language we see that (14.20) or (14.21) requires that the matrix of elements T_q^k be the inverse of the matrix of elements $\hat{T}_k^q$. It is easy to verify that the reciprocal bases are related by

$$\mathbf{e}^k = \sum_{q=1}^{N} \overline{T_q^k}\, \hat{\mathbf{e}}^q, \qquad \hat{\mathbf{e}}^q = \sum_{k=1}^{N} \overline{\hat{T}_k^q}\, \mathbf{e}^k \tag{14.22}$$

The covariant and contravariant components of $v \in V$ relative to the two pairs of reciprocal bases are given by

$$\mathbf{v} = \sum_{k=1}^{N} v_k \mathbf{e}^k = \sum_{k=1}^{N} v^k \mathbf{e}_k = \sum_{q=1}^{N} \hat{v}^q \hat{\mathbf{e}}_q = \sum_{q=1}^{N} \hat{v}_q \hat{\mathbf{e}}^q \tag{14.23}$$

To obtain a relationship between, say, covariant components of $\mathbf{v}$, one can substitute $(14.22)_1$ into (14.23), thus

$$\sum_{k=1}^{N} \sum_{q=1}^{N} v_k \overline{T_q^k}\, \hat{\mathbf{e}}^q = \sum_{q=1}^{N} \hat{v}_q \hat{\mathbf{e}}^q$$

This equation can be rewritten in the form

$$\sum_{q=1}^{N} \left(\sum_{k=1}^{N} \overline{T_q^k} v_k - \hat{v}_q \right) \hat{\mathbf{e}}^q = \mathbf{0}$$

and it follows from the linear independence of the basis $\{\hat{\mathbf{e}}^q\}$ that

$$\hat{v}_q = \sum_{k=1}^{N} \overline{T_q^k} v_k \tag{14.24}$$

In a similar manner the following formulas can be obtained:

$$\hat{v}^q = \sum_{k=1}^{N} \hat{T}_k^q v^k \tag{14.25}$$

$$v^k = \sum_{q=1}^{N} T_q^k \hat{v}^q \tag{14.26}$$

and

$$v_k = \sum_{q=1}^{N} \overline{\hat{T}_k^q} \hat{v}_q \tag{14.27}$$

Exercises

14.1 Show that the quantities T_q^s and $\hat{T}_q^k$ introduced in formulas (14.17) and (14.18) are given by the expressions

$$T_q^s = \hat{\mathbf{e}}_k \cdot \mathbf{e}^s, \qquad \hat{T}_q^k = \mathbf{e}_q \cdot \hat{\mathbf{e}}^k \tag{14.28}$$

14.2 Derive equation (14.21).

14.3 Derive equations (14.25)-(14.27).

14.4 Given the change of basis in a three-dimensional inner product space $\mathscr{V}$,

$$\mathbf{f}_1 = 2\mathbf{e}_1 - \mathbf{e}_2 - \mathbf{e}_3, \quad \mathbf{f}_2 = -\mathbf{e}_1 + \mathbf{e}_3, \quad \mathbf{f}_3 = 4\mathbf{e}_1 - \mathbf{e}_2 + 6\mathbf{e}_3$$

find the quantities T_k^s and $\hat{T}_q^k$.

14.5 Given the vector $\mathbf{v} = \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3$ in the vector space of the problem above, find the covariant and contravariant components of $\mathbf{v}$ relative to the basis $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$.

14.6 Show that under the basis transformation (14.17)

$$\hat{e}_{kj} = \hat{\mathbf{e}}_k \cdot \hat{\mathbf{e}}_j = \sum_{s,q=1}^{N} T_k^s \overline{T_j^q} e_{sq}$$

and

$$\hat{e}^{kj} = \hat{\mathbf{e}}^k \cdot \hat{\mathbf{e}}^j = \sum_{s,q=1}^{N} \overline{\hat{T}_s^k} \hat{T}_q^j e^{sq}$$

14.7 Prove Theorem 14.3.

Chapter 4

LINEAR TRANSFORMATIONS

Section 15 Definition of Linear Transformation

In this section and in the other sections of this chapter we shall introduce and study a special class of functions defined on a vector space. We shall assume that this space has an inner product, although this structure is not essential for the arguments in Sections 15-17. If $\mathscr{V}$ and $\mathscr{U}$ are vector spaces, a *linear transformation* is a function $\mathbf{A}:\mathscr{V}\to\mathscr{U}$ such that

(a) $\mathbf{A}(\mathbf{u}+\mathbf{v})=\mathbf{A}(\mathbf{u})+\mathbf{A}(\mathbf{v})$

(b) $\mathbf{A}(\lambda\mathbf{u})=\lambda\mathbf{A}(\mathbf{u})$

for all $\mathbf{u},\mathbf{v}\in\mathscr{V}$ and $\lambda\in\mathscr{C}$. Condition (a) asserts that $\mathbf{A}$ is a *homomorphism* on $\mathscr{V}$ with respect to the operation of addition and thus the theorems of Section 6 can be applied here. Condition (b) shows that $\mathbf{A}$, in addition to being a homomorphism, is also *homogeneous* with respect to the operation of scalar multiplication. Observe that the $+$ symbol on the left side of (a) denotes addition in $\mathscr{V}$, while on the right side it denotes addition in $\mathscr{U}$. It would be extremely cumbersome to adopt different symbols for these quantities. Further, it is customary to omit the parentheses and write simply $\mathbf{Au}$ for $\mathbf{A}(\mathbf{u})$ when $\mathbf{A}$ is a linear transformation.

Theorem 15.1. If $\mathbf{A}:\mathscr{V}\to\mathscr{U}$ is a function from a vector space $\mathscr{V}$ to a vector space $\mathscr{U}$, then A is a linear transformation if and only if

$$\mathbf{A}(\lambda\mathbf{u}+\mu\mathbf{v})=\lambda\mathbf{A}(\mathbf{u})+\mu\mathbf{A}(\mathbf{v})$$

for all $\mathbf{u},\mathbf{v}\in\mathscr{V}$ and $\lambda,\mu\in\mathscr{C}$.

The proof of this theorem is an elementary application of the definition. It is also possible to show that

$$\mathbf{A}\left(\lambda^1\mathbf{v}_1+\lambda^2\mathbf{v}_2+\cdots+\lambda^R\mathbf{v}_R\right)=\lambda^1\mathbf{A}\mathbf{v}_1+\lambda^2\mathbf{A}\mathbf{v}_2+\cdots+\lambda^R\mathbf{A}\mathbf{v}_R \tag{15.1}$$

for all $\mathbf{v}_1,...,\mathbf{v}_R \in \mathscr{V}$ and $\lambda^1,...,\lambda^R \in \mathscr{C}$.

By application of Theorem 6.1, we see that for a linear transformation $\mathbf{A}:\mathscr{V}\to\mathscr{U}$

$$\mathbf{A0}=\mathbf{0} \qquad \text{and} \qquad \mathbf{A}(-\mathbf{v})=-\mathbf{Av} \tag{15.2}$$

Note that in $(15.2)_1$ we have used the same symbol for the zero vector in $\mathscr{V}$ as in $\mathscr{U}$.

Theorem 15.2. If $\{\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_R\}$ is a linearly dependent set in $\mathscr{V}$ and if $\mathbf{A}:\mathscr{V}\to\mathscr{U}$ is a linear transformation, then $\{\mathbf{Av}_1,\mathbf{Av}_2,\ldots,\mathbf{Av}_R\}$ is a linearly dependent set in $\mathscr{U}$.

Proof Since the vectors $\mathbf{v}_1,...,\mathbf{v}_R$ are linearly dependent, we can write

$$\sum_{j=1}^{R}\lambda^j\mathbf{v}_j=\mathbf{0}$$

where at least one coefficient is not zero. Therefore

$$\mathbf{A}\left(\sum_{j=1}^{R}\lambda^j\mathbf{v}_j\right)=\sum_{j=1}^{R}\lambda^j\mathbf{Av}_j=\mathbf{0}$$

where (15.1) and (15.2) have been used. The last equation proves the theorem.

If the vectors $\mathbf{v}_1,\ldots,\mathbf{v}_R$ are linearly independent, then their image set $\{\mathbf{Av}_1,\mathbf{Av}_2,\ldots,\mathbf{Av}_R\}$ may or may not be linearly independent. For example, $\mathbf{A}$ might map all vectors into $\mathbf{0}$. The *kernel* of a linear transformation $\mathbf{A}:\mathscr{V}\to\mathscr{U}$ is the set

$$K(\mathbf{A})=\{\mathbf{v}\,|\,\mathbf{Av}=\mathbf{0}\}$$

In other words, $K(\mathbf{A})$ is the preimage of the set $\{\mathbf{0}\}$ in $\mathscr{U}$. Since $\{\mathbf{0}\}$ is a subgroup of the additive group $\mathscr{U}$, it follows from Theorem 6.3 that $K(\mathbf{A})$ is a subgroup of the additive group $\mathscr{V}$. However, a stronger statement can be made.

Theorem 15.3. The set $K(\mathbf{A})$ is a subspace of $\mathscr{V}$.

Proof. Since $K(\mathbf{A})$ is a subgroup, we only need to prove that if $\mathbf{v} \in K(\mathbf{A})$, then $\lambda\mathbf{v} \in K(\mathbf{A})$ for all $\lambda \in \mathscr{C}$. This is clear since

$$\mathbf{A}(\lambda\mathbf{v}) = \lambda\mathbf{A}\mathbf{v} = \mathbf{0}$$

for all $\mathbf{v}$ in $K(\mathbf{A})$.

The kernel of a linear transformation is sometimes called the *null space*. The *nullity* of a linear transformation is the dimension of the kernel, i.e., $\dim K(\mathbf{A})$. Since $K(\mathbf{A})$ is a subspace of $\mathscr{V}$, we have, by Theorem 10.2,

$$\dim K(\mathbf{A}) \leq \dim \mathscr{V} \tag{15.3}$$

Theorem 15.4. A linear transformation $\mathbf{A}: \mathscr{V} \to \mathscr{U}$ is one-to-one if and only if $K(\mathbf{A}) = \{\mathbf{0}\}$.

Proof. This theorem is just a special case of Theorem 6.4. For ease of reference, we shall repeat the proof. If $\mathbf{Au} = \mathbf{Av}$, then by the linearity of $\mathbf{A}$, $\mathbf{A}(\mathbf{u} - \mathbf{v}) = \mathbf{0}$. Thus if $K(\mathbf{A}) = \{\mathbf{0}\}$, then $\mathbf{Au} = \mathbf{Av}$ implies $\mathbf{u} = \mathbf{v}$, so that $\mathbf{A}$ is one-to-one. Now assume $\mathbf{A}$ is one-to-one. Since $K(\mathbf{A})$ is a subspace, it must contain the zero in $\mathscr{V}$ and therefore $\mathbf{A0} = \mathbf{0}$. If $K(\mathbf{A})$ contained any other element $\mathbf{v}$, we would have $\mathbf{Av} = \mathbf{0}$, which contradicts the fact that $\mathbf{A}$ is one-to-one.

Linear transformations that are one-to-one are called *regular linear transformations.* The following theorem gives another condition for such linear transformations.

Theorem 15.5. A linear transformation $\mathbf{A}: \mathscr{V} \to \mathscr{U}$ is regular if and only if it maps linearly independent sets in $\mathscr{V}$ to linearly independent sets in $\mathscr{U}$.

Proof. Necessity. Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_R\}$ be a linearly independent set in $\mathscr{V}$ and $\mathbf{A}: \mathscr{V} \to \mathscr{U}$ be a regular linear transformation. Consider the sum

$$\sum_{j=1}^{R} \lambda^j \mathbf{A}\mathbf{v}_j = \mathbf{0}$$

This equation is equivalent to

$$\mathbf{A}\left(\sum_{j=1}^{R} \lambda^j \mathbf{v}_j \right) = \mathbf{0}$$

Since $\mathbf{A}$ is regular, we must have

$$\sum_{j=1}^{R} \lambda^j \mathbf{v}_j = \mathbf{0}$$

Since the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_R$ are linearly independent, this equation shows $\lambda^1 = \lambda^2 = \cdots = \lambda^R = 0$, which implies that the set $\{\mathbf{A}\mathbf{v}_1, \ldots, \mathbf{A}\mathbf{v}_R\}$ is linearly independent.

Sufficiency. The assumption that $\mathbf{A}$ preserves linear independence implies, in particular, that $\mathbf{A}\mathbf{v} \neq \mathbf{0}$ for every nonzero vector $\mathbf{v} \in \mathscr{V}$ since such a vector forms a linearly independent set. Therefore $K(\mathbf{A})$ consists of the zero vector only, and thus $\mathbf{A}$ is regular.

For a linear transformation $\mathbf{A}: \mathscr{V} \to \mathscr{U}$ we denote the range of $\mathbf{A}$ by

$$R(\mathbf{A}) = \{\mathbf{A}\mathbf{v} | \mathbf{v} \in \mathscr{V}\}$$

It follows form Theorem 6.2 that $R(\mathbf{A})$ is a *subgroup* of $\mathscr{U}$. We leave it to the reader to prove the stronger results stated below.

Theorem 15.6. The range $R(\mathbf{A})$ is a subspace of $\mathscr{U}$.

We have from Theorems 15.6 and 10.2 that

$$\dim R(\mathbf{A}) \le \dim \mathscr{U} \tag{15.4}$$

The *rank* of a linear transformation is defined as the dimension of $R(\mathbf{A})$, i.e., $\dim R(\mathbf{A})$. A stronger statement than (15.4) can be made regarding the rank of linear transformation $\mathbf{A}$.

Theorem 15.7. $\dim R(\mathbf{A}) \le \min(\dim \mathscr{V}, \dim \mathscr{U})$.

Proof. Clearly, it suffices to prove that $\dim R(\mathbf{A}) \le \dim \mathscr{V}$, since this inequality and (15.4) imply the assertion of the theorem. Let $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N\}$ be a basis for $\mathscr{V}$, where $N = \dim \mathscr{V}$. Then $\mathbf{v} \in \mathscr{V}$ can be written

$$\mathbf{v} = \sum_{j=1}^{N} \mathbf{v}^j \mathbf{e}_j$$

Therefore any vector $\mathbf{A}\mathbf{v} \in R(\mathbf{A})$ can be written

$$\mathbf{A}\mathbf{v} = \sum_{j=1}^{N} v^j \mathbf{A}\mathbf{e}_j$$

Hence the vectors $\{\mathbf{A}\mathbf{e}_1, \mathbf{A}\mathbf{e}_2, \ldots, \mathbf{A}\mathbf{e}_N\}$ generate $R(\mathbf{A})$. By Theorem 9.10, we can conclude

$$\dim R(\mathbf{A}) \le \dim \mathscr{V} \tag{15.5}$$

A result that improves on (15.5) is the following important theorem.

Theorem 15.8. If $\mathbf{A} : \mathscr{V} \to \mathscr{U}$ is a linear transformation, then

$$\dim \mathscr{V} = \dim R(\mathbf{A}) + \dim K(\mathbf{A}) \tag{15.6}$$

Proof. Let $P = \dim K(\mathbf{A})$, $R = \dim R(\mathbf{A})$, and $N = \dim \mathscr{V}$. We must prove that $N = P + R$. Select N vectors in $\mathscr{V}$ such that $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_P\}$ is a basis for $K(\mathbf{A})$ and $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_P, \mathbf{e}_{P+1}, \ldots, \mathbf{e}_N\}$ a basis for $\mathscr{V}$. As in the proof of Theorem 15.7, the vectors $\{\mathbf{Ae}_1, \mathbf{Ae}_2, \ldots, \mathbf{Ae}_P, \mathbf{Ae}_{P+1}, \ldots, \mathbf{Ae}_N\}$ generate $R(\mathbf{A})$. But, by the properties of the kernel, $\mathbf{Ae}_1 = \mathbf{Ae}_2 = \cdots = \mathbf{Ae}_p = \mathbf{0}$. Thus the vectors $\{\mathbf{Ae}_{P+1}, \mathbf{Ae}_{P+2}, \ldots, \mathbf{Ae}_N\}$ generate $R(\mathbf{A})$. If we can establish that these vectors are linearly independent, we can conclude from Theorem 9.10 that the vectors $\{\mathbf{Ae}_{P+1}, \mathbf{Ae}_{P+2}, \ldots, \mathbf{Ae}_N\}$ form a basis for $R(\mathbf{A})$ and that $\dim R(\mathbf{A}) = R = N - P$. Consider the sum

$$\sum_{j=1}^{R} \lambda^j \mathbf{Ae}_{P+j} = \mathbf{0}$$

Therefore

$$\mathbf{A}\left(\sum_{j=1}^{R} \lambda^j \mathbf{e}_{p+j} \right) = \mathbf{0}$$

which implies that the vector $\sum_{j=1}^{R} \lambda^j \mathbf{e}_{P+j} \in K(\mathbf{A})$. This fact requires that $\lambda^1 = \lambda^2 = \cdots = \lambda^R = 0$, or otherwise the vector $\sum_{j=1}^{R} \lambda^j \mathbf{e}_{P+j}$ could be expanded in the basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_P\}$, contradicting the linear independence of $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_N\}$. Thus the set $\{\mathbf{Ae}_{P+1}, \mathbf{Ae}_{P+2}, \ldots, \mathbf{Ae}_N\}$ is linearly independent and the proof of the theorem is complete.

As usual, a linear transformation $\mathbf{A} : \mathscr{V} \to \mathscr{U}$ is said to be *onto* if $R(\mathbf{A}) = \mathscr{U}$, i.e., for every vector $\mathbf{u} \in \mathscr{U}$ there exists $\mathbf{v} \in \mathscr{V}$ such that $\mathbf{Av} = \mathbf{u}$.

Theorem 15.9. $\dim R(\mathbf{A}) = \dim \mathscr{U}$ if and only if $\mathbf{A}$ is onto.

In the special case when $\dim \mathscr{V} = \dim \mathscr{U}$, it is possible to state the following important theorem.

Theorem 15.10. If $\mathbf{A} : \mathscr{V} \to \mathscr{U}$ is a linear transformation and if $\mathscr{V} = \dim \mathscr{U}$, then $\mathbf{A}$ is a linear transformation *onto* $\mathscr{U}$ if and only if $\mathbf{A}$ is regular.

Proof. Assume that $\mathbf{A}:\mathscr{V}\to\mathscr{U}$ is onto $\mathscr{U}$, then (15.6) and Theorem 15.9 show that

$$\dim\mathscr{V}=\dim\mathscr{U}=\dim\mathscr{U}+\dim K(\mathbf{A})$$

Therefore $\dim K(\mathbf{A})=0$ and thus $K(\mathbf{A})=\{\mathbf{0}\}$ and $\mathbf{A}$ is one-to-one. Next assume that $\mathbf{A}$ is one-to-one. By Theorem 15.4, $K(\mathbf{A})=\{\mathbf{0}\}$ and thus $\dim K(\mathbf{A})=0$. Then (15.6) shows that

$$\dim\mathscr{V}=\dim R(\mathbf{A})=\dim\mathscr{U}$$

By using Theorem 10.3, we can conclude that

$$R(\mathbf{A})=\mathscr{U}$$

and thus $\mathbf{A}$ is onto.

Exercises

15.1 Prove Theorems 15.1, 15.6, and 15.9

15.2 Let $\mathbf{A}:\mathscr{V}\to\mathscr{U}$ be a linear transformation, and let $\mathscr{V}^1$ be a subspace of $\mathscr{V}$. The restriction of $\mathbf{A}$ to $\mathscr{V}^1$ is a function $\mathbf{A}\big|_{\mathscr{V}^1}:\mathscr{V}^1\to\mathscr{U}$ defined by

$$\mathbf{A}\big|_{\mathscr{V}^1}\mathbf{v}=\mathbf{A}\mathbf{v}$$

for all $\mathbf{v}\in\mathscr{V}^1$. Show that $\mathbf{A}\big|_{\mathscr{V}^1}$ is a linear transformation and that $K\left(\mathbf{A}\big|_{\mathscr{V}^1}\right)=K(\mathbf{A})\cap\mathscr{V}^1$

15.3 Let $\mathbf{A}:\mathscr{V}\to\mathscr{U}$ be a linear transformation, and define a function $\bar{\mathbf{A}}:\mathscr{V}/K(\mathbf{A})\to\mathscr{U}$ by

$$\bar{\mathbf{A}}\bar{\mathbf{v}}=\mathbf{A}\mathbf{v}$$

for all $\mathbf{v} \in \mathscr{V}$. Here $\bar{\mathbf{v}}$ denotes the equivalence set of $\mathbf{v}$ in $\mathscr{V}/K(\mathbf{A})$. Show that $\bar{\mathbf{A}}$ is a linear transformation. Show also that $\bar{\mathbf{A}}$ is regular and that $R(\mathbf{A}) = R(\bar{\mathbf{A}})$.

15.4 Let $\mathscr{V}^1$ be a subspace of $\mathscr{V}$, and define a function $\mathbf{P}: \mathscr{V} \to \mathscr{V}/\mathscr{V}^1$ by

$$\mathbf{P}\mathbf{v} = \bar{\mathbf{v}}$$

for all $\mathbf{v}$ in $\mathscr{V}$. Prove that $\mathbf{P}$ is a linear transformation, onto, and that $K(\mathbf{P}) = \mathscr{V}^1$. The mapping $\mathbf{P}$ is called the *canonical projection* from $\mathscr{V}$ to $\mathscr{V}/\mathscr{V}^1$.

15.5 Prove the formula

$$\dim \mathscr{V} = \dim\left(\mathscr{V}/\mathscr{V}^1\right) + \dim \mathscr{V}^1 \tag{15.7}$$

of Exercise 11.5 by applying Theorem 15.8 to the canonical projection $\mathbf{P}$ defined in the preceding exercise. Conversely, prove the formula (15.6) of Theorem 15.8 by using the formula (15.7) and the result of Exercise 15.3.

15.6 Let $\mathscr{U}$ and $\mathscr{V}$ be vector spaces, and let $\mathscr{U} \oplus \mathscr{V}$ be their direct sum. Define mapping $\mathbf{P}_1: \mathscr{U} \oplus \mathscr{V} \to \mathscr{U}$ and $\mathbf{P}_2: \mathscr{U} \oplus \mathscr{V} \to \mathscr{V}$ by

$$\mathbf{P}_1(\mathbf{u}, \mathbf{v}) = \mathbf{u}, \quad \mathbf{P}_2(\mathbf{u}, \mathbf{v}) = \mathbf{v}$$

for all $\mathbf{u} \in \mathscr{U}$ and $\mathbf{v} \in \mathscr{V}$. Show that $\mathbf{P}_1$ and $\mathbf{P}_2$ are onto linear transformations. Show also the formula

$$\dim(\mathscr{U} \oplus \mathscr{V}) = \dim \mathscr{U} + \dim \mathscr{V}$$

of Theorem 10.9 by using these linear transformations and the formula (15.6). The mappings $\mathbf{P}_1$ and $\mathbf{P}_2$ are also called the *canonical projections* from $\mathscr{U} \oplus \mathscr{V}$ to $\mathscr{U}$ and $\mathscr{V}$, respectively.

Section 16. Sums and Products of Liner Transformations

In this section we shall assign meaning to the operations of *addition* and *scalar multiplication* for linear transformations. If $\mathbf{A}$ and $\mathbf{B}$ are linear transformations $\mathscr{V} \to \mathscr{U}$, then their *sum* $\mathbf{A}+\mathbf{B}$ is a linear transformation defined by

$$(\mathbf{A}+\mathbf{B})\mathbf{v} = \mathbf{A}\mathbf{v} + \mathbf{B}\mathbf{v} \tag{16.1}$$

for all $\mathbf{v} \in \mathscr{V}$. In a similar fashion, if $\lambda \in \mathscr{C}$, then $\lambda\mathbf{A}$ is a linear transformation $\mathscr{V} \to \mathscr{U}$ defined by

$$(\lambda\mathbf{A})\mathbf{v} = \lambda(\mathbf{A}\mathbf{v}) \tag{16.2}$$

for all $\mathbf{v} \in \mathscr{V}$. If we write $\mathscr{L}(\mathscr{V};\mathscr{U})$ for the set of linear transformations from $\mathscr{V}$ to $\mathscr{U}$, then (16.1) and (16.2) make $\mathscr{L}(\mathscr{V};\mathscr{U})$ a *vector space.* The zero element in $\mathscr{L}(\mathscr{V};\mathscr{U})$ is the linear transformation $\mathbf{0}$ defined by

$$\mathbf{0}\mathbf{v} = \mathbf{0} \tag{16.3}$$

for all $\mathbf{v} \in \mathscr{V}$. The *negative* of $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{U})$ is a linear transformation $-\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{U})$ defined by

$$-\mathbf{A} = -1\mathbf{A} \tag{16.4}$$

It follows from (16.4) that -$\mathbf{A}$ is the additive inverse of $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{U})$. This assertion follows from

$$\mathbf{A}+(-\mathbf{A}) = \mathbf{A}+(-1\mathbf{A}) = 1\mathbf{A}+(-1\mathbf{A}) = (1-1)\mathbf{A} = 0\mathbf{A} = \mathbf{0} \tag{16.5}$$

where (16.1) and (16.2) have been used. Consistent with our previous notation, we shall write $\mathbf{A}-\mathbf{B}$ for the sum $\mathbf{A}+(-\mathbf{B})$ formed from the linear transformations $\mathbf{A}$ and $\mathbf{B}$. The formal proof that $\mathscr{L}(\mathscr{V};\mathscr{U})$ is a vector space is left as an exercise to the reader.

Theorem 16.1. $\dim \mathscr{L}(\mathscr{V};\mathscr{U}) = \dim \mathscr{V} \dim \mathscr{U}$.

Proof. Let $\{\mathbf{e}_1,...,\mathbf{e}_N\}$ be a basis for $\mathscr{V}$ and $\{\mathbf{b}_1,...,\mathbf{b}_M\}$ be a basis for $\mathscr{U}$. Define NM linear transformations $\mathbf{A}_\alpha^k : \mathscr{V} \to \mathscr{U}$ by

$$\begin{aligned} &\mathbf{A}_\alpha^k \mathbf{e}_k = \mathbf{b}_\alpha, \qquad k = 1,\ldots,N; \quad \alpha = 1,\ldots,M \\ &\mathbf{A}_\alpha^k \mathbf{e}_p = \mathbf{0}, \qquad k \neq p \end{aligned} \tag{16.6}$$

If $\mathbf{A}$ is an arbitrary member of $\mathscr{L}(\mathscr{V};\mathscr{U})$, then $\mathbf{A}\mathbf{e}_k \in \mathscr{U}$, and thus

$$\mathbf{A}\mathbf{e}_k = \sum_{\alpha=1}^{M} \mathbf{A}_k^\alpha \mathbf{b}_\alpha, \quad k = 1,...,N$$

Based upon the properties of $\mathbf{A}_\alpha^k$, we can write the above equation as

$$\mathbf{A}\mathbf{e}_k = \sum_{\alpha=1}^{M} A_k^\alpha \mathbf{A}_\alpha^k \mathbf{e}_k = \sum_{\alpha=1}^{M}\sum_{s=1}^{N} A_s^\alpha \mathbf{A}_\alpha^s \mathbf{e}_k$$

Therefore, since the vectors $\mathbf{e}_1,...,\mathbf{e}_N$ generate $\mathscr{V}$, we find

$$\left(\mathbf{A} - \sum_{\alpha=1}^{M}\sum_{s=1}^{N} A_s^\alpha \mathbf{A}_\alpha^s\right)\mathbf{v} = \mathbf{0}$$

for all vectors $\mathbf{v} \in \mathscr{V}$. Thus, from (16.3),

$$\mathbf{A} = \sum_{\alpha=1}^{M}\sum_{s=1}^{N} A_s^\alpha \mathbf{A}_\alpha^s \tag{16.7}$$

This equation means that the MN linear transformations $\mathbf{A}_\alpha^s \ (s=1,...,N;\alpha=1,...,M)$ generate $\mathscr{L}(\mathscr{V};\mathscr{U})$. If we can prove that these linear transformations are linearly independent, then the proof of the theorem is complete. To this end, set

$$\sum_{\alpha=1}^{M}\sum_{s=1}^{N} A_s^\alpha \mathbf{A}_\alpha^s = \mathbf{0}$$

Then, from (16.6),

$$\sum_{\alpha=1}^{M}\sum_{s=1}^{N} A_s^\alpha \mathbf{A}_\alpha^s(\mathbf{e}_p) = \sum_{\alpha=1}^{M} A_p^\alpha \mathbf{b}_\alpha = \mathbf{0}$$

Thus $A_p^\alpha = 0$, $(p=1,...,N;\ \alpha=1,...,M)$ because the vectors $\mathbf{b}_1,...,\mathbf{b}_M$ are linearly independent in $\mathscr{U}$. Hence $\{\mathbf{A}_\alpha^s\}$ is a basis of $\mathscr{L}(\mathscr{V};\mathscr{U})$. As a result, we have

$$\dim \mathscr{L}(\mathscr{V};\mathscr{U}) = MN = \dim \mathscr{U} \dim \mathscr{V} \tag{16.8}$$

If $\mathbf{A}:\mathscr{V}\to\mathscr{U}$ and $\mathbf{B}:\mathscr{U}\to\mathscr{W}$ are linear transformations, their product is a linear transformation $\mathscr{V}\to\mathscr{W}$, written $\mathbf{BA}$, defined by

$$\mathbf{BAv} = \mathbf{B}(\mathbf{Av}) \tag{16.9}$$

for all $\mathbf{v}\in\mathscr{V}$. The properties of the product operation are summarized in the following theorem.

Theorem 16.2.

$$\begin{aligned} \mathbf{C}(\mathbf{BA}) &= (\mathbf{CB})\mathbf{A} \\ (\lambda\mathbf{A}+\mu\mathbf{B})\mathbf{C} &= \lambda\mathbf{AC}+\mu\mathbf{BC} \\ \mathbf{C}(\lambda\mathbf{A}+\mu\mathbf{B}) &= \lambda\mathbf{CA}+\mu\mathbf{CB} \end{aligned} \tag{16.10}$$

for all $\lambda, \mu \in \mathscr{C}$ and where it is understood that $\mathbf{A}, \mathbf{B}, \text{and } \mathbf{C}$ are defined on the proper vector spaces so as to make the indicated products defined.

The proof of Theorem 16.2 is left as an exercise to the reader.

Exercises

16.1 Prove that $\mathscr{L}(\mathscr{V};\mathscr{U})$ is a vector space

16.2 Prove Theorem 16.2.

16.3 Let $\mathscr{V}, \mathscr{U}, \text{and } \mathscr{W}$ be vector spaces. Given any linear mappings $\mathbf{A}: \mathscr{V} \to \mathscr{U}$ and $\mathbf{B}: \mathscr{U} \to \mathscr{W}$, show that

$$\dim R(\mathbf{BA}) \leq \min\left(\dim R(\mathbf{A}), \dim R(\mathbf{B})\right)$$

16.4 Let $\mathbf{A}: \mathscr{V} \to \mathscr{U}$ be a linear transformation and define $\bar{\mathbf{A}}: \mathscr{V}/K(\mathbf{A}) \to \mathscr{U}$ as in Exercises 15.3. If $\mathbf{P}: \mathscr{V} \to \mathscr{V}/K(\mathbf{A})$ is the canonical projection defined in Exercise 15.4, show that

$$\mathbf{A} = \bar{\mathbf{A}}\mathbf{P}$$

This result along with the results of Exercises 15.3 and 15.4 show that every linear transformation can be written as the composition of an onto linear transformation and a regular linear transformation.

Section 17. Special Types of Linear Transformations

In this section we shall examine the properties of several special types of linear transformations. The first of these is one called an *isomorphism.* In section 6 we discussed group isomorphisms. A vector space *isomorphism* is a regular onto linear transformation $\mathbf{A}:\mathscr{V}\to\mathscr{U}$. It immediately follows from Theorem 15.8 that if $\mathbf{A}:\mathscr{V}\to\mathscr{U}$ is an isomorphism, then

$$\dim\mathscr{V}=\dim\mathscr{U} \tag{17.1}$$

An isomorphism $\mathbf{A}:\mathscr{V}\to\mathscr{U}$ establishes a one-to-one correspondence between the elements of $\mathscr{V}$ and $\mathscr{U}$. Thus there exists a unique inverse function $\mathbf{B}:\mathscr{U}\to\mathscr{V}$ with the property that if

$$\mathbf{u}=\mathbf{A}\mathbf{v} \tag{17.2}$$

then

$$\mathbf{v}=\mathbf{B}(\mathbf{u}) \tag{17.3}$$

for all $\mathbf{u}\in\mathscr{U}$ and $\mathbf{v}\in\mathscr{V}$. We shall now show that $\mathbf{B}$ is a linear transformation. Consider the vectors $\mathbf{u}_1$ and $\mathbf{u}_2\in\mathscr{U}$ and the corresponding vectors [as a result of (17.2) and (17.3)] $\mathbf{v}_1$ and $\mathbf{v}_2\in\mathscr{V}$. Then by (17.2), (17.3), and the properties of the linear transformation $\mathbf{A}$,

$$\begin{aligned}\mathbf{B}(\lambda\mathbf{u}_1+\mu\mathbf{u}_2)&=\mathbf{B}(\lambda\mathbf{A}\mathbf{v}_1+\mu\mathbf{A}\mathbf{v}_2)\\&=\mathbf{B}(\mathbf{A}(\lambda\mathbf{v}_1+\mu\mathbf{v}_2))\\&=\lambda\mathbf{v}_1+\mu\mathbf{v}_2\\&=\lambda\mathbf{B}(\mathbf{u}_1)+\mu\mathbf{B}(\mathbf{u}_2)\end{aligned}$$

Thus $\mathbf{B}$ is a linear transformation. This linear transformation shall be written $\mathbf{A}^{-1}$. Clearly the linear transformation $\mathbf{A}^{-1}$ is also an isomorphism whose inverse is $\mathbf{A}$; i.e.,

$$\left(\mathbf{A}^{-1}\right)^{-1}=\mathbf{A} \tag{17.4}$$

Theorem 17.1. If $\mathbf{A}: \mathscr{V} \to \mathscr{U}$ and $\mathbf{B}: \mathscr{U} \to \mathscr{W}$ are isomorphisms, then $\mathbf{BA}: \mathscr{V} \to \mathscr{W}$ is an isomorphism whose inverse is computed by

$$(\mathbf{BA})^{-1} = \mathbf{A}^{-1}\mathbf{B}^{-1} \tag{17.5}$$

Proof. The fact that $\mathbf{BA}$ is an isomorphism follows directly from the corresponding properties of $\mathbf{A}$ and $\mathbf{B}$. The fact that the inverse of $\mathbf{BA}$ is computed by (17.5) follows directly because if

$$\mathbf{u} = \mathbf{Av} \qquad \text{and} \qquad \mathbf{w} = \mathbf{Bu}$$

then

$$\mathbf{v} = \mathbf{A}^{-1}\mathbf{u} \qquad \text{and} \qquad \mathbf{u} = \mathbf{B}^{-1}\mathbf{w}$$

Thus

$$\mathbf{v} = (\mathbf{BA})^{-1}\mathbf{w} = \mathbf{A}^{-1}\mathbf{B}^{-1}\mathbf{w}$$

Therefore $\left((\mathbf{BA})^{-1} - \mathbf{A}^{-1}\mathbf{B}^{-1}\right)\mathbf{w} = \mathbf{0}$ for all $\mathbf{w} \in \mathscr{W}$, which implies (17.5).

The *identity* linear transformation $\mathbf{I}: \mathscr{V} \to \mathscr{V}$ is defined by

$$\mathbf{Iv} = \mathbf{v} \tag{17.6}$$

for all $\mathbf{v}$ in $\mathscr{V}$. Often it is desirable to distinguish the identity linear transformations on different vector spaces. In these cases we shall denote the identity linear transformation by $\mathbf{I}_{\mathscr{V}}$. It follows from (17.1) and (17.3) that if $\mathbf{A}$ is an isomorphism, then

$$\mathbf{AA}^{-1} = \mathbf{I}_{\mathscr{U}} \qquad \text{and} \qquad \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_{\mathscr{V}} \tag{17.7}$$

Conversely, if $\mathbf{A}$ is a linear transformation from $\mathscr{V}$ to $\mathscr{U}$, and if there exists a linear transformation $\mathbf{B}: \mathscr{U} \to \mathscr{V}$ such that $\mathbf{AB} = \mathbf{I}_{\mathscr{U}}$ and $\mathbf{BA} = \mathbf{I}_{\mathscr{V}}$, then $\mathbf{A}$ is an isomorphism and

$\mathbf{B} = \mathbf{A}^{-1}$. The proof of this assertion is left as an exercise to the reader. Isomorphisms are often referred to as *invertible* or *nonsingular* linear transformations.

A vector space $\mathscr{V}$ and a vector space $\mathscr{U}$ are said to be *isomorphic* if there exists at least one isomorphism from $\mathscr{V}$ to $\mathscr{U}$.

Theorem 17.2. Two finite-dimensional vector spaces $\mathscr{V}$ and $\mathscr{U}$ are isomorphic if and only if they have the same dimension.

Proof. Clearly, if $\mathscr{V}$ and $\mathscr{U}$ are isomorphic, by virtue of the properties of isomorphisms, $\dim \mathscr{V} = \dim \mathscr{U}$. If $\mathscr{U}$ and $\mathscr{V}$ have the same dimension, we can construct a regular onto linear transformation $\mathbf{A} : \mathscr{V} \to \mathscr{U}$ as follows. If $\{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$ is a basis for $\mathscr{V}$ and $\{\mathbf{b}_1, \ldots, \mathbf{b}_N\}$ is a basis for $\mathscr{U}$, define $\mathbf{A}$ by

$$\mathbf{A}\mathbf{e}_k = \mathbf{b}_k, \qquad k = 1, ..., N \tag{17.8}$$

Or, equivalently, if

$$\mathbf{v} = \sum_{k=1}^{N} v^k \mathbf{e}^k$$

then define $\mathbf{A}$ by

$$\mathbf{A}\mathbf{v} = \sum_{k=1}^{N} v^k \mathbf{b}_k \tag{17.9}$$

$\mathbf{A}$ is regular because if $\mathbf{A}\mathbf{v} = \mathbf{0}$, then $\mathbf{v} = \mathbf{0}$. Theorem 15.10 tells us $\mathbf{A}$ is onto and thus is an isomorphism.

As a corollary to Theorem 17.2, we see that $\mathscr{V}$ and the vector space $\mathscr{C}^N$, where $N = \dim \mathscr{V}$, are isomorphic.

In Section 16 we introduced the notation $\mathscr{L}(\mathscr{V}; \mathscr{U})$ for the vector space of linear transformations from $\mathscr{V}$ to $\mathscr{U}$. The set $\mathscr{L}(\mathscr{V}; \mathscr{V})$ corresponds to the vector space of linear transformations $\mathscr{V} \to \mathscr{V}$. An element of $\mathscr{L}(\mathscr{V}; \mathscr{V})$ is called an *endomorphism* of $\mathscr{V}$. This nomenclature parallels the previous usage of the word endomorphism introduced in Section 6. If

an endomorphism is regular (and thus onto), it is called an *automorphism.* The identity linear transformation defined by (17.6) is an example of an automorphism. If $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$, then it is easily seen that

$$\mathbf{AI} = \mathbf{IA} = \mathbf{A} \tag{17.10}$$

Also, if $\mathbf{A}$ and are in $\mathscr{L}(\mathscr{V};\mathscr{V})$, it is meaningful to compute the products $\mathbf{AB}$ and $\mathbf{BA}$; however,

$$\mathbf{AB} \neq \mathbf{BA} \tag{17.11}$$

in general. For example, let $\mathscr{V}$ be a two-dimensional vector space with basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ and define $\mathbf{A}$ and $\mathbf{B}$ by the rules

$$\mathbf{Ae}_k = \sum_{j=1}^{2} A^j{}_k \mathbf{e}_j \qquad \text{and} \qquad \mathbf{Be}_k = \sum_{j=1}^{2} B^j{}_k \mathbf{e}_j$$

where $A^j{}_k$ and $B^j{}_k$. $k, j = 1,2$, are prescribed. Then

$$\mathbf{BAe}_k = \sum_{j,l=1}^{2} A^j{}_k B^l{}_j \mathbf{e}_l \quad \text{and} \qquad \mathbf{ABe}_k = \sum_{j,l=1}^{2} B^j{}_k A^l{}_j \mathbf{e}_l$$

An examination of these formulas shows that it is only for special values of $A^j{}_k$ and $B^j{}_k$ that $\mathbf{AB} = \mathbf{BA}$.

The set $\mathscr{L}(\mathscr{V};\mathscr{V})$ has defined on it three operations. They are (a) addition of elements of $\mathscr{L}(\mathscr{V};\mathscr{V})$, (b) multiplication of an element of $\mathscr{L}(\mathscr{V};\mathscr{V})$ by a scalar, and (c) the product of a pair of elements of $\mathscr{L}(\mathscr{V};\mathscr{V})$. The operations (a) and (b) make $\mathscr{L}(\mathscr{V};\mathscr{V})$ into a vector space, while it is easily shown that the operations (a) and (c) make $\mathscr{L}(\mathscr{V};\mathscr{V})$ into a *ring*. The structure of $\mathscr{L}(\mathscr{V};\mathscr{V})$ is an example of an *associative algebra.*

The subset of $\mathscr{L}(\mathscr{V};\mathscr{V})$ that consists of all automorphisms of $\mathscr{V}$ is denoted by $\mathscr{GL}(\mathscr{V})$. It is immediately apparent that $\mathscr{GL}(\mathscr{V})$ is *not* a subspace of $\mathscr{L}(\mathscr{V};\mathscr{V})$, because the sum of two of its elements need not be an automorphism. However, this set is easily shown to be a *group* with respect to the product operation. This group is called the *general linear group.* Its identity element is $\mathbf{I}$ and if $\mathbf{A} \in \mathscr{GL}(\mathscr{V})$, its inverse is $\mathbf{A}^{-1} \in \mathscr{GL}(\mathscr{V})$.

A *projection* is an endomorphism $\mathbf{P} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ which satisfies the condition

$$\mathbf{P}^2 = \mathbf{P} \tag{17.12}$$

The following theorem gives an important property of a projection.

Theorem 17.3. If $\mathbf{P}: \mathscr{V} \to \mathscr{V}$ is a projection, then

$$\mathscr{V} = R(\mathbf{P}) \oplus K(\mathbf{P}) \tag{17.13}$$

Proof. Let $\mathbf{v}$ be an arbitrary vector in $\mathscr{V}$. Let

$$\mathbf{w} = \mathbf{v} - \mathbf{Pv} \tag{17.14}$$

Then, by (17.12), $\mathbf{Pw} = \mathbf{Pv} - \mathbf{P}(\mathbf{Pv}) = \mathbf{Pv} - \mathbf{Pv} = \mathbf{0}$. Thus, $\mathbf{w} \in K(\mathbf{P})$. Since $\mathbf{Pv} \in R(\mathbf{P})$, (17.14) implies that

$$\mathscr{V} = R(\mathbf{P}) + K(\mathbf{P})$$

To show that $R(\mathbf{P}) \cap K(\mathbf{P}) = \{\mathbf{0}\}$, let $\mathbf{u} \in R(\mathbf{P}) \cap K(\mathbf{P})$. Then, since $\mathbf{u} \in R(\mathbf{P})$ for some $\mathbf{v} \in \mathscr{V}$, $\mathbf{u} = \mathbf{Pv}$. But, since $\mathbf{u}$ is also in $K(\mathbf{P})$,

$$\mathbf{0} = \mathbf{Pu} = \mathbf{P}(\mathbf{Pv}) = \mathbf{Pv} = \mathbf{u}$$

which completes the proof.

The name *projection* arises from the geometric interpretation of (17.13). Given any $\mathbf{v} \in \mathscr{V}$, then there are unique vectors $\mathbf{u} \in R(\mathbf{P})$ and $\mathbf{w} \in K(\mathbf{P})$ such that

$$\mathbf{v} = \mathbf{u} + \mathbf{w} \tag{17.15}$$

where

$$\mathbf{Pu} = \mathbf{u} \qquad \text{and} \qquad \mathbf{Pw} = \mathbf{0} \tag{17.16}$$

Geometrically, $\mathbf{P}$ takes $\mathbf{v}$ and projects in onto the subspace $R(\mathbf{P})$ along the subspace $K(\mathbf{P})$. Figure 5 illustrates this point for $\mathscr{V} = \mathscr{R}^2$.

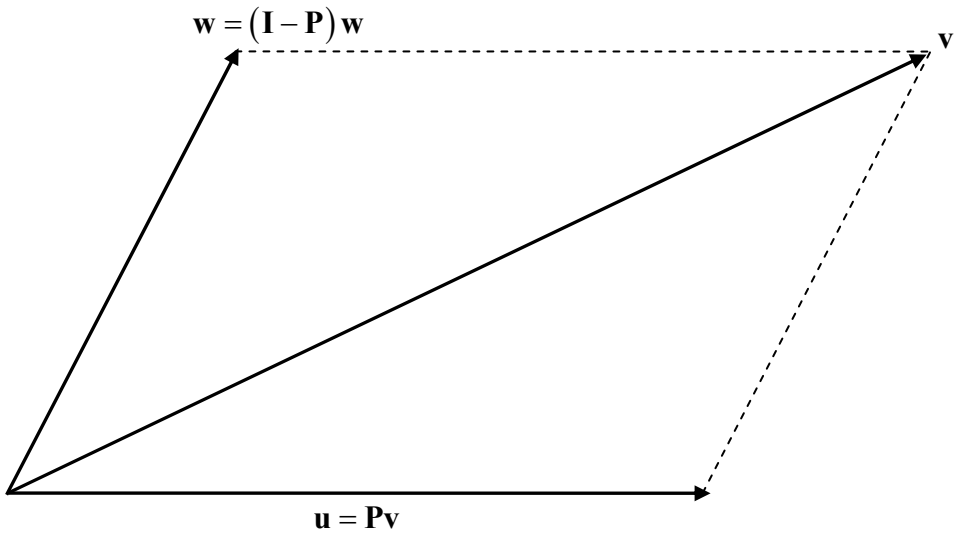

Figure 5

Given a projection $\mathbf{P}$, the linear transformation $\mathbf{I} - \mathbf{P}$ is also a projection. It is easily shown that

$$\mathscr{V} = R(\mathbf{I} - \mathbf{P}) \oplus K(\mathbf{I} - \mathbf{P})$$

and

$$R(\mathbf{I}-\mathbf{P})=K(\mathbf{P}),\quad K(\mathbf{I}-\mathbf{P})=R(\mathbf{P})$$

It follows from (17.16) that the restriction of $\mathbf{P}$ to $R(\mathbf{P})$ is the identity linear transformation on the subspace $R(\mathbf{P})$. Likewise, the restriction of $\mathbf{I}-\mathbf{P}$ to $K(\mathbf{P})$ is the identity linear transformation on $K(\mathbf{P})$. Theorem 17.3 is a special case of the following theorem.

Theorem 17.4. If $\mathbf{P}_k$, $k=1,...,R$, are projection operators with the properties that

$$\begin{aligned}\mathbf{P}_k^2&=\mathbf{P}_k, &\quad k&=1,...,R\\ \mathbf{P}_k\mathbf{P}_q&=\mathbf{0}, &\quad k&\neq q\end{aligned} \tag{17.17}$$

and

$$\mathbf{I}=\sum_{k=1}^{R}\mathbf{P}_k \tag{17.18}$$

then

$$\mathscr{V}=R(\mathbf{P}_1)\oplus R(\mathbf{P}_2)\oplus\cdots\oplus R(\mathbf{P}_R) \tag{17.19}$$

The proof of this theorem is left as an exercise for the reader. As a converse of Theorem 17.4, if $\mathscr{V}$ has the decomposition

$$\mathscr{V}=\mathscr{V}_1\oplus\cdots\oplus\mathscr{V}_R \tag{17.20}$$

then the endomorphisms $\mathbf{P}_k:\mathscr{V}\to\mathscr{V}$ defined by

$$\mathbf{P}_k\mathbf{v}=\mathbf{v}_k,\qquad k=1,...,R \tag{17.21}$$

where

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2 + \cdots + \mathbf{v}_R \tag{17.22}$$

are projections and satisfy (17.18). Moreover, $\mathscr{V}_k = R(\mathbf{P}_k)$, $k = 1, \ldots, R$.

Exercises

17.1 Let $\mathbf{A} : \mathscr{V} \to \mathscr{U}$ and $\mathbf{B} : \mathscr{U} \to \mathscr{V}$ be linear transformations. If $\mathbf{AB} = \mathbf{I}$, then $\mathbf{B}$ is the *right inverse* of $\mathbf{A}$. If $\mathbf{BA} = \mathbf{I}$, then $\mathbf{B}$ is the *left inverse* of $\mathbf{A}$. Show that $\mathbf{A}$ is an isomorphism if and only if it has a right inverse and a left inverse.

17.2 Show that if an endomorphism $\mathbf{A}$ of $\mathscr{V}$ commutes with *every* endomorphism $\mathbf{B}$ of $\mathscr{V}$, then $\mathbf{A}$ is a scalar multiple of $\mathbf{I}$.

17.3 Prove Theorem 17.4

17.4 An *involution* is an endomorphism $\mathbf{L}$ such that $\mathbf{L}^2 = \mathbf{I}$. Show that $\mathbf{L}$ is an involution if and only if $\mathbf{P} = \frac{1}{2}(\mathbf{L} + \mathbf{I})$ is a projection.

17.5 Consider the linear transformation $\mathbf{A}\big|_{\mathscr{V}^1} : \mathscr{V}^1 \to \mathscr{U}$ defined in Exercise 15.2. Show that if $\mathscr{V} = \mathscr{V}^1 \oplus K(\mathbf{A})$, then $\mathbf{A}\big|_{\mathscr{V}^1}$ is am isomorphism from $\mathscr{V}^1$ to $R(\mathbf{A})$.

17.6 If $\mathbf{A}$ is a linear transformation from $\mathscr{V}$ to $\mathscr{U}$ where $\dim \mathscr{V} = \dim \mathscr{U}$, assume that there exists a linear transformation $\mathbf{B} : \mathscr{U} \to \mathscr{V}$ such that $\mathbf{AB} = \mathbf{I}$ $(\text{or } \mathbf{BA} = \mathbf{I})$. Show that $\mathbf{A}$ is an isomorphism and that $\mathbf{B} = \mathbf{A}^{-1}$.

Section 18. The Adjoint of a Linear Transformation

The results in the earlier sections of this chapter did not make use of the inner product structure on $\mathscr{V}$. In this section, however, a particular inner product is needed to study the adjoint of a linear transformation as well as other ideas associated with the adjoint.

Given a linear transformation $\mathbf{A}:\mathscr{V}\to\mathscr{U}$, a function $\mathbf{A}^*:\mathscr{U}\to\mathscr{V}$ is called the *adjoint* of $\mathbf{A}$ if

$$\mathbf{u}\cdot(\mathbf{A}\mathbf{v})=(\mathbf{A}^*\mathbf{u})\cdot\mathbf{v} \tag{18.1}$$

for all $\mathbf{v}\in\mathscr{V}$ and $\mathbf{u}\in\mathscr{U}$. Observe that in (18.1) the inner product on the left side is the one in $\mathscr{U}$, while the one for the right side is the one in $\mathscr{V}$. Next we will want to examine the properties of the adjoint. It is probably worthy of note here that for linear transformations defined on real inner product spaces what we have called the *adjoint* is often called the *transpose.* Since our later applications are for real vector spaces, the name transpose is actually more important.

Theorem 18.1. For every linear transformation $\mathbf{A}:\mathscr{V}\to\mathscr{U}$, there exists a unique adjoint $\mathbf{A}^*:\mathscr{U}\to\mathscr{V}$ satisfying the condition (18.1).

Proof. Existence. Choose a basis $\{\mathbf{e}_1,\ldots,\mathbf{e}_N\}$ for $\mathscr{V}$ and a basis $\{\mathbf{b}_1,\ldots,\mathbf{b}_M\}$ for $\mathscr{U}$. Then $\mathbf{A}$ can be characterized by the $M\times N$ matrix $\left[A^\alpha{}_k\right]$ in such a way that

$$\mathbf{A}\mathbf{e}_k=\sum_{\alpha=1}^{M}A^\alpha{}_k\mathbf{b}_\alpha,\qquad k=1,...,N \tag{18.2}$$

This system suffices to define $\mathbf{A}$ since for any $\mathbf{v}\in\mathscr{V}$ with the representation

$$\mathbf{v}=\sum_{k=1}^{N}v^k\mathbf{e}_k$$

the corresponding representation of $\mathbf{A}\mathbf{v}$ is determined by $\left[A^\alpha{}_k\right]$ and $\left[v^k\right]$ by

$$\mathbf{A}\mathbf{v} = \sum_{\alpha=1}^{M}\left(\sum_{k=1}^{N} A^{\alpha}{}_{k} v^{k}\right)\mathbf{b}_{\alpha}$$

Now let $\{\mathbf{e}^1,\ldots,\mathbf{e}^N\}$ and $\{\mathbf{b}^1,...,\mathbf{b}^M\}$ be the reciprocal bases of $\{\mathbf{e}_1,\ldots,\mathbf{e}_N\}$ and $\{\mathbf{b}_1,\ldots,\mathbf{b}_M\}$, respectively. We shall define a linear transformation $\mathbf{A}^*$ by a system similar to (18.2) except that we shall use the reciprocal bases. Thus we put

$$\mathbf{A}^*\mathbf{b}^{\alpha} = \sum_{k=1}^{N} A^*{}_{k}{}^{\alpha}\mathbf{e}^{k} \tag{18.3}$$

where the matrix $\left[A^*{}_{k}{}^{\alpha}\right]$ is defined by

$$A^*{}_{k}{}^{\alpha} \equiv \overline{A}^{\alpha}{}_{k} \tag{18.4}$$

for all $\alpha = 1,...,M$ and $k = 1,...,N$. For any vector $\mathbf{u} \in \mathscr{U}$ the representation of $\mathbf{A}^*\mathbf{u}$ relative to $\{\mathbf{e}^1,...,\mathbf{e}^N\}$ is then given by

$$\mathbf{A}^*\mathbf{u} = \sum_{k=1}^{N}\left(\sum_{\alpha=1}^{M} A^*{}_{k}{}^{\alpha} u_{\alpha}\right)\mathbf{e}^{k}$$

where the representation of $\mathbf{u}$ itself relative to $\{\mathbf{b}^1,...,\mathbf{b}^M\}$ is

$$\mathbf{u} = \sum_{\alpha=1}^{M} u_{\alpha}\mathbf{b}^{\alpha}$$

Having defined the linear transformation $\mathbf{A}^*$, we now verify that $\mathbf{A}$ and $\mathbf{A}^*$ satisfy the relation (18.1). Since $\mathbf{A}$ and $\mathbf{A}^*$ are both linear transformations, it suffices to check (18.1) for $\mathbf{u}$ equal to an arbitrary element of the basis $\{\mathbf{b}^1,...,\mathbf{b}^M\}$, say $\mathbf{u} = \mathbf{b}^{\alpha}$, and for $\mathbf{v}$ equal to an arbitrary element of the basis $\{\mathbf{e}_1,...,\mathbf{e}_N\}$, say $\mathbf{v} = \mathbf{e}_k$. For this choice of $\mathbf{u}$ and $\mathbf{v}$ we obtain from (18.2)

$$\mathbf{b}^{\alpha} \cdot \mathbf{A}\mathbf{e}_k = \mathbf{b}^{\alpha} \cdot \sum_{\beta=1}^{M} A^{\beta}{}_{k} \mathbf{b}_{\beta} = \sum_{\beta=1}^{M} \overline{A}^{\beta}{}_{k} \delta_{\beta}^{\alpha} = \overline{A}^{\alpha}{}_{k} \tag{18.5}$$

and likewise from (18.3)

$$\mathbf{A}^{*}\mathbf{b}^{\alpha} \cdot \mathbf{e}_k = \left(\sum_{l=1}^{N} A^{*}{}_{l}{}^{\alpha} \mathbf{e}^{l} \right) \cdot \mathbf{e}_k = \sum_{l=1}^{N} A^{*}{}_{l}{}^{\alpha} \delta_{k}^{l} = A^{*}{}_{k}{}^{\alpha} \tag{18.6}$$

Comparing (18.5) and (18.6) with (18.4), we see that the linear transformation $\mathbf{A}^{*}$ defined by (18.3) satisfies the condition

$$\mathbf{b}^{\alpha} \cdot \mathbf{A}\mathbf{e}_k = \mathbf{A}^{*}\mathbf{b}^{\alpha} \cdot \mathbf{e}_k$$

for all $\alpha = 1, ..., M$ and $k = 1, ..., N$, and hence also the condition (18.1) for all $\mathbf{u} \in \mathscr{U}$ and $\mathbf{v} \in \mathscr{V}$.

Uniqueness. Assume that there are two functions $\mathbf{A}_1^{*} : \mathscr{U} \to \mathscr{V}$ and $\mathbf{A}_2^{*} : \mathscr{U} \to \mathscr{V}$ which satisfy (18.1). Then

$$\left(\mathbf{A}_1^{*}\mathbf{u} \right) \cdot \mathbf{v} = \left(\mathbf{A}_2^{*}\mathbf{u} \right) \cdot \mathbf{v} = \mathbf{u} \cdot \left(\mathbf{A}\mathbf{v} \right)$$

Thus

$$\left(\mathbf{A}_1^{*}\mathbf{u} - \mathbf{A}_2^{*}\mathbf{u} \right) \cdot \mathbf{v} = 0$$

Since the last formula must hold for all $\mathbf{v} \in \mathscr{V}$, the inner product properties show that

$$\mathbf{A}_1^{*}\mathbf{u} = \mathbf{A}_2^{*}\mathbf{u}$$

This formula must hold for every $\mathbf{u} \in \mathscr{U}$ and thus

$$\mathbf{A}_1^{*} = \mathbf{A}_2^{*}$$

As a corollary to the preceding theorem we see that the adjoint $\mathbf{A}^*$ of a linear transformation $\mathbf{A}$ is a linear transformation. Further, the matrix $\left[A^*{}_k{}^\alpha\right]$ that characterizes $\mathbf{A}^*$ by (18.3) is related to the matrix $\left[A^k{}_\alpha\right]$ that characterizes $\mathbf{A}$ by (18.2). Notice the choice of bases in (18.2) and (18.3), however.

Other properties of the adjoint are summarized in the following theorem.

Theorem 18.2.

(a) $\left(\mathbf{A}+\mathbf{B}\right)^* = \mathbf{A}^* + \mathbf{B}^*$ $(18.7)_1$

(b) $\left(\mathbf{AB}\right)^* = \mathbf{B}^*\mathbf{A}^*$ $(18.7)_2$

(c) $\left(\lambda\mathbf{A}\right)^* = \bar{\lambda}\mathbf{A}^*$ $(18.7)_3$

(d) $\mathbf{0}^* = \mathbf{0}$ $(18.7)_4$

(e) $\mathbf{I}^* = \mathbf{I}$ $(18.7)_5$

(f) $\left(\mathbf{A}^*\right)^* = \mathbf{A}$ $(18.7)_6$

and

(g) If $\mathbf{A}$ is nonsingular, so is $\mathbf{A}^*$ and in addition,

$$\mathbf{A}^{*-1} = \mathbf{A}^{-1*} \tag{18.8}$$

In (a), $\mathbf{A}$ and $\mathbf{B}$ are in $\mathscr{L}\left(\mathscr{V};\mathscr{U}\right)$; in (b), $\mathbf{B}\in\mathscr{L}\left(\mathscr{V};\mathscr{U}\right)$ and $\mathbf{A}\in\mathscr{L}\left(\mathscr{U};\mathscr{W}\right)$; in (c), $\lambda\in\mathscr{C}$; in(d), $\mathbf{0}$ is the zero element; and in (e), $\mathbf{I}$ is the identity element in $\mathscr{L}\left(\mathscr{V};\mathscr{V}\right)$. The proof of the above theorem is straightforward and is left as an exercise for the reader.

Theorem 18.3. If $\mathbf{A}:\mathscr{V}\to\mathscr{U}$ is a linear transformation, then $\mathscr{V}$ and $\mathscr{U}$ have the orthogonal decompositions

$$\mathscr{V} = R\left(\mathbf{A}^*\right)\oplus K\left(\mathbf{A}\right) \tag{18.9}$$

and

$$\mathscr{U} = R(\mathbf{A}) \oplus K(\mathbf{A}^*) \tag{18.10}$$

where

$$R(\mathbf{A}^*) = K(\mathbf{A})^\perp \tag{18.11}$$

and

$$R(\mathbf{A}) = K(\mathbf{A}^*)^\perp \tag{18.12}$$

Proof. We shall prove (18.11) and (18.12) and then apply Theorem 13.4 to obtain (18.9) and (18.10). Let $\mathbf{u}$ be and arbitrary element in $K(\mathbf{A}^*)$. Then for every $\mathbf{v} \in \mathscr{V}$, $\mathbf{u} \cdot (\mathbf{A}\mathbf{v}) = (\mathbf{A}^*\mathbf{u}) \cdot \mathbf{v} = 0$. Thus $K(\mathbf{A}^*)$ is contained in $R(\mathbf{A})^\perp$. Conversely, take $\mathbf{u} \in R(\mathbf{A})^\perp$; then for every $\mathbf{v} \in \mathscr{V}$, $(\mathbf{A}^*\mathbf{u}) \cdot \mathbf{v} = \mathbf{u} \cdot (\mathbf{A}\mathbf{v}) = 0$, and thus $(\mathbf{A}^*\mathbf{u}) = \mathbf{0}$, which implies that $\mathbf{u} \in K(\mathbf{A}^*)$ and that $R(\mathbf{A})^\perp$ is in $K(\mathbf{A}^*)$. Therefore, $R(\mathbf{A})^\perp = K(\mathbf{A}^*)$, which by Exercise 13.3 implies (18.11). Equation (18.11) follows by an identical argument with $\mathbf{A}$ replaced by $\mathbf{A}^*$. As mentioned above, (18.9) and (18.10) now follow from Theorem 13.4.

Theorem 18.4. Given a linear transformation $\mathbf{A} : \mathscr{V} \to \mathscr{U}$, then $\mathbf{A}$ and $\mathbf{A}^*$ have the same rank.

Proof. By application of Theorems 10.9 and 18.3,

$$\dim \mathscr{V} = \dim R(\mathbf{A}^*) + \dim K(\mathbf{A})$$

However, by (15.6),

$$\dim \mathscr{V} = \dim R(\mathbf{A}) + \dim K(\mathbf{A})$$

Therefore,

$$\dim R(\mathbf{A}) = \dim R(\mathbf{A}^*) \tag{18.13}$$

which is the desired result.

An endomorphism $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ is called *Hermitian* if $\mathbf{A} = \mathbf{A}^*$ and *skew-Hermitian* if $\mathbf{A} = -\mathbf{A}^*$. It should be pointed out, however, that the terms *symmetric* and *skew-symmetric* are often used instead of Hermitian and skew-Hermitian for linear transformations defined on real inner product spaces. The following theorem, which follows directly from the above definitions and from (18.1), characterizes Hermitian and skew-Hermitian endomorphisms.

Theorem 18.5. An endomorphism $\mathbf{A}$ is Hermitian if and only if

$$\mathbf{v}_1 \cdot (\mathbf{A}\mathbf{v}_2) = (\mathbf{A}\mathbf{v}_1) \cdot \mathbf{v}_2 \tag{18.14}$$

for all $\mathbf{v}_1, \mathbf{v}_2 \in \mathscr{V}$, and it is skew-Hermitian if and only if

$$\mathbf{v}_1 \cdot (\mathbf{A}\mathbf{v}_2) = -(\mathbf{A}\mathbf{v}_1) \cdot \mathbf{v}_2 \tag{18.15}$$

for all $\mathbf{v}_1, \mathbf{v}_2 \in \mathscr{V}$.

We shall denote by $\mathscr{S}(\mathscr{V};\mathscr{V})$ and $\mathscr{A}(\mathscr{V};\mathscr{V})$ the subsets of $\mathscr{L}(\mathscr{V};\mathscr{V})$ defined by

$$\mathscr{S}(\mathscr{V};\mathscr{V}) = \{\mathbf{A} \,|\, \mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V}) \text{ and } \mathbf{A} = \mathbf{A}^*\}$$

and

$$\mathscr{A}(\mathscr{V};\mathscr{V}) = \{\mathbf{A} \,|\, \mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V}) \text{ and } \mathbf{A} = -\mathbf{A}^*\}$$

In the special case of a *real* inner product space, it is easy to show that $\mathscr{S}(\mathscr{V};\mathscr{V})$ and $\mathscr{A}(\mathscr{V};\mathscr{V})$ are both *subspaces* of $\mathscr{L}(\mathscr{V};\mathscr{V})$. In particular, $\mathscr{L}(\mathscr{V};\mathscr{V})$ has the following decomposition:

Theorem 18.6. For a real inner product space,

$$\mathscr{L}(\mathscr{V};\mathscr{V}) = \mathscr{T}(\mathscr{V};\mathscr{V}) \oplus \mathscr{A}(\mathscr{V};\mathscr{V}) \tag{18.16}$$

Proof. An arbitrary element $\mathbf{L} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ can always be written

$$\mathbf{L} = \mathbf{S} + \mathbf{A} \tag{18.17}$$

where

$$\mathbf{S} = \frac{1}{2}\left(\mathbf{L} + \mathbf{L}^T\right) = \mathbf{S}^T \quad \text{and} \quad \mathbf{A} = \frac{1}{2}\left(\mathbf{L} - \mathbf{L}^T\right) = -\mathbf{A}^T$$

Here the superscript T denotes that transpose, which is the specialization of the adjoint for a real inner product space. Since $\mathbf{S} \in \mathscr{S}(\mathscr{V};\mathscr{V})$ and $\mathbf{A} \in \mathscr{A}(\mathscr{V};\mathscr{V})$, (18.17) shows that

$$\mathscr{L}(\mathscr{V};\mathscr{V}) = \mathscr{S}(\mathscr{V};\mathscr{V}) + \mathscr{A}(\mathscr{V};\mathscr{V})$$

Now, let $\mathbf{B} \in \mathscr{S}(\mathscr{V};\mathscr{V}) \cap \mathscr{A}(\mathscr{V};\mathscr{V})$. Then $\mathbf{B}$ must satisfy the conditions

$$\mathbf{B} = \mathbf{B}^T \quad \text{and} \quad \mathbf{B} = -\mathbf{B}^T$$

Thus, $\mathbf{B} = \mathbf{0}$ and the proof is complete.

We shall see in the exercises at the end of Section 19 that a real inner product can be defined on $\mathscr{L}(\mathscr{V};\mathscr{V})$ in such a fashion that the subspace of symmetric endomorphisms is the orthogonal complement of the subspace of skew-symmetric endomorphisms. For *complex* inner product spaces, however, $\mathscr{S}(\mathscr{V};\mathscr{V})$ and $\mathscr{A}(\mathscr{V};\mathscr{V})$ *are not subspaces* of $\mathscr{L}(\mathscr{V};\mathscr{V})$. For example, if $\mathbf{A} \in \mathscr{S}(\mathscr{V};\mathscr{V})$, then $i\mathbf{A}$ is in $\mathscr{A}(\mathscr{V};\mathscr{V})$ because $(i\mathbf{A})^* = \bar{i}\mathbf{A}^* = -i\mathbf{A}^* = -i\mathbf{A}$, where $(18.7)_3$ has been used.

A linear transformation $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{U})$ is *unitary* if

$$\mathbf{A}\mathbf{v}_2 \cdot \mathbf{A}\mathbf{v}_1 = \mathbf{v}_2 \cdot \mathbf{v}_1 \tag{18.18}$$

for all $\mathbf{v}_1, \mathbf{v}_2 \in \mathscr{V}$. However, for a real inner product space the above condition defines $\mathbf{A}$ to be *orthogonal*. Essentially, (18.18) asserts that unitary (or orthogonal) linear transformations preserve the inner products.

Theorem 18.7. If $\mathbf{A}$ is unitary, then it is regular.

Proof. Take $\mathbf{v}_1 = \mathbf{v}_2 = \mathbf{v}$ in (18.18), and by (12.1) we find

$$\|\mathbf{A}\mathbf{v}\| = \|\mathbf{v}\| \tag{18.19}$$

Thus, if $\mathbf{A}\mathbf{v} = \mathbf{0}$, then $\mathbf{v} = \mathbf{0}$, which proves the theorem.

Theorem 18.8. $\mathbf{A} \in \mathscr{L}(\mathscr{V}; \mathscr{U})$ is unitary if and only if $\|\mathbf{A}\mathbf{v}\| = \|\mathbf{v}\|$ for all $\mathbf{v} \in \mathscr{V}$.

Proof. If $\mathbf{A}$ is unitary, we saw in the proof of Theorem 18.7 that $\|\mathbf{A}\mathbf{v}\| = \|\mathbf{v}\|$. Thus, we shall assume $\|\mathbf{A}\mathbf{v}\| = \|\mathbf{v}\|$ for all $\mathbf{v} \in \mathscr{V}$ and attempt to derive (18.18). This derivation is routine because, by the polar identity of Exercise 12.1,

$$2\mathbf{A}\mathbf{v}_1 \cdot \mathbf{A}\mathbf{v}_2 = \|\mathbf{A}(\mathbf{v}_1 + \mathbf{v}_2)\|^2 + i\|\mathbf{A}(\mathbf{v}_1 + i\mathbf{v}_2)\|^2 - (1+i)\left(\|\mathbf{A}\mathbf{v}_1\|^2 + \|\mathbf{A}\mathbf{v}_2\|^2\right)$$

Therefore, by (18.19),

$$2\mathbf{A}\mathbf{v}_1 \cdot \mathbf{A}\mathbf{v}_2 = \|\mathbf{v}_1 + \mathbf{v}_2\|^2 + i\|\mathbf{v}_1 + i\mathbf{v}_2\|^2 - (1+i)\left(\|\mathbf{v}_1\|^2 + \|\mathbf{v}_2\|^2\right) = 2\mathbf{v}_1 \cdot \mathbf{v}_2$$

This proof cannot be specialized directly to a real inner product space since the polar identity is valid for a complex inner product space only. We leave the proof for the real case as an exercise.

If we require $\mathscr{V}$ and $\mathscr{U}$ to have the same dimension, then Theorem 15.10 ensures that a unitary transformation $\mathbf{A}$ is an isomorphism. In the case we can use (18.1) and (18.18) and conclude the following

Theorem 18.9. Given a linear transformation $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{U})$, where $\dim \mathscr{V} = \dim \mathscr{U}$; then $\mathbf{A}$ is unitary if and only if it is an isomorphism whose inverse satisfies

$$\mathbf{A}^{-1} = \mathbf{A}^{*} \tag{18.20}$$

Recall from Theorem 15.5 that a regular linear transformation maps linearly independent vectors into linearly independent vectors. Therefore if $\{\mathbf{e}_1,\ldots,\mathbf{e}_N\}$ is a basis for $\mathscr{V}$ and $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{U})$ is regular, then $\{\mathbf{A}\mathbf{e}_1,\ldots,\mathbf{A}\mathbf{e}_N\}$ is basis for $R(\mathbf{A})$ which is a subspace in $\mathscr{U}$. If $\{\mathbf{e}_1,\ldots,\mathbf{e}_N\}$ is orthonormal and $\mathbf{A}$ is unitary, it easily follows that $\{\mathbf{A}\mathbf{e}_1,\ldots,\mathbf{A}\mathbf{e}_N\}$ is also orthonormal. Thus the image of an orthonormal basis under a unitary transformation is also an orthonormal set. Conversely, a linear transformation which sends an orthonormal basis of $\mathscr{V}$ into an orthonormal basis of $R(\mathbf{A})$ must be unitary. To prove this assertion, let $\{\mathbf{e}_1,\ldots,\mathbf{e}_N\}$ be orthonormal, and let $\{\mathbf{b}_1,\ldots,\mathbf{b}_N\}$ be an orthonormal set in $\mathscr{U}$, where $\mathbf{b}_k = \mathbf{A}\mathbf{e}_k$, $k = 1,\ldots,N$. Then, if $\mathbf{v}_1$ and $\mathbf{v}_2$ are arbitrary elements of $\mathscr{V}$,

$$\mathbf{v}_1 = \sum_{k=1}^{N} v_1{}^k \mathbf{e}_k \qquad \text{and} \qquad \mathbf{v}_2 = \sum_{l=1}^{N} v_2{}^l \mathbf{e}_l$$

we have

$$\mathbf{A}\mathbf{v}_1 = \sum_{k=1}^{N} v_1{}^k \mathbf{b}_k \qquad \text{and} \qquad \mathbf{A}\mathbf{v}_2 = \sum_{l=1}^{N} v_2{}^k \mathbf{b}_l$$

and, thus,

$$\mathbf{A}\mathbf{v}_1 \cdot \mathbf{A}\mathbf{v}_2 = \sum_{k=1}^{N}\sum_{l=1}^{N} v_1{}^k \overline{v}_2{}^l \mathbf{b}_k \cdot \mathbf{b}_l = \sum_{k=1}^{N} v_1{}^k \overline{v}_2{}^l = \mathbf{v}_1 \cdot \mathbf{v}_2 \tag{18.21}$$

Equation (18.21) establishes the desired result.

Recall that in Section 17 we introduced the *general linear group* $\mathscr{GL}(\mathscr{V})$. We define a subset $\mathscr{U}(\mathscr{V})$ of $\mathscr{GL}(\mathscr{V})$ by

$$\mathscr{U}(\mathscr{V}) = \left\{ \mathbf{A} \middle| \mathbf{A} \in \mathscr{GL}(\mathscr{V}) \text{ and } \mathbf{A}^{-1} = \mathbf{A}^* \right\}$$

which is easily shown to be a *subgroup*. This subgroup is called the *unitary* group of $\mathscr{V}$.

In Section 17 we have defined projections $\mathbf{P} : \mathscr{V} \to \mathscr{V}$ by the characteristic property

$$\mathbf{P}^2 = \mathbf{P} \tag{18.22}$$

In particular, we have showed in Theorem 17.3 that $\mathscr{V}$ has the decomposition

$$\mathscr{V} = R(\mathbf{P}) \oplus K(\mathbf{P}) \tag{18.23}$$

There are several additional properties of projections which are worthy of discussion here.

Theorem 18.10. If $\mathbf{P}$ is a projection, then $\mathbf{P} = \mathbf{P}^* \Leftrightarrow R(\mathbf{P}) = K(\mathbf{P})^{\perp}$.

Proof. First take $\mathbf{P} = \mathbf{P}^*$ and let $\mathbf{v}$ be an arbitrary element of $\mathscr{V}$. Then by (18.23)

$$\mathbf{v} = \mathbf{u} + \mathbf{w}$$

where $\mathbf{u} = \mathbf{P}\mathbf{v}$ and $\mathbf{w} = \mathbf{v} - \mathbf{P}\mathbf{v}$. Then

$$\begin{aligned} \mathbf{u} \cdot \mathbf{w} &= \mathbf{P}\mathbf{v} \cdot (\mathbf{v} - \mathbf{P}\mathbf{v}) \\ &= (\mathbf{P}\mathbf{v}) \cdot \mathbf{v} - (\mathbf{P}\mathbf{v}) \cdot \mathbf{P}\mathbf{v} \\ &= (\mathbf{P}\mathbf{v}) \cdot \mathbf{v} - (\mathbf{P}^*\mathbf{P}\mathbf{v}) \cdot \mathbf{v} \\ &= (\mathbf{P}\mathbf{v}) \cdot \mathbf{v} - (\mathbf{P}^2\mathbf{v}) \cdot \mathbf{v} \\ &= 0 \end{aligned}$$

where (18.22), (18.1), and the assumption that $\mathbf{P}$ is Hermitian have been used. Conversely, assume $\mathbf{u} \cdot \mathbf{w} = 0$ for all $\mathbf{u} \in R(\mathbf{P})$ and all $\mathbf{w} \in K(\mathbf{P})$. Then if $\mathbf{v}_1$ and $\mathbf{v}_2$ are arbitrary vectors in $\mathscr{V}$

$$\mathbf{P}\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{P}\mathbf{v}_1 \cdot \left(\mathbf{P}\mathbf{v}_2 + \mathbf{v}_2 - \mathbf{P}\mathbf{v}_2\right) = \mathbf{P}\mathbf{v}_1 \cdot \mathbf{P}\mathbf{v}_2$$

and, by interchanging $\mathbf{v}_1$ and $\mathbf{v}_2$

$$\mathbf{P}\mathbf{v}_2 \cdot \mathbf{v}_1 = \mathbf{P}\mathbf{v}_2 \cdot \mathbf{P}\mathbf{v}_1$$

Therefore,

$$\mathbf{P}\mathbf{v}_1 \cdot \mathbf{v}_2 = \overline{\mathbf{P}\mathbf{v}_2 \cdot \mathbf{v}_1} = \mathbf{v}_1 \cdot \left(\mathbf{P}\mathbf{v}_2\right)$$

This last result and Theorem 18.5 show that $\mathbf{P}$ is Hermitian.

Because of Theorem 18.10, Hermitian projections are called *perpendicular projections.* In Section 13 we introduced the concept of an orthogonal complement of a subspace of an inner product space. A similar concept is that of an orthogonal pair of subspaces. If $\mathscr{V}_1$ and $\mathscr{V}_2$ are subspaces of $\mathscr{V}$, they are orthogonal, written $\mathscr{V}_1 \perp \mathscr{V}_2$, if $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$ for all $\mathbf{v}_1 \in \mathscr{V}_1$ and $\mathbf{v}_2 \in \mathscr{V}_2$

Theorem 18.11. If $\mathscr{V}_1$ and $\mathscr{V}_2$ are subspaces of $\mathscr{V}$, $\mathbf{P}_1$ is the perpendicular projection of $\mathscr{V}$ onto $\mathscr{V}_1$, and $\mathbf{P}_2$ is the perpendicular projection of $\mathbf{P}$ onto $\mathscr{V}_2$, then $\mathscr{V}_1 \perp \mathscr{V}_2$ if and only if $\mathbf{P}_2\mathbf{P}_1 = \mathbf{0}$.

Proof. Assume that $\mathscr{V}_1 \perp \mathscr{V}_2$; then $\mathbf{P}_1\mathbf{v} \in \mathscr{V}_2^{\perp}$ for all $\mathbf{v} \in \mathscr{V}$ and thus $\mathbf{P}_2\mathbf{P}_1\mathbf{v} = \mathbf{0}$ for all $\mathbf{v} \in \mathscr{V}$ which yields $\mathbf{P}_2\mathbf{P}_1 = \mathbf{0}$. Next assume $\mathbf{P}_2\mathbf{P}_1 = \mathbf{0}$; this implies that $\mathbf{P}_1\mathbf{v} \in \mathscr{V}_2^{\perp}$ for every $\mathbf{v} \in \mathscr{V}$. Therefore $\mathscr{V}_1$ is contained in $\mathscr{V}_2^{\perp}$ and, as a result, $\mathscr{V}_1 \perp \mathscr{V}_2$.

Exercises

18.1 Prove Theorem 18.2

18.2 For a real inner product space, prove that $\dim \mathscr{S}\left(\mathscr{V};\mathscr{V}\right) = \frac{1}{2}N\left(N+1\right)$ and $\dim \mathscr{A}\left(\mathscr{V};\mathscr{V}\right) = \frac{1}{2}N\left(N-1\right)$, where $N = \dim \mathscr{V}$.

18.3 Define $\mathbf{\Phi} : \mathscr{L}\left(\mathscr{V};\mathscr{V}\right) \to \mathscr{L}\left(\mathscr{V};\mathscr{V}\right)$ and $\mathbf{\Psi} : \mathscr{L}\left(\mathscr{V};\mathscr{V}\right) \to \mathscr{L}\left(\mathscr{V};\mathscr{V}\right)$ by

$$\mathbf{\Phi A} = \frac{1}{2}\left(\mathbf{A} + \mathbf{A}^T\right) \qquad \text{and} \qquad \mathbf{\Psi A} = \frac{1}{2}\left(\mathbf{A} - \mathbf{A}^T\right)$$

where $\mathscr{V}$ is a *real* inner product space. Show that $\mathbf{\Phi}$ and $\mathbf{\Psi}$ are projections.

18.4 Let $\mathscr{V}_1$ and $\mathscr{V}_2$ be subspaces of $\mathscr{V}$ and let $\mathbf{P}_1$ be the perpendicular projection of $\mathscr{V}$ onto $\mathscr{V}_1$ and $\mathbf{P}_2$ be the perpendicular projection of $\mathscr{V}$ onto $\mathscr{V}_2$. Show that $\mathbf{P}_1 + \mathbf{P}_2$ is a perpendicular projection if and only if $\mathscr{V}_1 \perp \mathscr{V}_2$.

18.5 Let $\mathbf{P}_1$ and $\mathbf{P}_2$ be the projections introduced in Exercise 18.4. Show that $\mathbf{P}_1 - \mathbf{P}_2$ is a projection if and only if $\mathscr{V}_2$ is a subspace of $\mathscr{V}_1$.

18.6 Show that $\mathbf{A} \in \mathscr{L}\left(\mathscr{V};\mathscr{V}\right)$ is skew-symmetric $\Leftrightarrow \mathbf{v} \cdot \mathbf{Av}=0$ for all $\mathbf{v} \in \mathscr{V}$, where $\mathscr{V}$ is real vector space

18.7 A linear transformation $\mathbf{A} \in \mathscr{L}\left(\mathscr{V};\mathscr{V}\right)$ is *normal* if $\mathbf{A}^*\mathbf{A} = \mathbf{A}\mathbf{A}^*$. Show that $\mathbf{A}$ is normal if and only if

$$\mathbf{Av}_1 \cdot \mathbf{Av}_2 = \mathbf{A}^*\mathbf{v}_1 \cdot \mathbf{A}^*\mathbf{v}_2$$

for all $\mathbf{v}_1, \mathbf{v}_2$ in $\mathscr{V}$.

18.8 Show that $\mathbf{v} \cdot \left(\left(\mathbf{A} + \mathbf{A}^*\right)\mathbf{v}\right)$ is real for every $\mathbf{v} \in \mathscr{V}$ and every $\mathbf{A} \in \mathscr{L}\left(\mathscr{V};\mathscr{V}\right)$.

18.9 Show that every linear transformation $\mathbf{A} \in \mathscr{L}\left(\mathscr{V};\mathscr{V}\right)$, where $\mathscr{V}$ is a complex inner product space, has the unique decomposition

$$\mathbf{A} = \mathbf{B} + i\mathbf{C}$$

where $\mathbf{B}$ and $\mathbf{C}$ are both Hermitian.

18.10 Prove Theorem 18.8 for real inner product spaces. Hint: A polar identity for a real inner product space is

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

18.11 Let $\mathbf{A}$ be an endomorphism of $\mathscr{R}^3$ whose matrix relative to the standard basis is

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

What is $K(\mathbf{A})$? What is $R(\mathbf{A}^T)$? Check the results of Theorem 18.3 for this particular endomorphism.

Section 19. Component Formulas

In this section we shall introduce the components of a linear transformation and several related ideas. Let $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{U})$, $\{\mathbf{e}_1,\ldots,\mathbf{e}_N\}$ a basis for $\mathscr{V}$, and $\{\mathbf{b}_1,\ldots,\mathbf{b}_M\}$ a basis for $\mathscr{U}$. The vector $\mathbf{A}\mathbf{e}_k$ is in $\mathscr{U}$ and, as a result, can be expanded in a basis of $\mathscr{U}$ in the form

$$\mathbf{A}\mathbf{e}_k = \sum_{\alpha=1}^{M} A_k{}^{\alpha}\mathbf{b}_\alpha \tag{19.1}$$

The MN scalars $A_k{}^{\alpha}$ $(\alpha = 1,\ldots,M; k = 1,\ldots,N)$ are called the *components* of $\mathbf{A}$ with respect to the bases. If $\{\mathbf{b}^1,\ldots,\mathbf{b}^M\}$ is a basis of $\mathscr{U}$ which is reciprocal to $\{\mathbf{b}_1,\ldots,\mathbf{b}_M\}$, (19.1) yields

$$A^{\alpha}{}_k = (\mathbf{A}\mathbf{e}_k)\cdot\mathbf{b}^{\alpha} \tag{19.2}$$

Under change of bases in $\mathscr{V}$ and $\mathscr{U}$ defined by [cf. (14.18) and $(14.22)_1$]

$$\mathbf{e}_k = \sum_{j=1}^{N} \hat{T}_k{}^{j}\hat{\mathbf{e}}_j \tag{19.3}$$

and

$$\mathbf{b}^{\alpha} = \sum_{\beta=1}^{M} \overline{S^{\alpha}_{\beta}}\hat{\mathbf{b}}^{\beta} \tag{19.4}$$

(19.2) can be used to derive the following transformation rule for the components of $\mathbf{A}$:

$$A^{\alpha}{}_k = \sum_{\beta=1}^{M}\sum_{j=1}^{N} S^{\alpha}_{\beta}\hat{A}^{\beta}{}_j\hat{T}_k{}^{j} \tag{19.5}$$

where

$$\hat{A}^{\beta}{}_{j} = \left(\mathbf{A}\hat{\mathbf{e}}_{j}\right)\cdot\hat{\mathbf{b}}^{\beta}$$

If $\mathbf{A}\in\mathscr{L}\left(\mathscr{V};\mathscr{V}\right)$, (19.1) and (19.5) specialize to

$$\mathbf{A}\mathbf{e}_{k} = \sum_{q=1}^{N} A^{q}{}_{k}\mathbf{e}_{q} \tag{19.6}$$

and

$$A^{q}{}_{k} = \sum_{s,j=1}^{N} T_{s}^{q}\hat{A}^{s}{}_{j}\hat{T}_{k}^{j} \tag{19.7}$$

The *trace* of an endomorphism is a function $\mathrm{tr}:\mathscr{L}\left(\mathscr{V};\mathscr{V}\right)\to\mathscr{C}$ defined by

$$\mathrm{tr}\,\mathbf{A} = \sum_{k=1}^{N} A^{k}{}_{k} \tag{19.8}$$

It easily follows from (19.7) and (14.21) that $\mathrm{tr}\,\mathbf{A}$ is independent of the choice of basis of $\mathscr{V}$. Later we shall give a definition of the trace which does not employ the use of a basis.

If $\mathbf{A}\in\mathscr{L}\left(\mathscr{V};\mathscr{U}\right)$, then $\mathbf{A}^{*}\in\mathscr{L}\left(\mathscr{U};\mathscr{V}\right)$. The components of $\mathbf{A}^{*}$ are obtained by the same logic as was used in obtaining (19.1). For example,

$$\mathbf{A}^{*}\mathbf{b}^{\alpha} = \sum_{k=1}^{N} A^{*}{}_{k}{}^{\alpha}\mathbf{e}^{k} \tag{19.9}$$

where the MN scalars $A^{*}{}_{k}{}^{\alpha}$ $\left(k=1,...,N;\alpha=1,...,M\right)$ are the components of $\mathbf{A}^{*}$ with respect to $\left\{\mathbf{b}^{1},...,\mathbf{b}^{M}\right\}$ and $\left\{\mathbf{e}^{1},...,\mathbf{e}^{N}\right\}$. From the proof of Theorem 18.1 we can relate these components of $\mathbf{A}^{*}$ to those of $\mathbf{A}$ in (19.1); namely,

$$A^{*\,\alpha}_{\ s} = \overline{A^{\alpha}_{\ s}} \tag{19.10}$$

If the inner product spaces $\mathscr{U}$ and $\mathscr{V}$ are real, (19.10) reduces to

$$A^{T\,\alpha}_{\ s} = A^{\alpha}_{\ s} \tag{19.11}$$

By the same logic which produced (19.1), we can also write

$$\mathbf{A}\mathbf{e}_k = \sum_{\alpha=1}^{M} A_{\alpha k}\mathbf{b}^{\alpha} \tag{19.12}$$

and

$$\mathbf{A}\mathbf{e}_k = \sum_{\alpha=1}^{M} A^{\alpha k}\mathbf{b}_{\alpha} = \sum_{\alpha=1}^{M} A_{\alpha}^{\ k}\mathbf{b}^{\alpha} \tag{19.13}$$

If we use (14.6) and (14.9), the various components of $\mathbf{A}$ are related by

$$A^{\alpha}_{\ k} = \sum_{s=1}^{N} A^{\alpha s}e_{ks} = \sum_{\beta=1}^{M}\sum_{s=1}^{N} b^{\beta\alpha}A_{\beta}^{\ s}e_{ks} = \sum_{\beta=1}^{M} A_{\beta k}b^{\beta\alpha} \tag{19.14}$$

where

$$b_{\alpha\beta} = \mathbf{b}_{\alpha}\cdot\mathbf{b}_{\beta} = \overline{b_{\beta\alpha}} \qquad \text{and} \qquad b^{\alpha\beta} = \mathbf{b}^{\alpha}\cdot\mathbf{b}^{\beta} = \overline{b^{\beta\alpha}} \tag{19.15}$$

A similar set of formulas holds for the components of $\mathbf{A}^*$. Equations (19.14) and (19.10) can be used to obtain these formulas. The transformation rules for the components defined by (19.12) and (19.13) are easily established to be

$$A_{\alpha k} = \left(\mathbf{A}\mathbf{e}_k\right)\cdot\mathbf{b}_{\alpha} = \sum_{\beta=1}^{M}\sum_{j=1}^{N} \overline{\hat{S}^{\beta}_{\alpha}}\hat{A}_{\beta j}\hat{T}^{j}_{k} \tag{19.16}$$

$$A^{\alpha k} = \left(\mathbf{A}\mathbf{e}^k\right)\cdot\mathbf{b}^\alpha = \sum_{\beta=1}^{M}\sum_{j=1}^{N} S_\beta^\alpha \hat{A}^{\beta j}\overline{T_j^k} \tag{19.17}$$

and

$$A_\alpha{}^k = \left(\mathbf{A}\mathbf{e}^k\right)\cdot\mathbf{b}_\alpha = \sum_{\beta=1}^{M}\sum_{j=1}^{N} \overline{\hat{S}_\alpha^\beta}\hat{A}_\beta{}^j\overline{T_j^k} \tag{19.18}$$

where

$$\hat{A}_{\beta j} = \left(\mathbf{A}\hat{\mathbf{e}}_j\right)\cdot\hat{\mathbf{b}}_\beta \tag{19.19}$$

$$\hat{A}^{\beta j} = \left(\mathbf{A}\hat{\mathbf{e}}^j\right)\cdot\hat{\mathbf{b}}^\beta \tag{19.20}$$

and

$$\hat{A}_\beta{}^j = \left(\mathbf{A}\hat{\mathbf{e}}^j\right)\cdot\hat{\mathbf{b}}_\beta \tag{19.21}$$

The quantities $\hat{S}_\alpha^\beta$ introduced in (19.16) and (19.18) above are related to the quantities S_β^α by formulas like (14.20) and (14.21).

Exercises

19.1 Express (18.20), written in the form $\mathbf{A}^*\mathbf{A} = \mathbf{I}$, in components.
19.2 Verify (19.14)
19.3 Establish the following properties of the trace of endomorphisms of $\mathscr{V}$:

$$\begin{aligned}
\operatorname{tr}\mathbf{I} &= \dim\mathscr{V}\\
\operatorname{tr}\left(\mathbf{A}+\mathbf{B}\right) &= \operatorname{tr}\mathbf{A} + \operatorname{tr}\mathbf{B}\\
\operatorname{tr}\left(\lambda\mathbf{A}\right) &= \lambda\operatorname{tr}\mathbf{A}\\
\operatorname{tr}\,\mathbf{A}\mathbf{B} &= \operatorname{tr}\mathbf{B}\mathbf{A}
\end{aligned}$$

and

$$\mathrm{tr}\mathbf{A}^* = \mathrm{tr}\mathbf{A}$$

19.4 If $\mathbf{A}, \mathbf{B} \in \mathscr{L}(\mathscr{V};\mathscr{V})$, where $\mathscr{V}$ is an inner product space, define

$$\mathbf{A} \cdot \mathbf{B} = \mathrm{tr}\,\mathbf{A}\mathbf{B}^*$$

Show that this definition makes $\mathscr{L}(\mathscr{V};\mathscr{V})$ into an inner product space.

19.5 In the special case when $\mathscr{V}$ is a real inner product space, show that the inner product of Exercise 19.4 implies that $\mathscr{A}(\mathscr{V};\mathscr{V}) \perp \mathscr{S}(\mathscr{V};\mathscr{V})$.

19.6 Verify formulas (19.16)-(19.18).

19.7 Use the results of Exercise 14.6 along with (19.14) to derive the transformation rules (19.16)-(19.18)

19.8 Given an endomorphism $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$, show that

$$\mathrm{tr}\,\mathbf{A} = \sum_{k=1}^{N} A_k{}^k$$

where $A_k{}^q = (\mathbf{A}\mathbf{e}^q) \cdot \mathbf{e}_k$. Thus, we can compute the trace of an endomorphism from (19.8) or from the formula above and be assured the same result is obtained in both cases. Of course, the formula above is also independent of the basis of $\mathscr{V}$.

19.9 Exercise 19.8 shows that $\mathrm{tr}\,\mathbf{A}$ can be computed from two of the four possible sets of components of $\mathbf{A}$. Show that the quantities $\sum_{k=1}^{N} A^{kk}$ and $\sum_{k=1}^{N} A_{kk}$ are not equal to each other in general and, moreover, do not equal $\mathrm{tr}\,\mathbf{A}$. In addition, show that each of these quantities depends upon the choice of basis of $\mathscr{V}$.

19.10 Show that

$$A^{*k\alpha} = (\mathbf{A}^*\mathbf{b}^\alpha) \cdot \mathbf{e}^k = \overline{A^{\alpha k}}$$

$$A^*{}_{k\alpha} = (\mathbf{A}^*\mathbf{b}_\alpha) \cdot \mathbf{e}_k = \overline{A_{\alpha k}}$$

and

$$A^{*k}{}_{\alpha} = \left(\mathbf{A}^{*}\mathbf{b}_{\alpha}\right)\cdot\mathbf{e}^{k} = \overline{A_{\alpha}{}^{k}}$$

Chapter 5

DETERMINANTS AND MATRICES

In Chapter 0 we introduced the concept of a matrix and examined certain manipulations one can carry out with matrices. In Section 7 we indicated that the set of 2 by 2 matrices with real numbers for its elements forms a ring with respect to the operations of matrix addition and matrix multiplication. In this chapter we shall consider further the concept of a matrix and its relation to a linear transformation.

Section 20. The Generalized Kronecker Deltas and the Summation Convention

Recall that a $M \times N$ matrix is an array written in anyone of the forms

$$A = \begin{bmatrix} A^1{}_1 & \cdot & \cdot & \cdot & A^1{}_N \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ A^M{}_1 & \cdot & \cdot & \cdot & A^M{}_N \end{bmatrix} = \left[A^\alpha{}_j \right], \qquad A = \begin{bmatrix} A_{11} & \cdot & \cdot & \cdot & A_{1N} \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ A_{M1} & \cdot & \cdot & \cdot & A_{MN} \end{bmatrix} = \left[A_{\alpha j} \right]$$

$$A = \begin{bmatrix} A_1{}^1 & \cdot & \cdot & \cdot & A_1{}^N \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ A_M{}^1 & \cdot & \cdot & \cdot & A_M{}^N \end{bmatrix} = \left[A_\alpha{}^j \right], \qquad A = \begin{bmatrix} A^{11} & \cdot & \cdot & \cdot & A^{1N} \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ A^{M1} & \cdot & \cdot & \cdot & A^{MN} \end{bmatrix} = \left[A^{\alpha j} \right] \tag{20.1}$$

Throughout this section the placement of the indices is of no consequence. The components of the matrix are allowed to be complex numbers. The set of $M \times N$ matrices shall be denoted by $\mathscr{M}^{M\times N}$. It is an elementary exercise to show that the rules of addition and scalar multiplication insure that $\mathscr{M}^{M\times N}$ is a vector space. The reader will recall that in Section 8 we used the set $\mathscr{M}^{M\times N}$ as one of several examples of a vector space. We leave as an exercise to the reader the fact that the dimension of $\mathscr{M}^{M\times N}$ is MN.

In order to give a definition of a determinant of a square matrix, we need to define a permutation and consider certain of its properties. Consider a set of K elements $\{\alpha_1,...,\alpha_K\}$. A *permutation* is a one-to-one function from $\{\alpha_1,...,\alpha_K\}$ to $\{\alpha_1,...,\alpha_K\}$. If σ is a permutation, it is customary to write

$$\sigma = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdot\ \cdot\ \cdot & \alpha_K \\ \sigma(\alpha_1) & \sigma(\alpha_2) & \cdot\ \cdot\ \cdot & \sigma(\alpha_K) \end{pmatrix} \tag{20.2}$$

It is a known result in algebra that permutations can be classified into even and odd ones: The permutation σ in (20.2) is *even* if an even number of pairwise interchanges of the bottom row is required to order the bottom row exactly like the top row, and σ is *odd* if that number is odd. For a given permutation σ, the number of pairwise interchanges required to order the bottom row the same as the top row is not unique, but it is proved in algebra that those numbers are either all even or all odd. Therefore the definition for σ to be even or odd is meaningful. For example, the permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

is odd, while the permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$$

is even.

The *parity* of σ, denoted by ε_σ, is defined by

$$\varepsilon_\sigma = \begin{cases} +1 \text{ if } \sigma \text{ is an even permutation} \\ \\ -1 \text{ if } \sigma \text{ is an odd permutation} \end{cases}$$

All of the applications of permutations we shall have are for permutations defined on K $(K \le N)$ positive integers selected from the set of N positive integers $\{1, 2, 3, ..., N\}$. Let $\{i_1, ..., i_K\}$ and $\{j_1, ..., j_K\}$ be two subsets of $\{1, 2, 3, ..., N\}$. If we order these two subsets and construct the two K-tuples $(i_1, ..., i_K)$ and $(j_1, ..., j_K)$, we can define the *generalized Kronecker delta* as follows: The generalized Kronecker delta, denoted by

$$\delta^{i_1 i_2 \cdots i_K}_{j_1 j_2 \cdots j_K}$$

is defined by

$$\delta^{i_1 i_2 \ldots i_K}_{j_1 j_2 \ldots j_K} = \begin{cases} 0 \text{ if the integers } (i_1,...,i_K) \text{ or } (j_1,...,j_K) \text{ are not distinct} \\ \\ 0 \text{ if the integers } (i_1,...,i_K) \text{ and } (j_1,...,j_K) \text{ are distinct} \\ \quad \text{but the sets } \{i_1,...,i_K\} \text{ and } \{j_1,...,j_K\} \text{ are not equal} \\ \\ \varepsilon_\sigma \text{ if the integers } (i_1,...,i_K) \text{ and } (j_1,...,j_K) \text{ are distinct} \\ \quad \text{and the sets } \{i_1,...,i_K\} \text{ and } \{j_1,...,j_K\} \text{ are equal, where} \\ \quad \sigma = \begin{pmatrix} i_1 & i_2 & \cdot & \cdot & \cdot & i_K \\ j_1 & j_2 & \cdot & \cdot & \cdot & j_K \end{pmatrix} \end{cases}$$

It follows from this definition that the generalized Kronecker delta is zero whenever the superscripts are not the same set of integers as the subscripts, or when the superscripts are not distinct, or when the subscripts are not distinct. Naturally, when $K = 1$ the generalized Kronecker delta reduces to the usual one. As an example, $\delta^{i_1 i_2}_{j_1 j_2}$ has the values

$$\delta^{12}_{12} = 1, \quad \delta^{12}_{21} = -1, \quad \delta^{13}_{12} = 0, \quad \delta^{13}_{21} = 0, \quad \delta^{11}_{12} = 1, \quad \text{etc.}$$

It can be shown that there are $N!K!/(N-K)!$ nonzero generalized Kronecker deltas for given positive integers N and N .

An ε symbol is one of a pair of quantities

$$\varepsilon^{i_1 i_2 \ldots i_N} = \delta^{i_1 i_2 \ldots i_N}_{12 \ldots N} \qquad \text{or} \qquad \varepsilon_{j_1 j_2 \ldots j_N} = \delta^{12 \ldots N}_{j_1 j_2 \ldots j_N} \tag{20.3}$$

For example, take $N = 3$; then

$$\begin{aligned} &\varepsilon^{123} = \varepsilon^{312} = \varepsilon^{231} = 1 \\ &\varepsilon^{132} = \varepsilon^{321} = \varepsilon^{213} = -1 \\ &\varepsilon^{112} = \varepsilon^{221} = \varepsilon^{222} = \varepsilon^{233} = 0, \quad \text{etc.} \end{aligned}$$

As an exercise, the reader is asked to confirm that

$$\varepsilon^{i_1 \ldots i_N} \varepsilon_{j_1 \ldots j_N} = \delta^{i_1 \ldots i_N}_{j_1 \ldots j_N} \tag{20.4}$$

An identity involving the quantity δ^{ijq}_{lms} is

$$\sum_{q=1}^{N} \delta^{ijq}_{lmq} = (N-1)\delta^{ij}_{lm} \tag{20.5}$$

To establish this identity, expand the left side of (20.5) in the form

$$\sum_{q=1}^{N} \delta_{lmq}^{ijq} = \delta_{lm1}^{ij1} + \cdots + \delta_{lmN}^{ijN} \tag{20.6}$$

Clearly we need to verify (20.5) for the cases $i \neq j$ and $\{i, j\} = \{l, m\}$ only, since in the remaining cases (20.5) reduces to the trivial equation $0 = 0$. In the nontrivial cases, exactly two terms on the right-hand side of (20.6) are equal to zero: one for $i = q$ and one for $j = q$. For i, j, l, and m not equal to q, δ_{lmq}^{ijq} has the same value as δ_{lm}^{ij}. Therefore,

$$\sum_{q=1}^{N} \delta_{lmq}^{ijq} = (N-1)\delta_{lm}^{ij}$$

which is the desired result (20.5). By the same procedure used above, it is clear that

$$\sum_{j=1}^{N} \delta_{lj}^{ij} = (N-1)\delta_{l}^{i} \tag{20.7}$$

Combining (20.5)and (20.7), we have

$$\sum_{j=1}^{N}\sum_{q=1}^{N} \delta_{ljq}^{ijq} = (N-2)(N-1)\delta_{j}^{i} \tag{20.8}$$

Since $\sum_{j=1}^{N} \delta_{i}^{i} = N$, we have from (20.8)

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{q=1}^{N} \delta_{ijq}^{ijq} = (N-2)(N-1)(N) = \frac{N!}{(N-3)!} \tag{20.9}$$

Equation (20.9) is a special case of

$$\sum_{i_1, i_2, \ldots, i_K = 1}^{N} \delta_{i_1 i_2 \ldots i_K}^{i_1 i_2 \ldots i_K} = \frac{N!}{(N-K)!} \tag{20.10}$$

Several other numerical relationships are

$$\sum_{i_{R+1}, i_{R+2}, \ldots, i_K = 1}^{N} \delta_{j_1 \ldots j_R i_{R+1} \ldots i_K}^{i_1 \ldots i_R i_{R+1} \ldots i_K} = \frac{(N-R)!}{(N-K)!} \delta_{j_1 \ldots j_R}^{i_1 \ldots i_R} \tag{20.11}$$

$$\sum_{i_{K+1}, \ldots, i_N = 1}^{N} \varepsilon^{i_1 \ldots i_K i_{K+1} \ldots i_N} \varepsilon_{j_1 \ldots j_K i_{K+1} \ldots i_N} = (N-K)! \delta_{j_1 \ldots j_K}^{i_1 \ldots i_K} \tag{20.12}$$

$$\sum_{i_{K+1},\ldots,i_N=1}^{N} \varepsilon^{i_1\ldots i_K i_{K+1}\ldots i_N}\delta^{j_{K+1}\ldots j_N}_{i_{K+1}\ldots i_N} = (N-K)!\,\varepsilon^{i_1\ldots i_K j_{K+1}\ldots j_N} \tag{20.13}$$

$$\sum_{j_{K+1},\ldots,j_R=1}^{N} \delta^{i_1\ldots i_K i_{K+1}\ldots i_R}_{j_1\ldots j_K j_{K+1}\ldots j_R}\delta^{j_{K+1}\ldots j_R}_{l_{K+1}\ldots l_R} = (R-K)!\,\delta^{i_1\ldots i_K i_{K+1}\ldots i_R}_{j_1\ldots j_K l_{K+1}\ldots l_R} \tag{20.14}$$

$$\sum_{j_{K+1},\ldots,j_R=1}^{N}\sum_{i_{K+1},\ldots,i_R=1}^{N} \delta^{i_1\ldots i_K i_{K+1}\ldots i_R}_{j_1\ldots j_K j_{K+1}\ldots j_R}\delta^{j_{K+1}\ldots j_R}_{i_{K+1}\ldots i_R} = \frac{(N-K)!}{(N-R)!}(R-K)!\,\delta^{i_1\ldots i_K}_{j_1\ldots j_K} \tag{20.15}$$

It is possible to simplify the formal appearance of many of our equations if we adopt a *summation convention*: We automatically sum every repeated index without writing the summation sign. For example, (20.5) is written

$$\delta^{ijq}_{lmq} = (N-1)\delta^{ij}_{lm} \tag{20.16}$$

The occurrence of the subscript q and the superscript q implies the summation indicated in (20.5). We shall try to arrange things so that we always sum a superscript on a subscript. It is important to know the range of a given summation, so we shall use the summation convention only when the range of summation is understood. Also, observe that the repeated indices are *dummy* indices in the sense that it is unimportant which symbol is used for them. For example,

$$\delta^{ijs}_{lms} = \delta^{ijt}_{lmt} = \delta^{ijq}_{lmq}$$

Naturally there is no meaning to the occurrence of the same index more than twice in a given term. Other than the summation or dummy indices, many equations have *free* indices whose values in the given range $\{1,\ldots,N\}$ are arbitrary. For example, the indices j,k,l and m are free indices in (20.16). Notice that every term in a given equation must have the same free indices; otherwise, the equation is meaningless.

The summation convention will be adopted in the remaining portion of this text. If we feel it might be confusing in some context, summations will be indicated by the usual summation sign.

Exercises

20.1 Verify equation (20.4).

20.2 Prove that $\dim \mathscr{M}^{M\times N} = MN$.

Section 21. Determinants

In this section we shall use the generalized Kronecker deltas and the ε symbols to define the determinant of a square matrix.

The *determinant* of the $N \times N$ matrix $A = \left[A_{ij} \right]$, written $\det A$, is a complex number defined by

$$\det A = \begin{vmatrix} A_{11} & A_{12} & \cdot & \cdot & \cdot & A_{1N} \\ A_{21} & A_{22} & \cdot & \cdot & \cdot & A_{2N} \\ A_{31} & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ A_{N1} & A_{N2} & \cdot & \cdot & \cdot & A_{NN} \end{vmatrix} = \varepsilon^{i_1 \cdots i_N} A_{i_1 1} A_{i_2 2} \cdots A_{i_N N} \tag{21.1}$$

where all summations are from 1 to *N*. If the elements of the matrix are written $\left[A^{ij} \right]$, its determinant is defined to be

$$\det A = \begin{vmatrix} A^{11} & A^{12} & \cdot & \cdot & \cdot & A^{1N} \\ A^{21} & A^{22} & \cdot & \cdot & \cdot & A^{2N} \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ A^{N1} & A^{N2} & \cdot & \cdot & \cdot & A^{NN} \end{vmatrix} = \varepsilon_{i_1 \cdots i_N} A^{i_1 1} A^{i_2 2} \cdots A^{i_N N} \tag{21.2}$$

Likewise, if the elements of the matrix are written $\left[A^i{}_j \right]$ its determinant is defined by

$$\det A = \begin{vmatrix} A^1{}_1 & A^1{}_2 & \cdot & \cdot & \cdot & A^1{}_N \\ A^2{}_1 & A^2{}_2 & \cdot & \cdot & \cdot & A^2{}_N \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ A^N{}_1 & A^N{}_2 & \cdot & \cdot & \cdot & A^N{}_N \end{vmatrix} = \varepsilon_{i_1 \cdots i_N} A^{i_1}{}_1 A^{i_2}{}_2 \cdots A^{i_N}{}_N \tag{21.3}$$

A similar formula holds when the matrix is written $A = \left[A_i{}^j \right]$. The generalized Kronecker delta can be written as the determinant of ordinary Kronecker deltas. For example, one can show, using (21.3), that

$$\delta^{ij}_{kl} = \begin{vmatrix} \delta^i_k & \delta^i_l \\ \delta^j_k & \delta^j_l \end{vmatrix} = \delta^i_k \delta^j_l - \delta^i_l \delta^j_k \tag{21.4}$$

Equation (21.4) is a special case of the general result

$$\delta^{i_1 \ldots i_K}_{j_1 \ldots j_K} = \begin{vmatrix} \delta^{i_1}_{j_1} & \delta^{i_1}_{j_2} & \cdot & \cdot & \cdot & \delta^{i_1}_{j_K} \\ \delta^{i_2}_{j_1} & \delta^{i_2}_{j_2} & \cdot & \cdot & \cdot & \delta^{i_2}_{j_K} \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ \delta^{i_K}_{j_1} & \cdot & \cdot & \cdot & \cdot & \delta^{i_K}_{j_K} \end{vmatrix} \tag{21.5}$$

It is possible to use (21.1)-(21.3) to show that

$$\varepsilon_{j_1 \ldots j_N} \det A = \varepsilon^{i_1 \ldots i_N} A_{i_1 j_1} \cdots A_{i_N j_N} \tag{21.6}$$

$$\varepsilon^{j_1 \ldots j_N} \det A = \varepsilon_{i_1 \ldots i_N} A^{i_1 j_1} \cdots A^{i_N j_N} \tag{21.7}$$

and

$$\varepsilon_{j_1 \ldots j_N} \det A = \varepsilon_{i_1 \ldots i_N} A^{i_1}{}_{j_1} \cdots A^{i_N}{}_{j_N} \tag{21.8}$$

It is possible to use (21.6)-(21.8) and (20.10) with $N = K$ to show that

$$\det A = \frac{1}{N!} \varepsilon^{j_1 \ldots j_N} \varepsilon^{i_1 \ldots i_N} A_{i_1 j_1} \cdots A_{i_N j_N} \tag{21.9}$$

$$\det A = \frac{1}{N!} \varepsilon_{j_1 \ldots j_N} \varepsilon_{i_1 \ldots i_N} A^{i_1 j_1} \cdots A^{i_N j_N} \tag{21.10}$$

and

$$\det A = \frac{1}{N!} \delta^{j_1 \ldots j_N}_{i_1 \ldots i_N} A^{i_1}{}_{j_1} \cdots A^{i_N}{}_{j_N} \tag{21.11}$$

Equations (21.9), (21.10) and (21.11) confirm the well-known result that a matrix and its transpose have the same determinant. The reader is cautioned that the transpose mentioned here is that of the

matrix and not of a linear transformation. By use of this fact it follows that (21.1)-(21.3) could have been written

$$\det A = \varepsilon^{i_1 \cdots i_N} A_{1 i_1} \cdots A_{N i_N} \tag{21.12}$$

$$\det A = \varepsilon_{i_1 \cdots i_N} A^{1 i_1} \cdots A^{N i_N} \tag{21.13}$$

and

$$\det A = \varepsilon^{i_1 \cdots i_N} A^1{}_{i_1} \cdots A^N{}_{i_N} \tag{21.14}$$

A similar logic also yields

$$\varepsilon_{j_1 \cdots j_N} \det A = \varepsilon^{i_1 \cdots i_N} A_{j_1 i_1} \cdots A_{j_N i_N} \tag{21.15}$$

$$\varepsilon^{j_1 \cdots j_N} \det A = \varepsilon_{i_1 \cdots i_N} A^{j_1 i_1} \cdots A^{j_N i_N} \tag{21.16}$$

and

$$\varepsilon^{j_1 \cdots j_N} \det A = \varepsilon^{i_1 \cdots i_N} A^{j_1}{}_{i_1} \cdots A^{j_N}{}_{i_N} \tag{21.17}$$

The *cofactor* of the element $A^s{}_t$ in the matrix $A = \left[A^i{}_j \right]$ is defined by

$$\operatorname{cof} A^s{}_t = \varepsilon_{i_1 i_2 \cdots i_N} A^{i_1}{}_1 \cdots A^{i_{t-1}}{}_{t-1} \delta^{i_t}_s A^{i_{t+1}}{}_{t+1} \cdots A^{i_N}{}_N \tag{21.18}$$

As an illustration of the application of (21.18), let $N = 3$ and $s = t = 1$. Then

$$\begin{aligned} \operatorname{cof} A^1{}_1 &= \varepsilon_{ijk} \delta^i_1 A^j{}_2 A^k{}_3 \\ &= \varepsilon_{1jk} A^j{}_2 A^k{}_3 \\ &= A^2{}_2 A^3{}_3 - A^3{}_2 A^2{}_3 \end{aligned}$$

Theorem 21.1.

$$\sum_{s=1}^{N} A^s{}_q \operatorname{cof} A^s{}_t = \delta^t_q \det A \qquad \text{and} \quad \sum_{t=1}^{N} A^q{}_t \operatorname{cof} A^s{}_t = \delta^q_s \det A \tag{21.19}$$

Proof. It follows from (21.18) that

$$\begin{aligned}\sum_{s=1}^{N} A^s{}_q \operatorname{cof} A^s{}_t &= \varepsilon_{i_1 i_2 \ldots i_N} A^{i_1}{}_1 \cdots A^{i_{t-1}}{}_{t-1} A^s{}_q \delta^{i_t}_s A^{i_{t+1}}{}_{t+1} \cdots A^{i_N}{}_N \\ &= \varepsilon_{i_1 i_2 \ldots i_N} A^{i_1}{}_1 \cdots A^{i_{t-1}}{}_{t-1} A^{i_t}{}_q A^{i_{t+1}}{}_{t+1} \cdots A^{i_N}{}_N \\ &= \varepsilon_{i_1 i_2 \ldots i_N} A^{i_1}{}_1 \cdots A^{i_{t-1}}{}_{t-1} A^{i_t}{}_t A^{i_{t+1}}{}_{t+1} \cdots A^{i_N}{}_N \delta^t_q \\ &= \delta^t_q \det A\end{aligned}$$

where the definition (21.3) has been used. Equation $(21.19)_2$ follows by a similar argument.

Equations (21.19) represent the classical Laplace expansion of a determinant. In Chapter 0 we promised to present a proof for (0.29) for matrices of arbitrary order. Equations (21.19) fulfill this promise.

Exercises

21.1 Verify equation (21.4).
21.2 Verify equations (21.6)-(21.8).
21.3 Verify equationa (21.9)-(21.11).
21.4 Show that the cofactor of an element of an $N \times N$ matrix A can be written

$$\operatorname{cof} A^{i_1}{}_{j_1} = \frac{1}{(N-1)!} \delta^{j_1 \ldots j_N}_{i_1 \ldots i_N} A^{i_2}{}_{j_2} \cdots A^{i_N}{}_{j_N} \tag{21.20}$$

21.5 If A and B are $N \times N$ matrices use (21.8) to prove that

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

21.6 If I is the $N \times N$ identity matrix show that $\det I = 1$.
21.7 If A is a nonsingular matrix show, that $\det A \neq 0$ and that

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

21.8 If A is an $N \times N$ matrix, we define its $K \times K$ *minor* $(1 \leq K \leq N)$ to be the determinant of any $K \times K$ submatrix of A, e.g.,

$$A^{i_1\cdots i_K}_{j_1\cdots j_K} \equiv \det\begin{bmatrix} A^{i_1}{}_{j_1} & \cdot & \cdot & \cdot & A^{i_1}{}_{j_K} \\ \cdot & & & & \\ \cdot & & & & \\ \cdot & & & & \\ A^{i_K}{}_{j_1} & \cdot & \cdot & \cdot & A^{i_K}{}_{j_K} \end{bmatrix} = \delta^{k_1\cdots k_K}_{j_1\cdots j_K} A^{i_1}{}_{k_1}\cdots A^{i_K}{}_{k_K}$$

$$= \delta^{i_1\cdots i_K}_{k_1\cdots k_K} A^{k_1}{}_{j_1}\cdots A^{k_K}{}_{j_K} \tag{21.21}$$

In particular, any element $A^i{}_j$ of A is a 1×1 minor of A. The cofactor of the $K\times K$ minor $A^{i_1\cdots i_K}_{j_1\cdots j_K}$ is defined by

$$\text{cof } A^{i_1\cdots i_K}_{j_1\cdots j_K} \equiv \frac{1}{(N-K)!}\delta^{j_1\cdots j_N}_{i_1\cdots i_N} A^{i_{K+1}}{}_{j_{K+1}}\cdots A^{i_N}{}_{j_N} \tag{21.22}$$

which is equal to the $(N-K)\times(N-K)$ minor complementary to the $K\times K$ minor $A^{i_1\cdots i_K}_{j_1\cdots j_K}$ and is assigned an appropriate sign. Specifically, if $(i_{K+1},...,i_N)$ and $(j_{K+1},...,j_N)$ are complementary sets of $(i_1,...,i_K)$ and $(j_1,...,j_K)$ in $(1,...,N)$, then

$$\text{cof } A^{i_1\cdots i_K}_{j_1\cdots j_K} = \delta^{i_1\cdots i_N}_{j_1\cdots j_N} A^{i_{K+1}\cdots i_N}_{j_{K+1}\cdots j_N} \qquad \text{(no summation)} \tag{21.23}$$

Clearly (21.22) generalizes (21.20). Show that the formulas that generalize (21.19) to $K\times K$ minors in general are

$$\delta^{i_1\cdots i_K}_{j_1\cdots j_K}\det A = \frac{1}{K!}\sum_{k_1,...,k_K=1}^{N} A^{i_1\cdots i_K}_{k_1\cdots k_K}\,\text{cof } A^{j_1\cdots j_K}_{k_1\cdots k_K}$$

$$\delta^{j_1\cdots j_K}_{i_1\cdots i_K}\det A = \frac{1}{K!}\sum_{k_1,...,k_K=1}^{N} A^{k_1\cdots k_K}_{i_1\cdots i_K}\,\text{cof } A^{k_1\cdots k_K}_{j_1\cdots j_K} \tag{21.24}$$

21.9 If A is nonsingular, show that

$$A^{i_1\cdots i_K}_{j_1\cdots j_K} = (\det A)\,\text{cof}(A^{-1})^{j_1\cdots j_K}_{i_1\cdots i_K} \tag{21.25}$$

and that

$$\frac{1}{K!}\sum_{k_1,...,k_K=1}^{N} A^{i_1\cdots i_K}_{k_1\cdots k_K}(A^{-1})^{k_1\cdots k_K}_{j_1\cdots j_K} = \delta^{i_1\cdots i_K}_{j_1\cdots j_K} \tag{21.26}$$

In particular, if $K = 1$ and A is replaced by A^{-1} then (21.25) reduces to

$$(A^{-1})^i{}_j = (\det A)^{-1} \operatorname{cof} A^j{}_i \tag{21.27}$$

which is equation (0.30).

Section 22. The Matrix of a Linear Transformation

In this section we shall introduce the matrix of a linear transformation with respect to a basis and investigate certain of its properties. The formulas of Chapter 0 and Section 21 are purely numerical formulas independent of abstract vectors or linear transformations. Here we show how a certain matrix can be associated with a linear transformation, and, more importantly, we show to what extent the relationship is basis dependent.

If $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{U})$, $\{\mathbf{e}_1,...,\mathbf{e}_N\}$ is a basis for $\mathscr{V}$, and $\{\mathbf{b}_1,...,\mathbf{b}_M\}$ is a basis for $\mathscr{U}$, then we can characterize $\mathbf{A}$ by the system (19.1):

$$\mathbf{A}\mathbf{e}_k = A^{\alpha}{}_k \mathbf{b}_\alpha \tag{22.1}$$

where the summation is in force with the Greek indices ranging from 1 to *M*. The *matrix* of $\mathbf{A}$ with respect to the bases $\{\mathbf{e}_1,...,\mathbf{e}_N\}$ and $\{\mathbf{b}_1,...,\mathbf{b}_M\}$, denoted by $M(\mathbf{A},\mathbf{e}_k,\mathbf{b}_\alpha)$ is

$$M(\mathbf{A},\mathbf{e}_k,\mathbf{b}_\alpha) = \begin{bmatrix} A^1{}_1 & A^1{}_2 & \cdot & \cdot & \cdot & A^1{}_N \\ A^2{}_1 & A^2{}_2 & \cdot & \cdot & \cdot & A^2{}_N \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ A^M{}_1 & A^M{}_2 & \cdot & \cdot & \cdot & A^M{}_N \end{bmatrix} \tag{22.2}$$

As the above argument indicates, the matrix of $\mathbf{A}$ depends upon the choice of basis for $\mathscr{V}$ and $\mathscr{U}$. However, unless this point needs to be stressed, we shall often write $M(\mathbf{A})$ for the matrix of $\mathbf{A}$ and the basis dependence is understood. We can always regard M as a function $M : \mathscr{L}(\mathscr{U};\mathscr{V}) \to \mathscr{M}^{M\times N}$. It is a simple exercise to confirm that

$$M(\lambda\mathbf{A} + \mu\mathbf{B}) = \lambda M(\mathbf{A}) + \mu M(\mathbf{B}) \tag{22.3}$$

for all $\lambda, \mu \in \mathscr{C}$ and $\mathbf{A},\mathbf{B} \in \mathscr{L}(\mathscr{U};\mathscr{V})$. Thus M is a linear transformation. Since

$$M(\mathbf{A}) = 0$$

implies $\mathbf{A} = \mathbf{0}$, M is one-to-one. Since $\mathscr{L}(\mathscr{U};\mathscr{V})$ and $\mathscr{M}^{M\times N}$ have the same dimension (see Theorem 16.1: and Exercise 20.2), Theorem 15.10 tells us that M is an automorphism and, thus, $\mathscr{L}(\mathscr{U};\mathscr{V})$ and $\mathscr{M}^{M\times N}$ are isomorphic. The dependence of $M(\mathbf{A},\mathbf{e}_k,\mathbf{b}_\alpha)$ on the bases can be exhibited by use of the transformation rule (19.5). In matrix form (19.5) is

$$M(\mathbf{A},\mathbf{e}_k,\mathbf{b}_\alpha) = SM(\mathbf{A},\hat{\mathbf{e}}_k,\hat{\mathbf{b}}_\alpha)T^{-1} \tag{22.4}$$

where S is the $M \times M$ matrix

$$S = \begin{bmatrix} S_1^1 & S_2^1 & \cdot & \cdot & \cdot & S_M^1 \\ S_1^2 & S_2^2 & \cdot & \cdot & \cdot & S_M^2 \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ S_1^M & S_2^M & \cdot & \cdot & \cdot & S_M^M \end{bmatrix} \tag{22.5}$$

and T is the $N \times N$ matrix

$$T = \begin{bmatrix} T_1^1 & T_2^1 & \cdot & \cdot & \cdot & T_N^1 \\ T_1^2 & T_2^2 & \cdot & \cdot & \cdot & T_N^2 \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ T_1^N & T_2^N & \cdot & \cdot & \cdot & T_N^N \end{bmatrix} \tag{22.6}$$

Of course, in constructing (22.4) from (19.5) we have used the fact expressed by (14.20) and (14.21) that the matrix T^{-1} has components $\hat{T}_k^j$, $j,k = 1,...N$.

If $\mathbf{A}$ is an endomorphism, the transformation formula (22.4) becomes

$$M(\mathbf{A}, \mathbf{e}_k, \mathbf{e}_q) = TM(\mathbf{A}, \hat{\mathbf{e}}_k, \hat{\mathbf{e}}_q)T^{-1} \tag{22.7}$$

We shall use (22.7) to motivate the concept of the *determinant* of an endomorphism. The determinant of $M(\mathbf{A}, \mathbf{e}_k, \mathbf{e}_q)$, written $\det M(\mathbf{A}, \mathbf{e}_k, \mathbf{e}_q)$, can be computed by (21.3). It follows from (22.7) and Exercises 21.5 and 21.7 that

$$\begin{aligned} \det M(\mathbf{A}, \mathbf{e}_k, \mathbf{e}_q) &= (\det T)(\det M(\mathbf{A}, \hat{\mathbf{e}}_k, \hat{\mathbf{e}}_q))(\det T^{-1}) \\ &= \det M(\mathbf{A}, \hat{\mathbf{e}}_k, \hat{\mathbf{e}}_q) \end{aligned}$$

Thus, we obtain the important result that $\det M(\mathbf{A}, \mathbf{e}_k, \mathbf{e}_q)$ is *independent* of the choice of basis for $\mathscr{V}$. With this fact we define the determinant of an endomorphism $\mathbf{A} \in \mathscr{L}(\mathscr{V}; \mathscr{V})$, written $\det \mathbf{A}$, by

$$\det \mathbf{A} = \det M(\mathbf{A}, \mathbf{e}_k, \mathbf{e}_q) \tag{22.8}$$

By the above argument, we are assured that $\det \mathbf{A}$ is a property of $\mathbf{A}$ alone. In Chapter 8, we shall introduce directly a definition for the determinant of an endomorphism without use of a basis.

Given a linear transformation $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{U})$, the adjoint $\mathbf{A}^* \in \mathscr{L}(\mathscr{U};\mathscr{V})^*$ is defined by the component formula (19.9) Consistent with equations (22.1) and (22.2), equation (19.9) implies that

$$M(\mathbf{A}^*, \mathbf{b}^\alpha, \mathbf{e}^k) = \begin{bmatrix} A^{*1}_{\ 1} & A^{*2}_{\ 1} & \cdot & \cdot & \cdot & A^{*M}_{\ 1} \\ A^{*1}_{\ 2} & A^{*2}_{\ 2} & \cdot & \cdot & \cdot & A^{*M}_{\ 2} \\ \cdot & & & & & \\ \cdot & & & & & \\ \cdot & & & & & \\ A^{*1}_{\ N} & A^{*2}_{\ N} & \cdot & \cdot & \cdot & A^{*M}_{\ N} \end{bmatrix} \tag{22.9}$$

Notice that the matrix of $\mathbf{A}^*$ is referred to the reciprocal bases $\{\mathbf{e}^k\}$ and $\{\mathbf{b}^\alpha\}$. If we now use (19.10) and the definition (22.2) we see that

$$M(\mathbf{A}^*, \mathbf{b}^\alpha, \mathbf{e}^k) = \overline{M(\mathbf{A}, \mathbf{e}_k, \mathbf{b}_\alpha)^T} \tag{22.10}$$

where the complex conjugate of a matrix is the matrix formed by taking the complex conjugate of each component of the given matrix. Note that in (22.10) we have used the definition of the transpose of a matrix in Chapter 0. Equation (22.10) gives a simple comparison of the component matrices of a linear transformation and its adjoint. If the vector spaces are real, (22.10) reduces to

$$M(\mathbf{A}^T) = M(\mathbf{A})^T \tag{22.11}$$

where the basis dependence is understood. For an endomorphism $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$, we can use (22.10) and (22.8) to show that

$$\det \mathbf{A}^* = \overline{\det \mathbf{A}} \tag{22.12}$$

Given a $M \times N$ matrix $A = \left[A^\alpha_{\ k} \right]$, there correspond M $1 \times N$ row matrices and N $M \times 1$ column matrices. The *row rank* of the matrix A is equal to the number of linearly independent row matrices and the *column rank* of A is equal to the number of linearly independent column matrices. It is a property of matrices, which we shall prove, that the row rank *equals* the column rank. This common rank in turn is equal to the rank of the linear transformation whose matrix is A. The theorem we shall prove is the following.

Theorem 22.1. $A = \left[A^\alpha_{\ k} \right]$ is an $M \times N$ matrix, the row rank of A equals the column rank of A.

Proof. Let $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{U})$ and let $M(\mathbf{A}) = A = \left[A^{\alpha}{}_{k} \right]$ with respect to bases $\{\mathbf{e}_1,...,\mathbf{e}_N\}$ for $\mathscr{V}$ and $\{\mathbf{b}_1,...,\mathbf{b}_M\}$ for $\mathscr{U}$. We can define a linear transformation $\mathbf{B}: \mathscr{U} \to \mathscr{C}^M$ by

$$\mathbf{Bu} = (u^1, u^2, ..., u^M)$$

where $\mathbf{u} = u^{\alpha}\mathbf{b}_{\alpha}$. Observe that $\mathbf{B}$ is an isomorphism. The product $\mathbf{BA}$ is a linear transformation $\mathscr{V} \to \mathscr{C}^M$; further,

$$\mathbf{BAe}_k = \mathbf{B}(A^{\alpha}{}_{k}\mathbf{b}_{\alpha}) = A^{\alpha}{}_{k}\mathbf{Bb}_{\alpha} = \left(A^1{}_{k}, A^2{}_{k}, ..., A^M{}_{k} \right)$$

Therefore $\mathbf{BAe}_k$ is an M-tuple whose elements are those of the kth column matrix of A. This means $\dim R(\mathbf{BA}) =$ column rank of A. Since $\mathbf{B}$ is an isomorphism, $\mathbf{BA}$ and $\mathbf{A}$ have the same rank and thus column rank of $A = \dim R(\mathbf{A})$. A similar argument applied to the adjoint mapping $\mathbf{A}^*$ shows that row rank of $A = \dim R(\mathbf{A}^*)$. If we now apply Theorem 18.4 we find the desired result.

In what follows we shall only refer to the rank of a matrix rather than to the row or the column rank.

Theorem 22.2. An endomorphism $\mathbf{A}$ is regular if and only if $\det \mathbf{A} \neq 0$.

Proof. ~ If $\det \mathbf{A} \neq 0$, equations (21.19) or, equivalently, equation (0.30), provide us with a formula for the direct calculation of $\mathbf{A}^{-1}$ so $\mathbf{A}$ is regular. If $\mathbf{A}$ is regular, then $\mathbf{A}^{-1}$ exists, and the results of Exercises 22.3 and 21.7 show that $\det \mathbf{A} \neq 0$.

Before closing this section, we mention here that among the four component matrices $\left[A^{\alpha}{}_{k} \right]$, $[A_{\alpha k}]$, $\left[A^{\alpha k} \right]$, and $\left[A_{\alpha}{}^{k} \right]$ defined by (19.1), (19.16), (19.17), and (19.18), respectively, the first one, $\left[A_{\alpha}{}^{k} \right]$, is most frequently used. For example, it is that particular component matrix of an endomorphism $\mathbf{A}$ that we used to define the determinant of $\mathbf{A}$, as shown in (22.8). In the sequel, if we refer to the component matrix of a transformation or an endomorphism, then unless otherwise specified, we mean the component matrix of the first kind.

Exercises

22.1 If $\mathbf{A}$ and $\mathbf{B}$ are endomorphisms, show that

$$M(\mathbf{AB}) = M(\mathbf{A})M(\mathbf{B}) \qquad \text{and} \qquad \det \mathbf{AB} = \det \mathbf{A} \det \mathbf{B}$$

22.2 If $\mathbf{A}: \mathscr{V} \to \mathscr{V}$ is a skew-Hermitian endomorphism, show that

$$\det \mathbf{A} = (-1)^{\dim \mathscr{V}} \overline{\det \mathbf{A}}$$

22.3 If $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ is regular, show that

$$M(\mathbf{A}^{-1}) = M(\mathbf{A})^{-1}$$

22.4 If $\mathbf{P} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ is a projection, show that there exists a basis for $\mathscr{V}$ such that

$$M(\mathbf{P}) = \begin{bmatrix} 1 & & & & & & & \\ & 1 & & & & & & \\ & & \ddots & & & & & \\ & & & 1 & & & & \\ & & & & 0 & & & \\ & & & & & 0 & & \\ & & & & & & \ddots & \\ & & & & & & & 0 \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}1\\1\\\ddots\\1\end{matrix}}\right\} R = \dim R(\mathbf{P}) \\ \left.\vphantom{\begin{matrix}0\\0\\\ddots\\0\end{matrix}}\right\} N - R \end{matrix}$$

22.5 If $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ is Hermitian, show that the component version of this fact is

$$A_{ij} = \overline{A_{ji}}, \qquad A^i{}_j = \overline{A_j{}^i}, \qquad \text{and} \qquad e_{ik}A^i{}_j = \overline{e_{ji}A_k{}^i}$$

where $e_{ik} = \mathbf{e}_i \cdot \mathbf{e}_k$, and where $\{\mathbf{e}_1,...,\mathbf{e}_N\}$ is the basis used to express the transformation formulas (19.16)-(19,18) in matrix notation.

22.6 If $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ we have seen it has four sets of components which are denoted by $A_i{}^j$, $A^i{}_j$, A^{ij}, and A_{ij}. Prove that $\det\left[A_{ij}\right]$ and $\det\left[A^{ij}\right]$ *do* depend upon the choice of basis for $\mathscr{V}$.

22.7 Show that

$$\det(\alpha\mathbf{A}) = \alpha^N \det \mathbf{A} \tag{22.13}$$

for all scalars α. In particular, if $\mathbf{A} = \mathbf{I}$, then from Exercise 21.6

$$\det(\alpha\mathbf{I}) = \alpha^N \tag{22.14}$$

22.8 If $\mathbf{A}$ is an endomorphism of an N-dimensional vector space $\mathscr{V}$, show that $\det(\mathbf{A} - \lambda\mathbf{I})$ is a polynomial of degree N in λ.

Section 23. Solution of Systems of Linear Equations

In this section we shall examine the problem system of M equations in N unknowns of the form

$$A^{\alpha}{}_{k}v^{k} = u^{\alpha}, \qquad \alpha = 1,...,M, \quad k = 1,...,N \tag{23.1}$$

where the MN coefficients $A^{\alpha}{}_{k}$ and the M data u^{α} are given. If we introduce bases bases $\{\mathbf{e}_1,...,\mathbf{e}_N\}$ for a vector space $\mathscr{V}$ and $\{\mathbf{b}_1,...,\mathbf{b}_M\}$ for a vector space $\mathscr{U}$, then (23.1) can be viewed as the component formula of a certain vector equation

$$\mathbf{Av} = \mathbf{u} \tag{23.2}$$

which immediately yields the following theorem.

Theorem 23.1. (23.1) has a solution if and only if $\mathbf{u}$ is in $R(\mathbf{A})$.

Another immediate implication of (23.2) is the following theorem.

Theorem 23.2. If (23.1) has a solution, the solution is unique if and only if $\mathbf{A}$ is regular.

Given the system system of equations (23.1), the *associated homogeneous system* is the set of equations

$$A^{\alpha}{}_{k}v^{k} = 0 \tag{23.3}$$

Theorem 23.3. The set of solutions of the homogeneous system (23.3) whose coefficient matrix is of rank R form a vector space of dimension $N - R$.

Proof. Equation (23.3) can be regarded as the component formula of the vector equation

$$\mathbf{Av = 0}$$

which implies that v^k solves (23.3) if and only if $\mathbf{v} \in K(\mathbf{A})$. By (15.6) $\dim K(\mathbf{A}) = \operatorname{div}\mathscr{V} - \dim R(\mathbf{A}) = N - R$.

From Theorems 15.7 and 23.3 we see that if there are fewer equations than unknowns (i.e. $M < N$), the system (23.3) always has a nonzero solution. This assertion is clear because $N - R \geq N - M > 0$, since $R(\mathbf{A})$ is a subspace of $\mathscr{U}$ and $M = \dim\mathscr{U}$.

If in (23.1) and (23.3) $M = N$, the system (23.1) has a solution for all u^k if and only if (23.3) has the trivial solution $v^1 = v^2 = \cdots = v^N = 0$ only. For, in this circumstance, $\mathbf{A}$ is regular

and thus invertible. This means $\det\left[A^{\alpha}{}_{k} \right] \neq 0$, and we can use (21.19) and write the solution of (23.1) in the form

$$v^{j} = \frac{1}{\det\left[A^{\alpha}{}_{k} \right]} \sum_{\alpha=1}^{N} (\text{cof } A^{\alpha}{}_{j}) u^{\alpha} \tag{23.4}$$

which is the classical Cramer's rule and is to *N* dimension of equation (0.32).

Exercise

23.1 Given the system of equations (23.1), the *augmented matrix* of the system is the matrix obtained from $\left[A^{\alpha}{}_{k} \right]$ by the addition of a co1umn formed from $u^{1},...,u^{N}$. Use Theorem 23.1 and prove that the system (23.1) has a solution if and only if the rank of $\left[A^{\alpha}{}_{k} \right]$ equals the rank of the augmented matrix.

Chapter 6

SPECTRAL DECOMPOSITIONS

In this chapter we consider one of the more advanced topics in the study of linear transformations. Essentially we shall consider the problem of analyzing an endomorphism by decomposing it into elementary parts.

Section 24. Direct Sum of Endomorphisms

If $\mathbf{A}$ is an endomorphism of a vector space $\mathscr{V}$, a subspace $\mathscr{V}_1$ of $\mathscr{V}$ is said to be $\mathbf{A}$-invariant if $\mathbf{A}$ maps $\mathscr{V}_1$ to $\mathscr{V}_1$. The most obvious example of an $\mathbf{A}$-invariant subspace is the null space $K(\mathbf{A})$. Let $\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_L$ be endomorphisms of $\mathscr{V}$; then an endomorphism $\mathbf{A}$ is the *direct sum* of $\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_L$ if

$$\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2 + \cdots + \mathbf{A}_L \tag{24.1}$$

and

$$\mathbf{A}_i \mathbf{A}_j = \mathbf{0} \qquad i \neq j \tag{24.2}$$

For example, equation (17.18) presents a direct sum decomposition for the identity linear transformation.

Theorem 24.1. If $\mathbf{A} \in \mathscr{L}(\mathscr{V}; \mathscr{U})$ and $\mathscr{V} = \mathscr{V}_1 \oplus \mathscr{V}_2 \oplus \cdots \oplus \mathscr{V}_L$, where each subspace $\mathscr{V}_j$ is $\mathbf{A}$-invariant, then $\mathbf{A}$ has the direct sum decomposition (24.1), where each $\mathbf{A}_i$ is given by

$$\mathbf{A}_i \mathbf{v}_i = \mathbf{A}\mathbf{v}_i \tag{24.3}_1$$

for all $\mathbf{v}_i \in \mathscr{V}_i$ and

$$\mathbf{A}_i \mathbf{v}_j = \mathbf{0} \tag{24.3}_2$$

For all $\mathbf{v}_j \in \mathscr{V}_j$, $i \neq j$, for $i = 1,...,L$. Thus the restriction of $\mathbf{A}$ to $\mathscr{V}_j$ coincides with that of $\mathbf{A}_j$; further, each $\mathscr{V}_i$ is $\mathbf{A}_j$-invariant, for all $j = 1,...,L$.

Proof. Given the decomposition $\mathscr{V} = \mathscr{V}_1 \oplus \mathscr{V}_2 \oplus \cdots \oplus \mathscr{V}_L$, then $\mathbf{v} \in \mathscr{V}$ has the unique representation $\mathbf{v} = \mathbf{v}_1 + \cdots + \mathbf{v}_L$, where each $\mathbf{v}_i \in \mathscr{V}_i$. By the converse of Theorem 17.4, there exist L projections $\mathbf{P}_1,...,\mathbf{P}_L$ (19.18) and also $\mathbf{v}_j \in \mathbf{P}_j(\mathscr{V})$. From (24.3), we have $\mathbf{A}_j = \mathbf{A}\mathbf{P}_j$, and since $\mathscr{V}_j$ is $\mathbf{A}$-invariant, $\mathbf{A}\mathbf{P}_j = \mathbf{P}_j\mathbf{A}$. Therefore, if $i \neq j$,

$$\mathbf{A}_i\mathbf{A}_j = \mathbf{A}\mathbf{P}_i\mathbf{A}\mathbf{P}_j = \mathbf{A}\mathbf{A}\mathbf{P}_i\mathbf{P}_j = \mathbf{0}$$

where $(17.17)_2$ has .been used. Also

$$\begin{aligned}\mathbf{A}_1 + \mathbf{A}_2 + \cdots + \mathbf{A}_L &= \mathbf{A}\mathbf{P}_1 + \mathbf{A}\mathbf{P}_2 + \cdots + \mathbf{A}\mathbf{P}_L \\ &= \mathbf{A}(\mathbf{P}_1 + \mathbf{P}_2 + \cdots + \mathbf{P}_L) \\ &= \mathbf{A}\end{aligned}$$

where (17.18) has been used.

When the assumptions of the preceding theorem are satisfied, the endomorphism $\mathbf{A}$ is said to be *reduced* by the subspaces $\mathscr{V}_1,...,\mathscr{V}_L$. An important result of this circumstance is contained in the following theorem.

Theorem 24.2. Under the conditions of Theorem 24.1, the determinant of the endomorphism $\mathbf{A}$ is given by

$$\det \mathbf{A} = \det \mathbf{A}_1 \det \mathbf{A}_2 \cdots \det \mathbf{A}_L \tag{24.4}$$

where $\mathbf{A}_k$ denotes the restriction of $\mathbf{A}$ to $\mathscr{V}_k$ for all $k = 1,...,L$.

The proof of this theorem is left as an exercise to the reader.

Exercises

24.1 Assume that $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ is reduced by $\mathscr{V}_1,...,\mathscr{V}_L$ and select for the basis of $\mathscr{V}$ the union of some basis for the subspaces $\mathscr{V}_1,...,\mathscr{V}_L$. Show that with respect to this basis $M(\mathbf{A})$ has the following *block* form:

$$M(\mathbf{A}) = \begin{bmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_L \end{bmatrix} \tag{24.5}$$

where A_j is the matrix of the restriction of $\mathbf{A}$ to $\mathscr{V}_j$.

24.2 Use the result of Exercise 24.1 and prove Theorem 24.2.

24.3 Show that if $\mathscr{V} = \mathscr{V}_1 \oplus \mathscr{V}_2 \oplus \cdots \oplus \mathscr{V}_L$, then $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ is *reduced* by $\mathscr{V}_1,...,\mathscr{V}_L$ if and only if $\mathbf{AP}_j = \mathbf{P}_j\mathbf{A}$ for $j = 1,...,L$, where $\mathbf{P}_j$ is the projection on $\mathscr{V}_j$ given by (17.21).

Section 25 Eigenvectors and Eigenvalues

Given an endomorphism $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$, the problem of finding a direct sum decomposition of $\mathbf{A}$ is closely related to the study of the *spectral* properties of $\mathbf{A}$. This concept is central in the discussion of *eigenvalue problems*.

A scalar λ is an *eigenvalue* of $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ if there exists a nonzero vector $\mathbf{v} \in \mathscr{V}$ such that

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \tag{25.1}$$

The vector $\mathbf{v}$ in (25.1) is called an *eigenvector* of $\mathbf{A}$ corresponding to the eigenvalue λ. Eigenvalues and eigenvectors are sometimes refered to as *latent roots* and *latent vectors*, *characteristic roots* and *characteristic vectors*, and *proper values* and *proper vectors*. The set of all eigenvalues of $\mathbf{A}$ is the spectrum of $\mathbf{A}$, denoted by $\sigma(\mathbf{A})$. For any $\lambda \in \sigma(\mathbf{A})$, the set

$$\mathscr{V}(\lambda) = \{\mathbf{v} \in \mathscr{V} \,|\, \mathbf{A}\mathbf{v} = \lambda\mathbf{v}\}$$

is a subspace of $\mathscr{V}$, called the *eigenspace* or *characteristic subspace* corresponding to λ. The geometric multiplicity of λ is the dimension of $\mathscr{V}(\lambda)$.

Theorem 25.1. Given any $\lambda \in \sigma(\mathbf{A})$, the corresponding eigenspace $\mathscr{V}(\lambda)$ is $\mathbf{A}$-invariant.

The proof of this theorem involves an elementary use of (25.1). Given any $\lambda \in \sigma(\mathbf{A})$, the restriction of $\mathbf{v}$ to $\mathscr{V}(\lambda)$, denoted by $\mathbf{A}_{\mathscr{V}(\lambda)}$, has the property that

$$\mathbf{A}_{\mathscr{V}(\lambda)}\mathbf{u} = \lambda\mathbf{u} \tag{25.2}$$

for all $\mathbf{u}$ in $\mathscr{V}(\lambda)$. Geometrically, $\mathbf{A}_{\mathscr{V}(\lambda)}$ simply amplifies $\mathbf{u}$ by the factor λ. Such linear transformations are often called *dilatations*.

Theorem 25.2. $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ is not regular if and only if 0 is an eigenvalue of $\mathbf{A}$; further, the corresponding eigenspace is $K(\mathbf{A})$.

The proof is obvious. Note that (25.1) can be written as $(\mathbf{A} - \lambda\mathbf{I})\mathbf{v} = \mathbf{0}$, which shows that

$$\mathscr{V}(\lambda) = K(\mathbf{A} - \lambda\mathbf{I}) \tag{25.3}$$

Equation (25.3) implies that λ is an eigenvalue of $\mathbf{A}$ if and only if $\mathbf{A} - \lambda\mathbf{I}$ is singular. Therefore, by Theorem 22.2,

$$\lambda \in \sigma(\mathbf{A}) \Leftrightarrow \det(\mathbf{A} - \lambda\mathbf{I}) = 0 \tag{25.4}$$

The polynomial $f(\lambda)$ of degree $N = \dim\mathscr{V}$ defined by

$$f(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) \tag{25.5}$$

is called the *characteristic polynomial* of $\mathbf{A}$. Equation (25.5) shows that the eigenvalues of $\mathbf{A}$ are roots of the characteristic equation

$$f(\lambda) = 0 \tag{25.6}$$

Exercises

25.1 Let $\mathbf{A}$ be an endomorphism whose matrix $M(\mathbf{A})$ relative to a particular basis is a triangular matrix, say

$$M(\mathbf{A}) = \begin{bmatrix} A^1{}_1 & A^1{}_2 & \cdot & \cdot & \cdot & A^1{}_N \\ & A^2{}_2 & & & & \\ & & \cdot & & & \cdot \\ & & & \cdot & & \cdot \\ & & & & \cdot & \cdot \\ 0 & & & & & A^N{}_N \end{bmatrix}$$

What is the characteristic polynomial of $\mathbf{A}$? Use (25.6) and determine the eigenvalues of $\mathbf{A}$.

25.2 Show that the eigenvalues of a Hermitian endomorphism of an inner product space are all real numbers and the eigenvalues of a skew-Hermitian endomorphism are all purely imaginary numbers (including the number $0 = i0$).

25.3 Show that the eigenvalues of a unitary endomorphism all have absolute value 1 and the eigenvalues of a projection are either 1 or 0.

25.4 Show that the eigenspaces corresponding to different eigenvalues of a Hermitian endomorphism are mutually orthogonal.

25.5 If $\mathbf{B} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ is regular, show that $\mathbf{B}^{-1}\mathbf{AB}$ and $\mathbf{A}$ have the same characteristic equation. How are the eigenvectors of $\mathbf{B}^{-1}\mathbf{AB}$ related to those of $\mathbf{A}$?

Section 26 The Characteristic Polynomial

In the last section we found that the eigenvalues of $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ are roots of the characteristic polynomial

$$f(\lambda) = 0 \tag{26.1}$$

where

$$f(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) \tag{26.2}$$

If $\{\mathbf{e}_1,...,\mathbf{e}_2\}$ is a basis for $\mathscr{V}$, then by (19.6)

$$\mathbf{A}\mathbf{e}_k = A^j{}_k\mathbf{e}_j \tag{26.3}$$

Therefore, by (26.2) and (21.11),

$$f(\lambda) = \frac{1}{N!}\delta^{j_1\cdots j_N}_{i_1\ldots i_N}(A^{i_1}{}_{j_1} - \lambda\delta^{i_1}_{j_1})\cdots(A^{i_N}{}_{j_N} - \lambda\delta^{i_N}_{j_N}) \tag{26.4}$$

If (26.4) is expanded and we use (20.11), the result is

$$f(\lambda) = (-\lambda)^N + \mu_1(-\lambda)^{N-1} + \cdots + \mu_{N-1}(-\lambda) + \mu_N \tag{26.5}$$

where

$$\mu_j = \frac{1}{j!}\delta^{q_1,...,q_j}_{i_1,...,i_j}A^{i_1}{}_{q_1}\cdots A^{i_j}{}_{q_j} \tag{26.6}$$

Since $f(\lambda)$ is defined by (26.2), the coefficients μ_j, $j = 1,...,N$, are independent of the choice of basis for $\mathscr{V}$. These coefficients are called the *fundamental invariants* of $\mathbf{A}$. Equation (26.6) specializes to

$$\mu_1 = \operatorname{tr}\mathbf{A}, \qquad \mu_2 = \frac{1}{2}\left\{(\operatorname{tr}\mathbf{A})^2 - \operatorname{tr}\mathbf{A}^2\right\}, \qquad \text{and} \qquad \mu_N = \det\mathbf{A} \tag{26.7}$$

where (21.4) has been used to obtain $(26.7)_2$. Since $f(\lambda)$ is a Nth degree polynomial, it can be factored into the form

$$f(\lambda) = (\lambda_1 - \lambda)^{d_1}(\lambda_2 - \lambda)^{d_2}\cdots(\lambda_L - \lambda)^{d_L} \tag{26.8}$$

where $\lambda_1,...,\lambda_L$ are the distinct foots of $f(\lambda) = 0$ and $d_1,...,d_L$ are positive integers which must satisfy $\sum_{j=1}^{L} d_j = N$. It is in writing (26.8) that we have made use of the assumption that the scalar field is complex. If the scalar field is real, the polynomial (26.5), generally, cannot be factored.

In general, a scalar field is said to be *algebraically closed* if every polynomial equation has at least one root in the field, or equivalently, if every polynomial, such as $f(\lambda)$, can be factored into the form (26.8). It is a theorem in algebra that the complex field is algebraically closed. The real field, however, is not algebraically closed. For example, if λ is real the polynomial equation $f(\lambda) = \lambda^2 + 1 = 0$ has no real roots. By allowing the scalar fields to be complex, we are assured that every endomorphism *has at least one eigenvector*. In the expression (26.8), the integer d_j is called the *algebraic multiplicity* of the eigenvalue λ_j. It is possible to prove that the algebraic multiplicity of an eigenvalue is not less than the geometric multiplicity of the same eigenvalue. However, we shall postpone the proof of this result until Section 30.

An expression for the invariant μ_j can be obtained in terms of the eigenvalues if (26.8) is expanded and the results are compared to (26.5). For example,

$$\begin{aligned} \mu_1 &= \operatorname{tr}\mathbf{A} = d_1\lambda_1 + d_2\lambda_2 + \cdots + d_L\lambda_L \\ \mu_N &= \det\mathbf{A} = \lambda_1^{d_1}\lambda_2^{d_2}\cdots\lambda_L^{d_L} \end{aligned} \tag{26.9}$$

The next theorem we want to prove is known as the *Cayley-Hamilton* theorem. Roughly speaking, this theorem asserts that an endomorphism satisfies its own characteristic equation. To make this statement clear, we need to introduce certain ideas associated with polynomials of endomorphisms. As we have done in several places in the preceding chapters, if $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$, $\mathbf{A}^2$ is defined by

$$\mathbf{A}^2 = \mathbf{A}\mathbf{A}$$

Similarly, we define by induction, starting from

$$\mathbf{A}^0 = \mathbf{I}$$

and in general

$$\mathbf{A}^k \equiv \mathbf{A}\mathbf{A}^{k-1} = \mathbf{A}^{k-1}\mathbf{A}$$

where k is any integer greater than one. If k and l are positive integers, it is easily established by induction that

$$\mathbf{A}^k\mathbf{A}^l = \mathbf{A}^l\mathbf{A}^k = \mathbf{A}^{l+k} \tag{26.10}$$

Thus $\mathbf{A}^k$ and $\mathbf{A}^l$ commute. A *polynomial* in $\mathbf{A}$ is an endomorphism of the form

$$g(\mathbf{A}) = \alpha_0\mathbf{A}^M + \alpha_1\mathbf{A}^{M-1} + \cdots + \alpha_{M-1}\mathbf{A} + \alpha_M\mathbf{I} \tag{26.11}$$

where M is a positive integer and $\alpha_0, ..., \alpha_M$ are scalars. Such polynomials have certain of the properties of polynomials of scalars. For example, if a scalar polynomial

$$g(t) = \alpha_0 t^M + \alpha_1 t^{M-1} + \cdots + \alpha_{M-1}t + \alpha_M$$

can be factored into

$$g(t) = \alpha_0(t-\eta_1)(t-\eta_2)\cdots(t-\eta_M)$$

then the polynomial (26.11) can be factored into

$$g(\mathbf{A}) = \alpha_0(\mathbf{A}-\eta_1\mathbf{I})(\mathbf{A}-\eta_2\mathbf{I})\cdots(\mathbf{A}-\eta_M\mathbf{I}) \tag{26.12}$$

The order of the factors in (26.12) is not important since, as a result of (26.10), the factors commute. Notice, however, that the product of two nonzero endomorphisms can be zero. Thus the formula

$$g_1(\mathbf{A})g_2(\mathbf{A}) = \mathbf{0} \tag{26.13}$$

for two polynomials g_1 and g_2 generally *does not* imply one of the factors is zero. For example, any projection $\mathbf{P}$ satisfies the equation $\mathbf{P(P - I)} = \mathbf{0}$, but generally $\mathbf{P}$ and $\mathbf{P - I}$ are both nonzero.

Theorem 26.1. (Cayley-Hamilton). If $f(\lambda)$ is the characteristic polynomial (26.5) for an endomorphism $\mathbf{A}$, then

$$f(\mathbf{A}) = (-\mathbf{A})^N + \mu_1(-\mathbf{A})^{N-1} + \cdots + \mu_{N-1}(-\mathbf{A}) + \mu_N\mathbf{I} = \mathbf{0} \tag{26.14}$$

Proof. : The proof which we shall now present makes major use of equation (0.29) or, equivalently, (21.19). If $\text{adj}(\mathbf{A} - \lambda\mathbf{I})$ is the endomorphism whose matrix is $\text{adj}\left[A^p{}_q - \lambda\delta^p_q\right]$, where $\left[A^p{}_q\right] = M(\mathbf{A})$, then by (26.2) and (0.29) (see also Exercise 40.5)

$$(\text{adj}(\mathbf{A} - \lambda\mathbf{I}))(\mathbf{A} - \lambda\mathbf{I}) = f(\lambda)\mathbf{I} \tag{26.15}$$

By (21.18) it follows that $\text{adj}(\mathbf{A} - \lambda\mathbf{I})$ is a polynomial of degree $N-1$ in λ. Therefore

$$\text{adj}(\mathbf{A} - \lambda\mathbf{I}) = \mathbf{B}_0(-\lambda)^{N-1} + \mathbf{B}_1(-\lambda)^{N-2} + \cdots + \mathbf{B}_{N-2}(-\lambda) + \mathbf{B}_{N-1} \tag{26.16}$$

where $\mathbf{B}_0, ..., \mathbf{B}_{N-1}$ are endomorphisms determined by $\mathbf{A}$. If we now substitute (26.16) and (26.15) into (26.14) and require the result to hold for all λ, we find

$$\begin{aligned}
\mathbf{B}_0 &= \mathbf{I} \\
\mathbf{B}_0\mathbf{A} + \mathbf{B}_1 &= \mu_1\mathbf{I} \\
\mathbf{B}_1\mathbf{A} + \mathbf{B}_2 &= \mu_2\mathbf{I} \\
&\vdots \\
\mathbf{B}_{N-2}\mathbf{A} + \mathbf{B}_{N-1} &= \mu_{N-1}\mathbf{I} \\
\mathbf{B}_{N-1}\mathbf{A} &= \mu_N\mathbf{I}
\end{aligned} \tag{26.17}$$

Now we multiply $(26.17)_1$. by $(-\mathbf{A})^N$, $(26.17)_2$ by $(-\mathbf{A})^{N-1}$, $(26.17)_3$ by $(-\mathbf{A})^{N-2}$,…,$(26.17)_k$ by $(-\mathbf{A})^{N-k+1}$, etc., and add the resulting N equations, to find

$$(-\mathbf{A})^N + \mu_1(-\mathbf{A})^{N-1} + \cdots + \mu_{N-1}(-\mathbf{A}) + \mu_N\mathbf{I} = \mathbf{0} \tag{26.18}$$

which is the desired result.

Exercises

26.1 If $N = \dim\mathscr{V}$ is odd, and the scalar field is $\mathscr{R}$, prove that the characteristic equation of any $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ has at least one eigenvalue.

26.2 If $N = \dim\mathscr{V}$ is odd and the scalar field is $\mathscr{R}$, prove that if $\det\mathbf{A} > 0 (< 0)$, then $\mathbf{A}$ has at least one positive (negative) eigenvalue.

26.3 If $N = \dim\mathscr{V}$ is even and the scalar field is $\mathscr{R}$ and $\det\mathbf{A} < 0$, prove that $\mathbf{A}$ has at least one positive and one negative eigenvalue.

26.4 Let $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ be regular. Show that

$$\det(\mathbf{A}^{-1} - \lambda^{-1}\mathbf{I}) = (-\lambda)^{-N} \det\mathbf{A}^{-1} \det(\mathbf{A} - \lambda\mathbf{I})$$

where $N = \dim\mathscr{V}$.

26.5 Prove that the characteristic polynomial of a projection $\mathbf{P} : \mathscr{V} \to \mathscr{V}$ is

$$f(\lambda) = (-1)^{-N} \lambda^{N-L}(1 - \lambda)$$

where $N = \dim\mathscr{V}$ and $L = \dim\mathbf{P}(\mathscr{V})$.

26.6 If $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ is regular, express $\mathbf{A}^{-1}$ as a polynomial in $\mathbf{A}$.

26.7 Prove directly that the quantity μ_j defined by (26.6) is independent of the choice of basis for $\mathscr{V}$.

26.8 Let $\mathbf{C}$ be an endomorphism whose matrix relative to a basis $\{\mathbf{e}_1,...,\mathbf{e}_N\}$ is a triangular matrix of the form

$$M(\mathbf{C}) = \begin{bmatrix} 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ & 0 & 1 & 0 & & & \\ & & \cdot & \cdot & \cdot & & \\ & & & \cdot & \cdot & \cdot & \\ & & & & \cdot & \cdot & 0 \\ & & & & & 0 & 1 \\ 0 & & & & & & 0 \end{bmatrix} \tag{26.19}$$

i.e. $\mathbf{C}$ maps $\mathbf{e}_1$ to $\mathbf{0}$, $\mathbf{e}_2$ to $\mathbf{e}_1$, $\mathbf{e}_3$ to $\mathbf{e}_2$,…. Show that a necessary and sufficient condition for the existence of a component matrix of the form (26.19) for an endomorphism $\mathbf{C}$ is that

$$\mathbf{C}^N = \mathbf{0} \qquad \text{but} \qquad \mathbf{C}^{N-1} \neq \mathbf{0} \tag{26.20}$$

We call such an endomorphism $\mathbf{C}$ a *nilcyclic* endomorphism and we call the basis $\{\mathbf{e}_1,...,\mathbf{e}_N\}$ for the form (26.19) a *cyclic basis* for $\mathbf{C}$.

26.9 If $\phi_1,...,\phi_N$ denote the fundamental invariants of $\mathbf{A}^{-1}$, use the results of Exercise 26.4 to show that

$$\phi_j = (\det \mathbf{A}^{-1})\mu_{N-j}$$

for $j = 1,...,N$.

26.10 It follows from (26.16) that

$$\mathbf{B}_{N-1} = \operatorname{adj}\mathbf{A}$$

Use this result along with (26.17) and show that

$$\operatorname{adj}\mathbf{A} = (-\mathbf{A})^{N-1} + \mu_1(-\mathbf{A})^{N-2} + \cdots + \mu_{N-1}\mathbf{I}$$

26.11 Show that

$$\mu_{N-1} = \operatorname{tr}(\operatorname{adj}\mathbf{A})$$

and, from Exercise (26.10), that

$$\mu_{N-1} = -\frac{1}{N-1}\left\{\operatorname{tr}(-\mathbf{A})^{N-1} + \mu_1 \operatorname{tr}(-\mathbf{A})^{N-2} + \cdots + \mu_{N-2} \operatorname{tr}(-\mathbf{A})\right\}$$

Section 27. Spectral Decomposition for Hermitian Endomorphisms

An important problem in mathematical physics is to find a basis for a vector space $\mathscr{V}$ in which the matrix of a given $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ is diagonal. If we examine (25.1), we see that if $\mathscr{V}$ has a basis of eigenvectors of $\mathbf{A}$, then $M(\mathbf{A})$ is diagonal and vice versa; further, in this case the diagonal elements of $M(\mathbf{A})$ are the eigenvalues of $\mathbf{A}$. As we shall see, not all endomorphisms have matrices which are diagonal. Rather than consider this problem in general, we specialize here to the case where $\mathbf{A}$,.then $M(\mathbf{A})$ is Hermitian and show that every Hermitian endomorphism has a matrix which takes on the diagonal form. The general case will be treated later, in Section 30. First, we prove a general theorem for arbitrary endomorphisms about the linear independence of eigenvectors.

Theorem 27.1. If $\lambda_1,...,\lambda_L$ are distinct eigenvalues of $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ and if $\mathbf{u}_1,...,\mathbf{u}_L$ are eigenvectors corresponding to them, then $\{\mathbf{u}_1,...,\mathbf{u}_L\}$ form a linearly independent set.

Proof. If $\{\mathbf{u}_1,...,\mathbf{u}_L\}$ is not linearly independent, we choose a maximal, linearly independent subset, say $\{\mathbf{u}_1,...,\mathbf{u}_S\}$, from the set $\{\mathbf{u}_1,...,\mathbf{u}_L\}$; then the remaining vectors can be expressed uniquely as linear combinations of $\{\mathbf{u}_1,...,\mathbf{u}_S\}$, say

$$\mathbf{u}_{S+1} = \alpha_1\mathbf{u}_1 + \cdots + \alpha_S\mathbf{u}_S \tag{27.1}$$

where $\alpha_1,...,\alpha_S$ are not all zero and unique, because $\{\mathbf{u}_1,...,\mathbf{u}_S\}$ is linearly independent. Applying $\mathbf{A}$ to (27.1) yields

$$\lambda_{S+1}\mathbf{u}_{S+1} = (\alpha_1\lambda_1)\mathbf{u}_1 + \cdots + (\alpha_S\lambda_S)\mathbf{u}_S \tag{27.2}$$

Now if $\lambda_{S+1} = 0$, then $\lambda_1,...,\lambda_S$ are nonzero because the eigenvalues are distinct and (27.2) contradicts the linear independence of $\{\mathbf{u}_1,...,\mathbf{u}_S\}$; on the other hand, if $\lambda_{S+1} \neq 0$, then we can divide (27.2) by λ_{S+1}, obtaining another expression of $\mathbf{u}_{S+1}$ as a linear combination of $\{\mathbf{u}_1,...,\mathbf{u}_S\}$ contradicting the uniqueness of the coefficients $\alpha_1,...,\alpha_S$. Hence in any case the maximal linearly independent subset cannot be a proper subset of $\{\mathbf{u}_1,...,\mathbf{u}_L\}$; thus $\{\mathbf{u}_1,...,\mathbf{u}_L\}$ is linearly independent.

As a corollary to the preceding theorem, we see that if the geometric multiplicity is equal to the algebraic multiplicity for each eigenvalue of $\mathbf{A}$, then the vector space $\mathscr{V}$ admits the direct sum representation

$$\mathscr{V} = \mathscr{V}(\lambda_1) \oplus \mathscr{V}(\lambda_2) \oplus \cdots \oplus \mathscr{V}(\lambda_L)$$

where $\lambda_1, ..., \lambda_L$ are the distinct eigenvalues of $\mathbf{A}$. The reason for this representation is obvious, since the right-hand side of the above equation is a subspace having the same dimension as $\mathscr{V}$; thus that subspace is equal to $\mathscr{V}$. Whenever the representation holds, we can always choose a basis of $\mathscr{V}$ formed by bases of the subspaces $\mathscr{V}(\lambda_1), ..., \mathscr{V}(\lambda_L)$. Then this basis consists entirely of eigenvectors of $\mathbf{A}$ becomes a diagonal matrix, namely,

$$M(\mathbf{A}) = \begin{bmatrix} \lambda_1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & \lambda_1 & & & & & & \\ & & & \lambda_2 & & & 0 & & \\ & & & & \ddots & & & & \\ & & & & & \lambda_2 & & & \\ & & & & & & \ddots & & \\ & & & & & & & \lambda_L & \\ & & 0 & & & & & \ddots & \\ & & & & & & & & \lambda_L \end{bmatrix} \qquad (27.3)_1$$

where each λ_K is repeated d_k times, d_k being the algebraic as well as the geometric multiplicity of λ_K. Of course, the representation of $\mathscr{V}$ by direct sum of eigenspaces of $\mathbf{A}$ is possible if $\mathbf{A}$ has $N = \dim \mathscr{V}$ distinct eigenvalues. In this case the matrix of $\mathbf{A}$ taken with respect to a basis of eigenvectors has the diagonal form

$$M(\mathbf{A}) = \begin{bmatrix} \lambda_1 & & & & & \\ & \lambda_2 & & & 0 & \\ & & \lambda_3 & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & 0 & & & \cdot & \\ & & & & & \lambda_N \end{bmatrix} \tag{27.3}_2$$

If the eigenvalues of $\mathbf{v}$ are not all distinct, then in general the geometric multiplicity of an eigenvalue may be less than the algebraic multiplicity. Whenever the two multiplicities are different for at least one eigenvalue of $\mathbf{A}$, it is no longer possible to find any basis in which the matrix of $\mathbf{A}$ is diagonal. However, if $\mathscr{V}$ is an inner product space and if $\mathbf{A}$ is Hermitian, then a diagonal matrix of $\mathbf{A}$ can always be found; we shall now investigate this problem.

Recall that if $\mathbf{u}$ and $\mathbf{v}$ are arbitrary vectors in $\mathscr{V}$, the adjoint $\mathbf{A}^*$ of $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ is defined by

$$\mathbf{Au} \cdot \mathbf{v} = \mathbf{u} \cdot \mathbf{A}^* \mathbf{v} \tag{27.4}$$

As usual, if the matrix of $\mathbf{A}$ is referred to a basis $\{\mathbf{e}_k\}$, then the matrix of $\mathbf{A}^*$ is referred to the reciprocal basis $\{\mathbf{e}^k\}$ and is given by [cf. (18.4)]

$$M(\mathbf{A}^*) = \overline{M(\mathbf{A})^T} \tag{27.5}$$

where the overbar indicates the complex conjugate as usual. If $\mathbf{A}$ Hermitian, i.e., if $\mathbf{A} = \mathbf{A}^*$, then (27.4) reduces to

$$\mathbf{Au} \cdot \mathbf{v} = \mathbf{u} \cdot \mathbf{Av} \tag{27.6}$$

for all $\mathbf{u}, \mathbf{v} \in \mathscr{V}$.

Theorem 27.2. The eigenvalues of a Hermitian endomorphism are all real.

Proof. Let $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ be Hermitian. Since $\mathbf{Au} = \lambda\mathbf{u}$ for any eigenvalue λ, we have

$$\lambda = \frac{\mathbf{Au} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}} \tag{27.7}$$

Therefore we must show that $\mathbf{Au} \cdot \mathbf{u}$ is real or, equivalently, we must show $\mathbf{Au} \cdot \mathbf{u} = \overline{\mathbf{Au} \cdot \mathbf{u}}$. By (27.6)

$$\mathbf{Au} \cdot \mathbf{u} = \mathbf{u} \cdot \mathbf{Au} = \overline{\mathbf{Au} \cdot \mathbf{u}} \tag{27.8}$$

where the rule $\mathbf{u} \cdot \mathbf{v} = \overline{\mathbf{v} \cdot \mathbf{u}}$ has been used. Equation (27.8) yields the desired result.

Theorem 27.3. If $\mathbf{A}$ is Hermitian and if $\mathscr{V}_1$ is an $\mathbf{A}$-invariant subspace of $\mathscr{V}$, then $\mathscr{V}_1^{\perp}$ is also $\mathbf{A}$-invariant .

Proof. If $\mathbf{v} \in \mathscr{V}_1$ and $\mathbf{u} \in \mathscr{V}_1^{\perp}$, then $\mathbf{Av} \cdot \mathbf{u} = \mathbf{0}$ because $\mathbf{Av} \in \mathscr{V}_1$. But since $\mathbf{A}$ is Hermitian, $\mathbf{Av} \cdot \mathbf{u} = \mathbf{v} \cdot \mathbf{Au} = 0$. Therefore, $\mathbf{Au} \in \mathscr{V}_1^{\perp}$, which proves the theorem.

Theorem 27.4. If $\mathbf{A}$ is Hermitian, the algebraic multiplicity of each eigenvalue equa1s the geometric multiplicity.

Proof Let $\mathscr{V}(\lambda_0)$ be the characteristic subspace associated with an eigenvalue λ_0. Then the geometric multiplicity of λ_0 is $M = \dim \mathscr{V}(\lambda_0)$. By Theorems 13.4 and 27.3

$$\mathscr{V} = \mathscr{V}(\lambda_0) \oplus \mathscr{V}(\lambda_0)^{\perp} \tag{27.9}$$

Where $\mathscr{V}(\lambda_0)$ and $\mathscr{V}(\lambda_0)^{\perp}$ are $\mathbf{A}$-invariant. By Theorem 24.1

$$\mathbf{A} = \mathbf{A}_1 + \mathbf{A}_2$$

and by Theorem 17.4

$$\mathbf{I} = \mathbf{P}_1 + \mathbf{P}_2$$

where $\mathbf{P}_1$ projects $\mathscr{V}$ onto $\mathscr{V}(\lambda_0)$, $\mathbf{P}_2$ projects $\mathscr{V}$ onto $\mathscr{V}(\lambda_0)^\perp$, $\mathbf{A}_1 = \mathbf{A}\mathbf{P}_1$, and $\mathbf{A}_2 = \mathbf{A}\mathbf{P}_2$. By Theorem 18.10, $\mathbf{P}_1$ and $\mathbf{P}_2$ are Hermitian and they also commute with $\mathbf{A}$. Indeed, for any $\mathbf{v} \in \mathscr{V}$, $\mathbf{P}_1\mathbf{v} \in \mathscr{V}(\lambda_0)$ and $\mathbf{P}_2\mathbf{v} \in \mathscr{V}(\lambda_0)^\perp$, and thus $\mathbf{A}\mathbf{P}_1\mathbf{v} \in \mathscr{V}(\lambda_0)$ and $\mathbf{A}\mathbf{P}_2\mathbf{v} \in \mathscr{V}(\lambda_0)^\perp$. But since

$$\mathbf{A}\mathbf{v} = \mathbf{A}(\mathbf{P}_1 + \mathbf{P}_2)\mathbf{v} = \mathbf{A}\mathbf{P}_1\mathbf{v} + \mathbf{A}\mathbf{P}_2\mathbf{v}$$

we see that $\mathbf{A}\mathbf{P}_1\mathbf{v}$ is the $\mathscr{V}(\lambda_0)$ component of $\mathbf{A}\mathbf{v}$ and $\mathbf{A}\mathbf{P}_2\mathbf{v}$ is the $\mathscr{V}(\lambda_0)^\perp$ component of $\mathbf{A}\mathbf{v}$. Therefore

$$\mathbf{P}_1\mathbf{A}\mathbf{v} = \mathbf{A}\mathbf{P}_1\mathbf{v}, \qquad \mathbf{P}_2\mathbf{A}\mathbf{v} = \mathbf{A}\mathbf{P}_2\mathbf{v}$$

for all $\mathbf{v} \in \mathscr{V}$, or, equivalently

$$\mathbf{P}_1\mathbf{A} = \mathbf{A}\mathbf{P}_1, \qquad \mathbf{P}_2\mathbf{A} = \mathbf{A}\mathbf{P}_2$$

Together with the fact that $\mathbf{P}_1$, and $\mathbf{P}_2$ are Hermitian, these equations imply that $\mathbf{A}_1$ and $\mathbf{A}_2$ are also Hermitian. Further, $\mathbf{A}$ is reduced by the subspaces $\mathscr{V}(\lambda_0)$ and $\mathscr{V}(\lambda_0)^\perp$, since

$$\mathbf{A}_1\mathbf{A}_2 = \mathbf{A}\mathbf{P}_1\mathbf{A}\mathbf{P}_2 = \mathbf{A}^2\mathbf{P}_1\mathbf{P}_2$$

Thus if we select a basis $\{\mathbf{e}_1, ..., \mathbf{e}_N\}$ such that $\{\mathbf{e}_1, ..., \mathbf{e}_M\}$ span $\mathscr{V}(\lambda_0)$ and $\{\mathbf{e}_{M+1}, ..., \mathbf{e}_N\}$ span $\mathscr{V}(\lambda_0)^\perp$, then by the result of Exercise 24.1 the matrix of $\mathbf{A}$ to $\{\mathbf{e}_k\}$ takes the form

$$M(\mathbf{A}) = \left[\begin{array}{ccc|ccc} A^1{}_1 & \cdots & A^1{}_M & & & \\ \vdots & & \vdots & & 0 & \\ A^M{}_1 & \cdots & A^M{}_M & & & \\ \hline & & & A^{M+1}{}_{M+1} & \cdots & A^{M+1}{}_N \\ & 0 & & \vdots & & \vdots \\ & & & A^N{}_{M+1} & \cdots & A^N{}_N \end{array}\right]$$

and the matrices of $\mathbf{A}_1$ and $\mathbf{A}_2$ are

$$M(\mathbf{A}_1) = \begin{bmatrix} \begin{matrix} A^1{}_1 & \cdot & \cdot & \cdot & A^1{}_M \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ A^M{}_1 & \cdot & \cdot & \cdot & A^M{}_M \end{matrix} & \begin{matrix} \\ \\ 0 \\ \\ \\ \end{matrix} \\ \begin{matrix} \\ 0 \\ \\ \\ \\ \end{matrix} & \begin{matrix} \\ 0 \\ \\ \\ \\ \end{matrix} \end{bmatrix}$$

$$M(\mathbf{A}_2) = \begin{bmatrix} \begin{matrix} \\ 0 \\ \\ \\ \\ \end{matrix} & \begin{matrix} \\ 0 \\ \\ \\ \\ \end{matrix} \\ \begin{matrix} \\ \\ 0 \\ \\ \\ \end{matrix} & \begin{matrix} A^{M+1}{}_{M+1} & \cdot & \cdot & \cdot & A^{M+1}{}_N \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ A^N{}_{M+1} & \cdot & \cdot & \cdot & A^N{}_N \end{matrix} \end{bmatrix} \tag{27.10}$$

which imply

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \det(\mathbf{A}_1 - \lambda\mathbf{I}_{\mathscr{V}(\lambda_0)}) \det(\mathbf{A}_2 - \lambda\mathbf{I}_{\mathscr{V}(\lambda_0)^\perp})$$

By (25.2), $\mathbf{A}_1 = \lambda_0\mathbf{I}_{\mathscr{V}(\lambda_0)}$; thus by (21.21)

$$\det(\mathbf{A} - \lambda\mathbf{I}) = (\lambda_0 - \lambda)^M \det(\mathbf{A}_2 - \lambda\mathbf{I}_{\mathscr{V}(\lambda_0)^\perp})$$

On the other hand, λ_0 is not an eigenvalue of $\mathbf{A}_2$. Therefore

$$\det(\mathbf{A}_2 - \lambda\mathbf{I}_{\mathscr{V}(\lambda_0)^\perp}) \neq 0$$

Hence the algebraic multiplicity of λ_0 equals *M*, the geometric multiplicity.

The preceding theorem implies immediately the important result that $\mathscr{V}$ can be represented by a direct sum of eigenspaces of $\mathbf{A}$ if $\mathbf{A}$ is Hermitian. In particular, there exists a basis consisting entirely in eigenvectors of $\mathbf{A}$, and the matrix of $\mathbf{A}$ relative to this basis is diagonal. However, before we state this result formally as a theorem, we first strengthen the result of Theorem 27.1 for the special case that $\mathbf{A}$ is Hermitian.

Theorem 27.5. If $\mathbf{A}$ is Hermitian, the eigenspaces corresponding to distinct eigenvalues λ_1 and λ_2 are orthogonal.

Proof. Let $\mathbf{A}\mathbf{u}_1 = \lambda_1\mathbf{u}_1$ and $\mathbf{A}\mathbf{u}_2 = \lambda_2\mathbf{u}_2$. Then

$$\lambda_1\mathbf{u}_1 \cdot \mathbf{u}_2 = \mathbf{A}\mathbf{u}_1 \cdot \mathbf{u}_2 = \mathbf{u}_1 \cdot \mathbf{A}\mathbf{u}_2 = \lambda_2\mathbf{u}_1 \cdot \mathbf{u}_2$$

Since $\lambda_1 \neq \lambda_2$, $\mathbf{u}_1 \cdot \mathbf{u}_2 = 0$, which proves the theorem.

The main theorem regarding Hermitian endomorphisms is the following.

Theorem 27.6. If $\mathbf{A}$ is a Hermitian endomorphism with (distinct) eigenvalues $\lambda_1, \lambda_2, ..., \lambda_N$, then $\mathscr{V}$ has the representation

$$\mathscr{V} = \mathscr{V}_1(\lambda_1) \oplus \mathscr{V}(\lambda_2) \oplus \cdots \oplus \mathscr{V}(\lambda_L) \tag{27.11}$$

where the eigenspaces $\mathscr{V}(\lambda_k)$ are mutually orthogonal.

The proof of this theorem is a direct consequence of Theorems 27.4 and 27.5 and the remark following the proof of Theorem 27.1.

Corollary (*Spectral Theorem*). If $\mathbf{A}$ is a Hermitian endomorphish with (distinct) eigenvalues $\lambda_1,...,\lambda_L$ then

$$\mathbf{A} = \sum_{j=1}^{N} \lambda_j \mathbf{P}_j \tag{27.12}$$

where $\mathbf{P}_j$ is the perpendicular projection of $\mathscr{V}$ onto $\mathscr{V}(\lambda_j)$, for $j = 1,...,L$.

Proof. By Theorem 27.6, $\mathbf{A}$ has a representation of the form (24.1). Let $\mathbf{u}$ be an arbitrary element of $\mathscr{V}$, then, by (27.11),

$$\mathbf{u} = \mathbf{u}_1 + \cdots + \mathbf{u}_L \tag{27.13}$$

where $\mathbf{u}_j \in \mathscr{V}(\lambda_j)$. By (24.3), (27.13), and (25.1)

$$\mathbf{A}_j\mathbf{u} = \mathbf{A}_j\mathbf{u}_j = \mathbf{A}\mathbf{u}_j = \lambda_j\mathbf{u}_j$$

But $\mathbf{u}_j = \mathbf{P}_j\mathbf{u}$; therefore

$$\mathbf{A}_j = \lambda_j\mathbf{P}_j \qquad \text{(no sum)}$$

which, with (24.1) proves the corollary.

The reader is reminded that the L perpendicular projections satisfy the equations

$$\sum_{j=1}^{L} \mathbf{P}_j = \mathbf{I} \tag{27.14}$$

$$\mathbf{P}_j{}^2 = \mathbf{P}_j \tag{27.15}$$

$$\mathbf{P}_j = \mathbf{P}_j^* \tag{27.16}$$

and

$$\mathbf{P}_j\mathbf{P}_i = \mathbf{0} \qquad i \neq j \tag{27.17}$$

These equations follow from Theorems 14.4, 18.10 and 18.11. Certain other endomorphisms also have a spectral representation of the form (27.12); however, the projections are not perpendicular ones and do not obey the condition (27.17).

Another Corollary of Theorem 27.6 is that if $\mathbf{A}$ is Hermitian, there exists an orthogonal basis for $\mathscr{V}$ consisting entirely of eigenvectors of $\mathbf{A}$. This corollary is clear because each eigenspace is orthogonal to the other and within each eigenspace an orthogonal basis can be selected. (Theorem 13.3 ensures that an orthonormal basis for each eigenspace can be found.) With respect to this basis of eigenvectors, the matrix of $\mathbf{A}$ is clearly diagonal. Thus the problem of finding a basis for $\mathscr{V}$ such that $M(\mathbf{A})$ is diagonal is solved for Hermitian endomorphisms.

If $f(\mathbf{A})$ is any polynomial in the Hermitian endomorphism, then (27.12),(27.15) and (27.17) can be used to show that

$$f(\mathbf{A}) = \sum_{j=1}^{L} f(\lambda_j)\mathbf{P}_j \tag{27.18}$$

where $f(\lambda)$ is the same polynomial except that the variable $\mathbf{A}$ is replaced by the scalar λ. For example, the polynomial $\mathbf{P}^2$ has the representation

$$\mathbf{A}^2 = \sum_{j=1}^{N} \lambda_j^2 \mathbf{P}_j$$

In general, $f(\mathbf{A})$ is Hermitian if and only if $f(\lambda_j)$ is real for all $j = 1,...,L$. If the eigenvalues of $\mathbf{A}$ are all nonnegative, then we can extract *Hermitian roots* of $\mathbf{A}$ by the following rule:

$$\mathbf{A}^{1/k} = \sum_{j=1}^{N} \lambda_j^{1/k} \mathbf{P}_j \tag{27.19}$$

where $\lambda_j^{1/k} \geq 0$. Then we can verify easily that $\left(\mathbf{A}^{1/k}\right)^k = \mathbf{A}$. If $\mathbf{A}$ has no zero eigenvalues, then

$$\mathbf{A}^{-1} = \sum_{j=1}^{L} \frac{1}{\lambda_j} \mathbf{P}_j \tag{27.20}$$

which is easily confirmed.

A Hermitian endomorphism $\mathbf{A}$ is defined to be

$$\left\{\begin{array}{l}\text{positive definite}\\ \text{positive semidefinite}\\ \text{negative semidefinite}\\ \text{negative definite}\end{array}\right\} \quad \text{if} \quad \mathbf{v}\cdot\mathbf{A}\mathbf{v}\left\{\begin{array}{l}>0\\ \geq 0\\ \leq 0\\ <0\end{array}\right\}$$

all nonzero $\mathbf{v}$, It follows from (27.12) that

$$\mathbf{v}\cdot\mathbf{A}\mathbf{v} = \sum_{j=1}^{L} \lambda_j \mathbf{v}_j \cdot \mathbf{v}_j \tag{27.21}$$

where

$$\mathbf{v} = \sum_{j=1}^{L} \mathbf{v}_j$$

Equation (27.21) implies the following important theorem.

Theorem 27.7 A Hermitian endomorphism $\mathbf{A}$ is

$$\left\{\begin{array}{l}\text{positive definite}\\ \text{positive semidefinite}\\ \text{negative semidefinite}\\ \text{negative definite}\end{array}\right\}$$

if and only if every eigenvalue of $\mathbf{A}$ is

$$\left\{\begin{matrix} >0 \\ \geq 0 \\ \leq 0 \\ <0 \end{matrix}\right\}$$

As corollaries to Theorem 27.7 it follows that positive-definite and negative-definite Hermitian endomorphisms are regular [see (27.20)], and positive-definite and positive-semidefinite endomorphisms possess Hermitian roots.

Every complex number z has a polar representation in the form $z = re^{i\theta}$, where $r \geq 0$. It turns out that an endomorphism also has polar decompositions, and we shall deduce an important special case here.

Theorem 27.8. (Polar Decomposition Theorem). Every automorphism $\mathbf{A}$ has two unique multiplicative decompositions

$$\mathbf{A} = \mathbf{RU} \qquad \text{and} \qquad \mathbf{A} = \mathbf{VR} \tag{27.22}$$

where $\mathbf{R}$ is unitary and $\mathbf{U}$ and $\mathbf{V}$ are Hermitian and positive definite.

Proof. : For each automorphism $\mathbf{A}$, $\mathbf{A}^*\mathbf{A}$ is a positive-definite Hermitian endomorphism, and hence it has a spectral decomposition of the form

$$\mathbf{A}^*\mathbf{A} = \sum_{j=1}^{L} \lambda_j \mathbf{P}_j \tag{27.23}$$

where $\lambda_j > 0, j = 1,...,L$. We define $\mathbf{U}$ by

$$\mathbf{U} = \left(\mathbf{A}^*\mathbf{A}\right)^{1/2} = \sum_{j=1}^{L} \lambda_j^{1/2} \mathbf{P}_j \tag{27.24}$$

Clearly $\mathbf{U}$ is Hermitian and positive definite. Since $\mathbf{U}$ is positive definite it is regular. We now define

$$\mathbf{R} = \mathbf{A}\mathbf{U}^{-1} \tag{27.25}$$

By this formula $\mathbf{R}$ is regular and satisfies

$$\mathbf{R}^*\mathbf{R} = \mathbf{U}^{-1}\mathbf{A}^*\mathbf{A}\mathbf{U}^{-1} = \mathbf{U}^{-1}\mathbf{U}^2\mathbf{U}^{-1} = \mathbf{I} \tag{27.26}$$

Therefore $\mathbf{R}$ is unitary. To prove the uniqueness, assume

$$\mathbf{R}\mathbf{U} = \mathbf{R}_1\mathbf{U}_1 \tag{27.27}$$

and we shall prove $\mathbf{R} = \mathbf{R}_1$ and $\mathbf{U} = \mathbf{U}_1$. From (27.27)

$$\mathbf{U}^2 = \mathbf{U}\mathbf{R}^*\mathbf{R}\mathbf{U} = (\mathbf{R}\mathbf{U})^*\mathbf{R}\mathbf{U} = (\mathbf{R}_1\mathbf{U}_1)\mathbf{R}_1\mathbf{U}_1 = \mathbf{U}_1^2$$

Since the positive-definite Hermitian square root of $\mathbf{U}^2$ is unique, we find $\mathbf{U} = \mathbf{U}_1$. Then (27.27) implies $\mathbf{R} = \mathbf{R}_1$. The decomposition $(27.22)_2$ follows either by defining $\mathbf{V}^2 = \mathbf{A}\mathbf{A}^*$, or, equivalently, by defining $\mathbf{V} = \mathbf{R}\mathbf{U}\mathbf{R}^*$ and repeating the above argument.

A more general theorem to the last one is true even if $\mathbf{A}$ is not required to be regular. However, in this case $\mathbf{R}$ is not unique.

Before closing this section we mention again that if the scalar field is real, then a Hermitian endomorphism is symmetric, and a unitary endomorphism is orthogonal. The reader may rephrase the theorems in this and other sections for the real case according to this simple rule.

Exercises

27.1 Show that Theorem 27.5 remains valid if $\mathbf{A}$ is unitary or skew-Hermitian rather than Hermitian.

27.2 If $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ is Hermitian, show that $\mathbf{A}$ is positive semidefinite if and only if the fundamental invariants of $\mathbf{A}$ are nonnegative.

27.3 Given an endomorphism $\mathbf{A}$, the *exponential* of $\mathbf{A}$ is an endomorphism $\exp\mathbf{A}$ defined by the series

$$\exp \mathbf{A} = \sum_{j=1}^{\infty} \frac{1}{j!} \mathbf{A}^j \tag{27.28}$$

It is possible to show that this series converges in a definite sense for all $\mathbf{A}$. We shall examine this question in Section 65. Show that if $\mathbf{A} \in \mathscr{L}(\mathscr{V}; \mathscr{V})$ is a Hermitian endomorphism given by the representation (27.12), then

$$\exp \mathbf{A} = \sum_{j=1}^{L} e^{\lambda_j} \mathbf{P}_j \tag{27.29}$$

This result shows that the series representation of $\exp \mathbf{A}$ is consistent with (27.18).

27.4 Suppose that $\mathbf{A}$ is a positive-definite Hermitian endomorphism; give a definition for $\log \mathbf{A}$ by power series and one by a formula similar to (27.18), then show that the two definitions are consistent. Further, prove that

$$\log \exp \mathbf{A} = \mathbf{A}$$

27.5 For any $\mathbf{A} \in \mathscr{L}(\mathscr{V}; \mathscr{V})$ show that $\mathbf{A}^* \mathbf{A}$ and $\mathbf{A}\mathbf{A}^*$ are Hermitian and positive semidefinite. Also, show that $\mathbf{A}$ is regular if and only if $\mathbf{A}^* \mathbf{A}$ and $\mathbf{A}\mathbf{A}^*$ are positive definite.

27.6 Show that $\mathbf{A}^k = \mathbf{B}^k$ where k is a positive integer, generally does not imply $\mathbf{A} = \mathbf{B}$ even if both $\mathbf{A}$ and $\mathbf{B}$ are Hermitian.

Section 28. Illustrative Examples

In this section we shall illustrate certain of the results of the preceding section by working selected numerical examples. For simplicity, the basis of $\mathscr{V}$ shall be taken to be the orthonormal basis $\{\mathbf{i}_1,...,\mathbf{i}_N\}$ introduced in (13.1). The vector equation (25.1) takes the component form

$$A_{kj}v_j = \lambda v_k \tag{28.1}$$

where

$$\mathbf{v} = v_j\mathbf{i}_j \tag{28.2}$$

and

$$\mathbf{A}\mathbf{i}_j = A_{kj}\mathbf{i}_k \tag{28.3}$$

Since the basis is orthonormal, we have written all indices as subscripts, and the summation convention is applied in the usual way.

Example 1. Consider a real three-dimensional vector space $\mathscr{V}$. Let the matrix of an endomorphism $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ be

$$M(\mathbf{A}) = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix} \tag{28.4}$$

Clearly $\mathbf{A}$ is symmetric and, thus, the theorems of the preceding section can be applied. By direct expansion of the determinant of $M(\mathbf{A}-\lambda\mathbf{I})$ the characteristic polynomial is

$$f(\lambda) = (-\lambda)(1-\lambda)(3-\lambda) \tag{28.5}$$

Therefore the three eigenvalues of $\mathbf{A}$ are distinct and are given by

$$\lambda_1 = 0,\ \lambda_2 = 1,\ \lambda_3 = 3 \tag{28.6}$$

The ordering of the eigenvalues is not important. Since the eigenvalues are distinct, their corresponding characteristic subspaces are one-dimensional. For definiteness, let $\mathbf{v}^{(p)}$ be an eigenvector associated with λ_p. As usual, we can represent $\mathbf{v}^{(p)}$ by

$$\mathbf{v}^{(p)} = v_k^{(p)}\mathbf{i}_k \tag{28.7}$$

Then (28.1), for $p = 1$, reduces to

$$v_1^{(1)} + v_2^{(1)} = 0, \quad v_1^{(1)} + 2v_2^{(1)} + v_3^{(1)} = 0, \quad v_2^{(1)} + v_3^{(1)} = 0 \tag{28.8}$$

The general solution of this linear system is

$$v_1^{(1)} = t, \qquad v_2^{(1)} = -t, \qquad v_3^{(1)} = t \tag{28.9}$$

for all $t \in \mathscr{R}$. In particular, if $\mathbf{v}^{(1)}$ is required to be a unit vector, then we can choose $t = \pm 1/\sqrt{3}$, where the choice of sign is arbitrary, say

$$v_1^{(1)} = 1/\sqrt{3}, \qquad v_2^{(1)} = -1/\sqrt{3}, \quad v_3^{(1)} = 1/\sqrt{3} \tag{28.10}$$

So

$$\mathbf{v}^{(1)} = (1/\sqrt{3})\mathbf{i}_1 - (1/\sqrt{3})\mathbf{i}_2 + (1/\sqrt{3})\mathbf{i}_3 \tag{28.11}$$

Likewise we find for $p = 2$

$$\mathbf{v}^{(2)} = (1/\sqrt{2})\mathbf{i}_1 - (1/\sqrt{2})\mathbf{i}_3 \tag{28.12}$$

and for $p = 3$

$$\mathbf{v}^{(3)} = (1/\sqrt{6})\mathbf{i}_1 + (2/\sqrt{6})\mathbf{i}_2 + (1/\sqrt{6})\mathbf{i}_3 \tag{28.13}$$

It is easy to check that $\{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \mathbf{v}^{(3)}\}$ is an orthonormal basis.

By (28.6) and (27.12), $\mathbf{A}$ has the spectral decomposition

$$\mathbf{A} = 0\mathbf{P}_1 + 1\mathbf{P}_2 + 3\mathbf{P}_3 \tag{28.14}$$

where $\mathbf{P}_k$ is the perpendicular projection defined by

$$\mathbf{P}_k \mathbf{v}^{(k)} = \mathbf{v}^{(k)}, \quad \mathbf{P}_k \mathbf{v}^{(j)} = \mathbf{0}, \qquad j \neq k \tag{28.15}$$

for $k = 1, 2, 3$. In component form relative to the original orthonormal basis $\{\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3\}$ these projections are given by

$$\mathbf{P}_k \mathbf{i}_j = v_j^{(k)} v_l^{(k)} \mathbf{i}_i \qquad \text{(no sum on } k\text{)} \tag{28.16}$$

This result follows from (28.15) and the transformation law (22.7) or directly from the representations

$$\begin{aligned}
\mathbf{i}_1 &= (1/\sqrt{3})\mathbf{v}^{(1)} + (1/2)\mathbf{v}^{(2)} + (1/\sqrt{6})\mathbf{v}^{(3)} \\
\mathbf{i}_2 &= -(1/\sqrt{3})\mathbf{v}^{(1)} + (2/\sqrt{6})\mathbf{v}^{(3)} \\
\mathbf{i}_3 &= (1/\sqrt{3})\mathbf{v}^{(1)} - (1/2)\mathbf{v}^{(2)} + (1/\sqrt{6})\mathbf{v}^{(3)}
\end{aligned} \tag{28.17}$$

since the coefficient matrix of $\{\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3\}$ relative to $\{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \mathbf{v}^{(3)}\}$ is the transpose of that of $\{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \mathbf{v}^{(3)}\}$ relative to $\{\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3\}$.

There is a result, known as Sylvester's Theorem, which enables one to compute the projections directly. We shall not prove this theorem here, but we shall state the formula in the case *when the eigenvalues are distinct*. The result is

$$\mathbf{P}_j = \frac{\coprod_{k=1;\, j\neq k}^{N} (\lambda_k \mathbf{I} - \mathbf{A})}{\coprod_{k=1;\, j\neq k}^{N} (\lambda_k - \lambda_j)} \tag{28.18}$$

The advantage of this formula is that one does not need to know the eigenvectors in order to find the projections. With respect to an arbitrary basis, (28.18) yields

$$M(\mathbf{P}_j) = \frac{\coprod_{k=1;\, j\neq k}^{N} (\lambda_k M(\mathbf{I}) - M(\mathbf{A}))}{\coprod_{k=1;\, j\neq k}^{N} (\lambda_k - \lambda_j)} \tag{28.19}$$

Example 2. To illustrate (28.19), let

$$M(\mathbf{A}) = \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix} \tag{28.20}$$

with respect to an orthonormal basis. The eigenvalues of this matrix are easily found to be $\lambda_1 = -2, \lambda_2 = 3$. Then, (28.19) yields

$$M(\mathbf{P}_1) = \left(\lambda_2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix} \right) \Big/ (\lambda_2 - \lambda_1) = \frac{1}{5} \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix}$$

and

$$M(\mathbf{P}_2) = \left(\lambda_1 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 2 & 2 \\ 2 & -1 \end{bmatrix} \right) \Big/ (\lambda_2 - \lambda_1) = \frac{1}{5} \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$$

The spectral theorem for this linear transformation yields the matrix equation

$$M(\mathbf{A}) = -\frac{2}{5} \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix} + \frac{3}{5} \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$$

Exercises

28.1 The matrix of $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ with respect to an orthonormal basis is

$$M(\mathbf{A}) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

(a) Express $M(\mathbf{A})$ in its spectral form.

(b) Express $M(\mathbf{A}^{-1})$ in its spectral form.

(c) Find the matrix of the square root of $\mathbf{A}$.

28.2 The matrix of $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ with respect to an orthonormal basis is

$$M(\mathbf{A}) = \begin{bmatrix} 3 & 0 & 2 \\ 0 & 2 & 0 \\ 1 & 0 & 3 \end{bmatrix}$$

Find an orthonormal basis for $\mathscr{V}$ relative to which the matrix of $\mathbf{A}$ is diagonal.

28.3 If $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ is defined by

$$\mathbf{A}\mathbf{i}_1 = 2\sqrt{2}\mathbf{i}_1 - 2\mathbf{i}_2$$

$$\mathbf{A}\mathbf{i}_2 = (\frac{3}{2}\sqrt{2} + 1)\mathbf{i}_1 + (3 - \frac{1}{2}\sqrt{2})\mathbf{i}_2$$

and

$$\mathbf{A}\mathbf{i}_3 = \mathbf{i}_3$$

where $\{\mathbf{i}_1, \mathbf{i}_2, \mathbf{i}_3\}$ is an orthonormal basis for $\mathscr{V}$, determine the linear transformations $\mathbf{R}$, $\mathbf{U}$, and $\mathbf{V}$ in the polar decomposition theorem.

Section 29 The Minimal Polynomial

In Section 26 we remarked that for any given endomorphism $\mathbf{A}$ there exist some polynomials $f(t)$ such that

$$f(\mathbf{A}) = \mathbf{0} \tag{29.1}$$

For example, by the Cayley-Hamilton theorem, we can always choose, $f(t)$ to be the characteristic polynomial of $\mathbf{A}$. Another obvious choice can be found by observing the fact that since $\mathscr{L}(\mathscr{V};\mathscr{V})$ has dimension N^2, the set

$$\left\{\mathbf{A}^0 = \mathbf{I}, \mathbf{A}^1, \mathbf{A}^2, ..., \mathbf{A}^{N^2}\right\}$$

is necessarily linearly dependent and, thus there exist scalars $\left\{\alpha_0, \alpha_1, ..., \alpha_{N^2}\right\}$, not all equal to zero, such that

$$\alpha_0\mathbf{I} + \alpha_1\mathbf{A}^1 + \cdots + \alpha_{N^2}\mathbf{A}^{N^2} = \mathbf{0} \tag{29.2}$$

For definiteness, let us denote the set of all polynomials f satisfying the condition (29.1) for a given $\mathbf{A}$ by the symbol $\mathscr{P}(\mathbf{A})$. We shall now show that $\mathscr{P}(\mathbf{A})$ has a very simple structure, called a *principal ideal.*

In general, an ideal $\mathscr{I}$ is a subset of an integral domain $\mathscr{D}$ such that the following two conditions are satisfied:

1. If f and g belong to $\mathscr{I}$, so is their sum $f + g$.
2. If f belongs to $\mathscr{I}$ and h arbitrary element of $\mathscr{D}$, then $fh = hf$ also belong to $\mathscr{I}$.

We recall that the definition of an integral domain is given in Section 7. Of course, $\mathscr{D}$ itself and the subset $\{0\}$ consisting in the zero element of $\mathscr{D}$ are obvious examples of ideals, and these are called *trivial ideals* or *improper ideals.* Another example of ideal is the subset $\mathscr{I} \subset \mathscr{D}$ consisting in all multiples of a particular element $g \in \mathscr{D}$, namely

$$\mathscr{I} = \{hg, h \in \mathscr{D}\} \tag{29.3}$$

It is easy to verily that this subset satisfies the two conditions for an ideal. Ideals of the special form (29.3) are called *principal ideals*.

For the set $\mathscr{P}(\mathbf{A})$ we choose the integral domain $\mathscr{D}$ to be the set of all polynomials with complex coefficients. (For real vector space, the coefficients are required to be real, of course.) Then it is obvious that $\mathscr{P}(\mathbf{A})$ is an ideal in $\mathscr{D}$, since if f and g satisfy the condition (29.1), so does their sum $f + g$ and similarly if f satisfies (29.1) and h is an arbitrary polynomial, then

$$(hf)(\mathbf{A}) = h(\mathbf{A})\, f(\mathbf{A}) = h(\mathbf{A})\mathbf{0} = \mathbf{0} \tag{29.4}$$

The fact that $\mathscr{P}(\mathbf{A})$ is actually a principal ideal is a standard result in algebra, since we have the following theorem.

Theorem 29.1. Every ideal of the polynomial domain is a principal ideal.

Proof. We assume that the reader is simi1iar with the operation of division for polynomials. If f and $g \neq 0$ are polynomials, we can divide f by g and obtain a remainder r having degree less than g, namely

$$r(t) = f(t) - h(t)g(t) \tag{29.5}$$

Now, to prove that $\mathscr{P}(\mathbf{A})$ can be represented by the form (29.3), we choose a polynomial $g \neq 0$ having the lowest degree in $\mathscr{P}(\mathbf{A})$. Then we claim that

$$\mathscr{P}(\mathbf{A}) = \{hg, h \in \mathscr{D}\} \tag{29.6}$$

To see this, we must show that every $f \in \mathscr{P}(\mathbf{A})$ can be devided through by g without a remainder. Suppose that the division of f by g yields a remainder r as shown in (29.5). Then since $\mathscr{P}(\mathbf{A})$ is an ideal and since $f, g \in \mathscr{P}(\mathbf{A})$, (29.5) shows that $r \in \mathscr{P}(\mathbf{A})$ also. But since the degree of r is less than the degree of $g, r \in \mathscr{P}(\mathbf{A})$ is possible if and only if $r = 0$. Thus $f = hg$, so the representation (29.6) is valid.

A corollary of the preceding theorem is the fact that the nonzero polynomial g having the lowest degree in $\mathscr{P}(\mathbf{A})$ is unique to within an arbitrary nonzero multiple of a scalar. If we require the leading coefficient of g to be 1, then g becomes unique, and we call this particular polynomial g the *minimal polynomial* of the endomorphism $\mathbf{A}$.

We pause here to give some examples of minimal polynomials.

Example 1. The minimal polynomial of the zero endomorphism $\mathbf{0}$ is the polynomial $f(t) = 1$ of zero degree, since by convention

$$f(\mathbf{0}) = 1\mathbf{0}^0 = \mathbf{0} \tag{29.7}$$

In general, if $\mathbf{A} \neq \mathbf{0}$, then the minimal polynomial of $\mathbf{A}$ is at least of degree 1, since in this case

$$1\mathbf{A}^0 = 1\mathbf{I} \neq \mathbf{0} \tag{29.8}$$

Example 2. Let $\mathbf{P}$ be a nontrivial projection. Then the minimal polynomial g of $\mathbf{P}$ is

$$g(t) = t^2 - t \tag{29.9}$$

For, by the definition of a projection,

$$\mathbf{P}^2 - \mathbf{P} = \mathbf{P}(\mathbf{P} - \mathbf{I}) = \mathbf{0} \tag{29.10}$$

and since $\mathbf{P}$ is assumed to be non trivial, the two lower degree divisors t and $t-1$ no longer satisfy the condition (29.1) for $\mathbf{P}$.

Example 3. For the endomorphism $\mathbf{C}$ whose matrix is given by (26.19) in Exercise 26.8, the minimal polynomial g is

$$g(t) = t^N \tag{29.11}$$

since we have seen in that exercise that

$$\mathbf{C}^N = \mathbf{0} \qquad \text{but} \qquad \mathbf{C}^{N-1} \neq \mathbf{0}$$

For the proof of some theorems in the next section we need several other standard results in the algebra of polynomials. We summarize these results here.

Theorem 29.2. If f and g are polynomials, then there exists a *greatest common divisor* d which is a divisor (i.e., a factor) of f and g and is also a multiple of every common divisor of f and g.

Proof. We define the ideal $\mathscr{I}$ in the polynomial domain $\mathscr{D}$ by

$$\mathscr{I} \equiv \{hf + kg,\ h,k \in \mathscr{D}\} \tag{29.12}$$

By Theorem 29.1, $\mathscr{I}$ is a principal ideal, and thus it has a representation

$$\mathscr{I} \equiv \{hd,\ h \in \mathscr{D}\} \tag{29.13}$$

We claim that d is a greatest common divisor of f and g. Clearly, d is a common divisor of f and g, since f and g are themselves members of $\mathscr{I}$, so by (29.13) there exist h and k in $\mathscr{D}$ such that

$$f = hd, \qquad g = kd \tag{29.14}$$

On the other hand, since d is also a member of $\mathscr{I}$, by (29.12) there exist also p and q in $\mathscr{D}$ such that

$$d = pf + qg \tag{29.15}$$

Therefore if c is any common divisor of f and g, say

$$f = ac, \qquad g = bc \tag{29.16}$$

then from (29.15)

$$d = (pa + qb)c \tag{29.17}$$

so d is a multiple of c. Thus d is a greatest common divisor of f and g.

By the same argument as before, we see that the greatest common divisor d of f and g is unique to within a nonzero scalar factor. So we can render d unique by requiring its leading coefficient to be 1. Also, it is clear that the preceding theorem can be extended in an obvious way to more than two polynomials. If the greatest common divisor of $f_1,..., f_L$ is the zero degree polynomial 1, then $f_1,..., f_L$ are said to be *relatively prime*. Similarly $f_1,..., f_L$ are *pairwise prime* if each pair $f_i, f_j, i \neq j$, from $f_1,..., f_L$ is relatively prime.

Another important concept associated with the algebra of polynomials is the concept of the *least common multiple*.

Theorem 29.3. If f and g are polynomials, then there exists a *least common multiple* m which is a multiple of f and g and is a divisor of every common multiple of f and g.

The proof of this theorem is based on the same argument as the proof of the preceding theorem, so it is left as an exercise.

Exercises

29.1 If the eigenvalues of an endomorphism $\mathbf{A}$ are all single roots of the characteristic equation, show that the characteristic polynomial of $\mathbf{A}$ is also a minimal polynomial of $\mathbf{A}$.

29.2 Prove Theorem 29.3.

29.3 If f and g are nonzero polynomials and if d is their greatest common divisor, show that then

$$m = fg / d$$

is their least common multiple and, conversely, if m is their least common multiplies, then

$$d = fg / m$$

is their greatest common divisor.

Section 30. Spectral Decomposition for Arbitrary Endomorphisms

As we have mentioned in Section 27, not all endomorphims have matrices which are diagonal. However, for Hermitian endomorphisms, a decomposition of the endomorphism into a linear combination of projections is possible and is given by (27.12). In this section we shall consider the problem in general and we shall find decompositions which are, in some sense, closest to the simple decomposition (27.12) for endomorphisms in general.

We shall prove some preliminary theorems first. In Section 24 we remarked that the null space $K(\mathbf{A})$ is always $\mathbf{A}$-invariant. This result can be generalized to the following

Theorem 30.1. If . f is any polynomial, then the null space $K(f(\mathbf{A}))$ is $\mathbf{A}$-invariant.

Proof. Since the multiplication of polynomials is a commutative operation, we have

$$\mathbf{A}f(\mathbf{A}) = f(\mathbf{A})\mathbf{A} \tag{30.1}$$

Hence if $\mathbf{v} \in K(f(\mathbf{A}))$, then

$$f(\mathbf{A})\mathbf{A}\mathbf{v} = \mathbf{A}f(\mathbf{A})\mathbf{v} = \mathbf{A}\mathbf{0} = \mathbf{0} \tag{30.2}$$

which shows that $\mathbf{A}\mathbf{v} \in K(f(\mathbf{A}))$. Therefore $K(f(\mathbf{A}))$ is $\mathbf{A}$-invariant.

Next we prove some theorems which describe the dependence of $K(f(\mathbf{A}))$ on the choice of f.

Theorem 30.2. If f is a multiple of g, say

$$f = hg \tag{30.3}$$

Then

$$K(f(\mathbf{A})) \supset K(g(\mathbf{A})) \tag{30.4}$$

Proof. This result is a general property of the null space. Indeed, for any endomorphisms $\mathbf{B}$ and $\mathbf{C}$ we always have

$$K(\mathbf{BC}) \subset K(\mathbf{C}) \tag{30.5}$$

So if we set $h(\mathbf{A}) = \mathbf{B}$ and $g(\mathbf{A}) = \mathbf{C}$, then (30.5) reduces to (30.4).

The preceding theorem does not imply that $K(g(\mathbf{A}))$ is necessarily a proper subspace of $K(f(\mathbf{A}))$, however. It is quite possible that the two subspaces, in fact, coincide. For example, if g and hence f both belong to $\mathscr{P}(\mathbf{A})$, then $g(\mathbf{A}) = f(\mathbf{A}) = \mathbf{0}$, and thus

$$K(f(\mathbf{A})) = K(g(\mathbf{A})) = K(\mathbf{0}) = \mathscr{V} \tag{30.6}$$

However, if m is the minimal polynomial of $\mathbf{A}$, and if f is a proper divisor of m (i.e., m is not a divisor of f) so that $f \notin \mathscr{P}(\mathbf{A})$, then $K(f(\mathbf{A}))$ is strictly a proper subspace of $\mathscr{V} = K(m(\mathbf{A}))$. We can strengthen this result to the following

Theorem 30.3. If f is a divisor (proper or improper) of the minimal polynomial m of $\mathbf{A}$, and if g is a proper divisor of f, then $K(g(\mathbf{A}))$ is strictly a proper subspace of $K(f(\mathbf{A}))$.

Proof. By assumption there exists a polynomial h such that

$$m = hf \tag{30.7}$$

We set

$$k = hg \tag{30.8}$$

Then k is a proper divisor of m, since by assumption, g is a proper divisor of f. By the remark preceding the theorem, $K(k(\mathbf{A}))$ is strictly a subspace of $\mathscr{V}$, which is equal to $K(m(\mathbf{A}))$. Thus there exists a vector $\mathbf{v}$ such that

$$k(\mathbf{A})\mathbf{v} = g(\mathbf{A})h(\mathbf{A})\mathbf{v} \neq \mathbf{0} \tag{30.9}$$

which implies that the vector $\mathbf{u} = h(\mathbf{A})\mathbf{v}$ does not belong to $K(g(\mathbf{A}))$. On the other hand, from (30.7)

$$f(\mathbf{A})\mathbf{u} == f(\mathbf{A})h(\mathbf{A})\mathbf{v} = m(\mathbf{A})\mathbf{v} = \mathbf{0} \tag{30.10}$$

which implies that $\mathbf{u}$ belongs to $K(f(\mathbf{A}))$. Thus $K(g(\mathbf{A}))$ is strictly a proper subspace of $K(f(\mathbf{A}))$.

The next theorem shows the role of the greatest common divisor in terms of the null space.

Theorem 30.4. Let f and g be any polynomials, and suppose that d is their greatest common divisor. Then

$$K(d(\mathbf{A})) = K(f(\mathbf{A})) \cap K(g(\mathbf{A})) \tag{30.11}$$

Obviously, this result can be generalized for more than two polynomials.

Proof. Since d is a common divisor of f and g, the inclusion

$$K(d(\mathbf{A})) \subset K(f(\mathbf{A})) \cap K(g(\mathbf{A})) \tag{30.12}$$

follows readily from Theorem 30.2. To prove the reversed inclusion,

$$K(d(\mathbf{A})) \supset K(f(\mathbf{A})) \cap K(g(\mathbf{A})) \tag{30.13}$$

recall that from (29.15) there exist polynomials p and q such that

$$d = pf + qg \tag{30.14}$$

and thus

$$d(\mathbf{A}) = p(\mathbf{A})f(\mathbf{A}) + q(\mathbf{A})g(\mathbf{A}) \tag{30.15}$$

This equation means that if $\mathbf{v} \in K(f(\mathbf{A})) \cap K(g(\mathbf{A}))$, then

$$\begin{aligned} d(\mathbf{A})\mathbf{v} &= p(\mathbf{A})f(\mathbf{A})\mathbf{v} + q(\mathbf{A})g(\mathbf{A})\mathbf{v} \\ &= p(\mathbf{A})\mathbf{0} + q(\mathbf{A})\mathbf{0} = \mathbf{0} \end{aligned} \tag{30.16}$$

so that $\mathbf{v} \in K(d(\mathbf{A}))$. Therefore (30.13) is valid and hence (30.11).

A corollary of the preceding theorem is the fact that if f and g are relatively prime then

$$K(f(\mathbf{A})) \cap K(g(\mathbf{A})) = \{\mathbf{0}\} \tag{30.17}$$

since in this case the greatest common divisor of f and g is $d(t) = 1$, so that

$$K(d(\mathbf{A})) = K(\mathbf{A}^0) = K(\mathbf{I}) = \{\mathbf{0}\} \tag{30.18}$$

Here we have assumed $\mathbf{A} \neq \mathbf{0}$ of course.

Next we consider the role of the least common multiple in terms of the null space.

Theorem 30.5. Let f and g be any polynomials, and suppose that l is their least common multiplier. Then

$$K(l(\mathbf{A})) = K(f(\mathbf{A})) + K(g(\mathbf{A})) \tag{30.19}$$

where the operation on the right-hand side of (30.19) is the sum of subspaces defined in Section 10. Like the result (30.11), the result (30.19) can be generalized in an obvious way for more than two polynomials.

The proof of this theorem is based on the same argument as the proof of the preceding theorem, so it is left as an exercise. As in the preceding theorem, a corollary of this theorem is that if f and g are relatively prime (pairwise prime if there are more than two polynomials) then (30.19) can be strengthened to

$$K(l(\mathbf{A})) = K(f(\mathbf{A})) \oplus K(g(\mathbf{A})) \tag{30.20}$$

Again, we have assumed that $\mathbf{A} \neq \mathbf{0}$. We leave the proof of (30.20) also as an exercise.

Having summarized the preliminary theorems, we are now ready to state the main theorem of this section.

Theorem 30.6. If m is the minimal polynomial of $\mathbf{A}$ which is factored into the form

$$m(t) = (t-\lambda_1)^{a_1} \cdots (t-\lambda_L)^{a_L} \equiv m_1(t) \cdots m_L(t) \tag{30.21}$$

where $\lambda_1, ..., \lambda_L$ are distinct and $a_1, ..., a_L$ are positive integers, then $\mathscr{V}$ has the representation

$$\mathscr{V} = K(m_1(\mathbf{A})) \oplus \cdots \oplus K(m_L(\mathbf{A})) \tag{30.22}$$

Proof. Since $\lambda_1, ..., \lambda_L$ are distinct, the polynomials $m_1, ..., m_L$ are pairwise prime and their least common multiplier is m. Hence by (30.20) we have

$$K(m(\mathbf{A})) = K(m_1(\mathbf{A})) \oplus \cdots \oplus K(m_L(\mathbf{A})) \tag{30.23}$$

But since $m(\mathbf{A}) = 0$, $K(m(\mathbf{A})) = \mathscr{V}$, so that (30.22) holds.

Now from Theorem 30.1 we know that each subspace $K(m_i(\mathbf{A})), .i = 1, ..., L$ is $\mathbf{A}$-invariant; then from (30.22) we see that $\mathbf{A}$ is reduced by the subspaces $K(m_1(\mathbf{A})), ..., K(m_L(\mathbf{A}))$. k{ ,{))) .1 kfmL(6)). Therefore, the results of Theorem 24.1 can be applied. In particular, if we choose a basis for each $K(m_i(\mathbf{A}))$ and form a basis for $\mathscr{V}$ by the union of these bases, then the matrix of $\mathbf{A}$ takes the form (24.5) in Exercise 24.1. The next theorem shows that, in some sense, the factorization (30.21) gives also the minimal polynomials of the restrictions of $\mathbf{A}$ to the various $\mathbf{A}$-invariant subspaces from the representation (30.22).

Theorem 30.7. Each factor m_k of m is the minimal polynomial of the restriction of $\mathbf{A}$ to the subspace $K(m_k(\mathbf{A}))$. More generally, any product of factors, say $m_1 \cdots m_M$ is the minimal polynomial of the restriction of $\mathbf{A}$ to the corresponding subspace $K(m_1(\mathbf{A})) \oplus \cdots \oplus K(m_M(\mathbf{A}))$

Proof : We prove the special case of one factor only, say m_1; the proof of the general case of several factors is similar and is left as an exercise. For definiteness, let $\bar{\mathbf{A}}$ denote the restriction of $\mathbf{A}$ to $K(m_1(\mathbf{A}))$ Then $m_1(\bar{\mathbf{A}})$ is equal to the restriction of $m_1(\mathbf{A})$ to $K(m_1(\mathbf{A}))$, and, thus $m_1(\bar{\mathbf{A}}) = \mathbf{0}$, which means that $m_1 \in \mathscr{P}(\bar{\mathbf{A}})$. Now if g is any proper divisor of m_1, then $g(\bar{\mathbf{A}}) \neq \mathbf{0}$; for otherwise, we would have $g(\mathbf{A})m_2(\mathbf{A})\cdots m_L(\mathbf{A})$, contradicting the fact that m is minimal for $\mathbf{A}$. Therefore m_1 is minimal for $\bar{\mathbf{A}}$ and the proof is complete.

The form (24.5), generally, is not diagonal. However, if the minimal polynomial (30.21) has simple roots only, i.e., the powers $a_1,...,a_L$ are all equal to 1, then

$$m_i(\mathbf{A}) = \mathbf{A} - \lambda_i \mathbf{I} \tag{30.24}$$

for all i. In this case, the restriction of $\mathbf{A}$ to $K(m_i(\mathbf{A}))$ coincides with $\lambda_i \mathbf{I}$ on that subspace, namely

$$\mathbf{A}_{\mathscr{V}(\lambda_i)} = \lambda_i \mathbf{I} \tag{30.25}$$

Here we have used the fact from (30.24), that

$$\mathscr{V}(\lambda_i) = K(m_i(\mathbf{A})) \tag{30.26}$$

Then the form (24.5) reduces to the diagonal form $(27.3)_1$ or $(27.3)_2$.

The condition that m has simple roots only turns out to be necessary for the existence of a diagonal matrix for $\mathbf{A}$ also, as we shall now see in the following theorem.

Theorem 30.8. An endomorphism $\mathbf{A}$ has a diagonal matrix if and only if its minimal polynomial can be factored into distinct factors all of the first degree.

Proof. Sufficiencyhas already been proven. To prove necessity, assume that the matrix of $\mathbf{A}$ relative to some basis $\{\mathbf{e}_1,...,\mathbf{e}_N\}$ has the form $(27.3)_1$. Then the polynomial

$$m(t) = (t - \lambda_1)\cdots(t - \lambda_L) \tag{30.27}$$

is the minimal polynomial of $\mathbf{A}$. Indeed, each basis vector $\mathbf{e}_i$ is contained in the null space $K(\mathbf{A}-\lambda_k\mathbf{I})$ for one particular k. Consequently $\mathbf{e}_i \in K(m(\mathbf{A}))$ for all $i=1,...,N$, and thus

$$K(m(\mathbf{A})) = \mathscr{V} \tag{30.28}$$

or equivalently

$$m(\mathbf{A}) = \mathbf{0} \tag{30.29}$$

which implies that $m \in \mathscr{P}(\mathbf{A})$. But since $\lambda_1,,,\lambda_L$ are distinct, no proper divisor of m still belongs to $\mathscr{P}(\mathbf{A})$. Therefore m is the minimal polynomial of $\mathbf{A}$.

As before, when the condition of the preceding theorem is satisfied, then we can define projections $\mathbf{P}_1,...,\mathbf{P}_L$ by

$$R(\mathbf{P}_i) = \mathscr{V}(\lambda_i), \qquad K(\mathbf{P}_i) = \bigoplus_{\substack{j=1 \\ j\neq i}}^{L} \mathscr{V}(\lambda_j) \tag{30.30}$$

for all $i=1,...,L$, and the diagonal form $(27.3)_1$ shows that $\mathbf{A}$ has the representation

$$\mathbf{A} = \lambda_1\mathbf{P}_1 + \cdots + \lambda_L\mathbf{P}_L = \sum_{i=1}^{L} \lambda_i\mathbf{P}_i \tag{30.31}$$

It should be noted, however, that in stating this result we have not made use of any inner product, so it is not meaningful to say whether or not the projections $\mathbf{P}_1,...,\mathbf{P}_L$ are perpendicular; further, the eigenvalues $\lambda_1,...,\lambda_L$ are generally complex numbers. In fact, the factorization (30.21) for m in general is possible only if the scalar field is algebraically closed, such as the complex field used here. If the scalar field is the real field, we should define the factors $m_1,...,m_L$ of m to be powers of *irreducible polynomials*, i.e., polynomials having no proper divisors. Then the decomposition (30.22) for $\mathscr{V}$ remains valid, since the argument of the proof is based entirely on the fact that the factors of m are pairwise prime.

Theorem 30.8 shows that in order to know whether or not $\mathbf{A}$ has a diagonal form, we must know the roots and their multiplicities in the minimal polynomial m of $\mathbf{A}$. Now since the

characteristic polynomial f of $\mathbf{A}$ belongs to $\mathscr{P}(\mathbf{A})$, m is a divisor of f. Hence the roots of m are always roots of f, The next theorem gives the converse of this result.

Theorem 30.9. Each eigenvalue of $\mathbf{A}$ is a root of the minimal polynomial m of $\mathbf{A}$ and vice versa.

Proof. Sufficiency has already been proved. To prove necessity, let λ be an eigenvalue of $\mathbf{A}$. Then we wish to show that the polynomial

$$g(t) = t - \lambda \tag{30.32}$$

is a divisor of m. Since g is of the first degree, if g is not a divisor of m, then m and g are relatively prime. By (30.17) and the fact that $m \in \mathscr{P}(\mathbf{A})$, we have

$$\{\mathbf{0}\} = K(m(\mathbf{A})) \cap K(g(\mathbf{A})) = \mathscr{V} \cap K(g(\mathbf{A})) = K(g(\mathbf{A})) \tag{30.33}$$

But this is impossible, since $K(g(\mathbf{A}))$, being the characteristic subspace corresponding to the eigenvalue λ, cannot be of zero dimension.

The preceding theorem justifies our using the same notations $\lambda_1,...,\lambda_L$ for the (distinct) roots of the minimal polynomial m, as shown in (30.21). However it should be noted that the root λ_i generally has a smaller multiplicity in m than in f, because m is a divisor of f. The characteristic polynomial f yields not only the (distinct) roots $\lambda_1,...,\lambda_L$ of m, it determines also the dimensions of their corresponding subspaces $K(m_1(\mathbf{A})),...,K(m_L(\mathbf{A}))$ in the decomposition (30.22) This result is made explicit in the following theorem.

Theorem 30.10. Let d_k denote the algebraic multiplicity of the eigenvalue λ_k as before; i.e., d_k is the multiplicity of λ_k in f [cf. (26.8)]. Then we have

$$\dim K(m_k(\mathbf{A})) = d_k \tag{30.34}$$

Proof We prove this result by induction. Clearly, it is valid for all $\mathbf{A}$ having a diagonal form, since in this case $m_k(\mathbf{A}) = \mathbf{A} - \lambda_k \mathbf{I}$, so that $K(m_k(\mathbf{A}))$ is the characteristic subspace corresponding to λ_k and its dimension is the geometric multiplicity as well as the algebraic multiplicity of λ_k. Now assuming that the result is valid for all $\mathbf{A}$ whose minimal polynomial has at most M multiple roots, where $M = 0$ is the starting induction hypothesis, we wish to show that the same holds for

all $\mathbf{A}$ whose minimal polynomial has $M+1$ multiple roots. To see this, we make use of the decomposition (30.23) and, for definiteness, we assume that λ_1 is a multiple root of m. We put

$$\mathscr{U} = K(m_2(\mathbf{A})) \oplus \cdots \oplus K(m_L(\mathbf{A})) \tag{30.35}$$

Then $\mathscr{U}$ is $\mathbf{A}$=invariant. From Theorem 30.7 we know that the minimal polynomial $m_{\mathscr{U}}$ of the restriction $\mathbf{A}_{\mathscr{U}}$ is

$$m_{\mathscr{U}} = m_2 \cdots m_L \tag{30.36}$$

which has at most M multiple roots. Hence by the induction hypothesis we have

$$\dim \mathscr{U} = \dim K(m_2(\mathbf{A})) + \cdots + \dim K(m_L(\mathbf{A})) = d_2 + \cdots + d_L \tag{30.37}$$

But from (26.8) and (30.23) we have also

$$\begin{aligned} N &= d_1 + d_2 + \cdots + d_L \\ N &= \dim K(m_1(\mathbf{A})) + \dim K(m_2(\mathbf{A})) + \cdots + \dim K(m_L(\mathbf{A})) \end{aligned} \tag{30.38}$$

Comparing (30.38) with (30.37), we see that

$$d_1 = \dim K(m_1(\mathbf{A})) \tag{30.39}$$

Thus the result (30.34) is valid for all $\mathbf{A}$ whose minimal polynomial has $M+1$ multiple roots, and hence in general.

An immediate consequence of the preceding theorem is the following.

Theorem 30.11. Let b_k be the geometric multiplicity of λ_k, namely

$$b_k \equiv \dim \mathscr{V}(\lambda_k) \equiv \dim K(g_k(\mathbf{A})) \equiv \dim K(\mathbf{A} - \lambda_k \mathbf{I}) \tag{30.40}$$

Then we have

$$1 \le b_k \le d_k - a_k + 1 \tag{30.41}$$

where d_k is the algebraic multiplicity of λ_k and a_k is the multiplicity of λ_k in m ,as shown in (30.21). Further, $b_k = d_k$ if and only if

$$K(g_k(\mathbf{A})) = K(g_k^2(\mathbf{A})) \tag{30.42}$$

Proof. If λ_k is a simple root of m , i.e., $a_k = 1$ and $g_k = m_k$, then from (30.34) and (30.40) we have $b_k = d_k$. On the other hand, if λ_k is a multiple root of m , i.e., $a_k > 1$ and $m_k = g_k^{a_k}$, then the polynomials $g_k, g_k^2, ..., g_k^{(a_k - 1)}$ are proper divisors of m_k . Hence by Theorem 30.3

$$\mathscr{V}(\lambda_k) = K(g_k(\mathbf{A})) \subset K(g_k^2(\mathbf{A})) \subset \cdots \subset K(g_k^{(a_k-1)}(\mathbf{A})) \subset K(m_k(\mathbf{A})) \tag{30.43}$$

where the inclusions are strictly proper and the dimensions of the subspaces change by at least one in each inclusion. Thus (30.41) holds.

The second part of the theorem can be proved as follows: If λ_k is a simple root of m , then $m_k = g_k$, and, thus $K(g_k(\mathbf{A})) = K(g_k^2(\mathbf{A})) = \mathscr{V}(\lambda_k)$. On the other hand, if λ_k is not a simple root of m , then m_k is at least of second degree. In this case g_k and g_k^2 are both divisors of m . But since g_k is also a proper divisor of g_k^2 , by Theorem 30.3, $K(g_k(\mathbf{A}))$ is strictly a proper subspace of $K(g_k^2(\mathbf{A}))$, so that (30.42) cannot hold, and the proof is complete.

The preceding three theorems show that for each eigenvalue λ_k of $\mathbf{A}$, generally there are two nonzero $\mathbf{A}$ -invariant subspaces, namely, the eigenspace $\mathscr{V}(\lambda_k)$ and the subspace $K(m_k(\mathbf{A}))$. For definiteness, let us call the latter subspace the *characteristic subspace* corresponding to λ_k and denote it ‘by the more compact notation $\mathscr{U}(\lambda_k)$. Then $\mathscr{V}(\lambda_k)$ is a subspace of $\mathscr{U}(\lambda_k)$ in general, and the two subspaces coincide if and only if λ_k is a simple root of m . Since λ_k is the only eigenvalue of the restriction of $\mathbf{A}$ to $\mathscr{U}(\lambda_k)$, by the Cayley-Hamilton theorem we have also

$$\mathscr{U}(\lambda_k) = K((\mathbf{A} - \lambda_k \mathbf{I})^{d_k}) \tag{30.44}$$

where d_k is the algebraic multiplicity of λ_k, which is also the dimension of $\mathscr{U}(\lambda_k)$. Thus we can determine the characteristic subspace directly from the characteristic polynomial of $\mathbf{A}$ by (30.44).

Now if we define $\mathbf{P}_k$ to be the projection on $\mathscr{U}(\lambda_k)$ in the direction of the remaining $\mathscr{U}(\lambda_j)$, $j \neq k$, namely

$$R(\mathbf{P}_k) = \mathscr{U}(\lambda_k), K(\mathbf{P}_k) = \bigoplus_{\substack{j=1 \\ j\neq k}}^{L} \mathscr{U}(\lambda_j) \tag{30.45}$$

and we define $\mathbf{B}_k$ to be $\mathbf{A} - \lambda_k \mathbf{P}_k$ on $\mathscr{U}(\lambda_k)$ and $\mathbf{0}$ on $\mathscr{U}(\lambda_j)$, $j \neq k$, then $\mathbf{A}$ has the spectral decomposition by a direct sum

$$\mathbf{A} = \sum_{j=1}^{L} (\lambda_j \mathbf{P}_j + \mathbf{B}_j) \tag{30.46}$$

where

$$\left.\begin{aligned} \mathbf{P}_j^2 &= \mathbf{P}_j \\ \mathbf{P}_j \mathbf{B}_j &= \mathbf{B}_j \mathbf{P}_j = \mathbf{B}_j \\ \mathbf{B}_j^{a_j} &= \mathbf{0}, \quad 1 \leq a_j \leq d_j \end{aligned}\right\} \qquad j = 1, \ldots, L \tag{30.47}$$

$$\left.\begin{aligned} \mathbf{P}_j \mathbf{P}_k &= \mathbf{0} \\ \mathbf{P}_j \mathbf{B}_k &= \mathbf{B}_k \mathbf{P}_j = \mathbf{0} \\ \mathbf{B}_j \mathbf{B}_k &= \mathbf{0}, \end{aligned}\right\} \qquad j \neq k, \quad j,k = 1, \ldots, L \tag{30.48}$$

In general, an endomorphism $\mathbf{B}$ satisfying the condition

$$\mathbf{B}^a = \mathbf{0} \tag{30.49}$$

for some power a is called *nilpotent*. From (30.49) or from Theorem 30.9 the only eigenvalue of a nilpotent endomorphism is 0, and the lowest power a satisfying (30.49) is an integer $a, 1 \leq a \leq N$, such that t^N is the characteristic polynomial of $\mathbf{B}$ and t^a is the minimal polynomial of $\mathbf{B}$. In view of (30.47) we see that each endomorphism $\mathbf{B}_j$ in the decomposition (30.46) is nilpotent and can be

regarded also as a nilpotent endomorphism on $\mathscr{U}(\lambda_j)$. In order to decompose $\mathbf{A}$ further from (30.46) we must determine a spectral decomposition for each $\mathbf{B}_j$. This problem is solved in general as follows.

First, we recall that in Exercise 26.8 we have defined a nilcylic endomorphism $\mathbf{C}$ to be a nilpotent endomorphism such that

$$\mathbf{C}^N = \mathbf{0} \qquad \text{but} \qquad \mathbf{C}^{N-1} \neq \mathbf{0} \tag{30.50}$$

Where N is the dimension of the underlying vector space $\mathscr{V}$. For such an endomorphism we can find a cyclic basis $\{\mathbf{e}_1,...,\mathbf{e}_N\}$ which satisfies the conditions

$$\mathbf{C}^{N-1}\mathbf{e}_1 = \mathbf{0}, \qquad \mathbf{C}^{N-1}\mathbf{e}_2 = \mathbf{e}_1,...,\mathbf{C}\mathbf{e}_N = \mathbf{e}_{N-1} \tag{30.51}$$

or, equivalently

$$\mathbf{C}^{N-1}\mathbf{e}_N = \mathbf{e}_1, \qquad \mathbf{C}^{N-2}\mathbf{e}_2 = \mathbf{e}_2,...,\mathbf{C}\mathbf{e}_N = \mathbf{e}_{N-1} \tag{30.52}$$

so that the matrix of $\mathbf{C}$ takes the simple form (26.19). Indeed, we can choose $\mathbf{e}_N$ to be any vector such that $\mathbf{C}^{N-1}\mathbf{e}_N \neq \mathbf{0}$; then the set $\{\mathbf{e}_1,...,\mathbf{e}_N\}$ defined by (30.52) is linearly independent and thus forms a cyclic basis for $\mathbf{C}$. Nilcyclic endomorphisms constitute only a special class of nilpotent endomorphisms, but in some sense the former can be regarded as the building blocks for the latter. The result is made precise by the following theorem.

Theorem 30.12. Let $\mathbf{B}$ be a nonzero nilpotent endomorphism of $\mathscr{V}$ in general, say $\mathbf{B}$ satisfies the conditions

$$\mathbf{B}^a = \mathbf{0} \qquad \text{but} \qquad \mathbf{B}^{a-1} \neq \mathbf{0} \tag{30.53}$$

for some integer a between 1 and N. Then there exists a direct sum decomposition for $\mathscr{V}$:

$$\mathscr{V} = \mathscr{V}_1 \oplus \cdots \oplus \mathscr{V}_M \tag{30.54}$$

and a corresponding direct sum decomposition for $\mathbf{B}$ (in the sense explained in Section 24):

$$\mathbf{B} = \mathbf{B}_1 + \cdots + \mathbf{B}_M \tag{30.55}$$

such that each $\mathbf{B}_j$ is nilpotent and its restriction to $\mathscr{V}_j$ is nilcyclic. The subspaces $\mathscr{V}_1,...,\mathscr{V}_M$ in the decomposition (30.54) are not unique, but their dimensions are unique and obey the following rules: The maximum of the dimension of $\mathscr{V}_1,...,\mathscr{V}_M$ is equal to the integer a in (30.53); the number N_a of subspaces among $\mathscr{V}_1,...,\mathscr{V}_M$ having dimension a is given by

$$N_a = N - \dim K(\mathbf{B}^{a-1}) \tag{30.56}$$

More generally, the number N_b of subspaces among $\mathscr{V}_1,...,\mathscr{V}_M$ having dimensions greater than or equal to b is given by

$$N_b = \dim K(\mathbf{B}^b) - \dim K(\mathbf{B}^{b-1}) \tag{30.57}$$

for all $b = 1,...,a$. In particular, when $b = 1$, N_1 is equal to the integer M in (30.54), and (30.57) reduces to

$$M = \dim K(\mathbf{B}) \tag{30.58}$$

Proof. We prove the theorem by induction on the dimension of $\mathscr{V}$. Clearly, the theorem is valid for one-dimensional space since a nilpotent endomorphism there is simply the zero endomorphism which is nilcylic. Assuming now the theorem is valid for vector spaces of dimension less than or equal to $N-1$, we shall prove that the same is valid for vector spaces of dimension N.

Notice first if the integer a in (30.53) is equal to N, then $\mathbf{B}$ is nilcyclic and the assertion is trivially satisfied with $M = 1$, so we can assume that $1 < a < N$. By $(30.53)_2$, there exists a vector $\mathbf{e}_a \in \mathscr{V}$ such that $\mathbf{B}^{a-1}\mathbf{e}_a \neq \mathbf{0}$. As in (30.52) we define

$$\mathbf{B}^{a-1}\mathbf{e}_a = \mathbf{e}_1,...,\mathbf{B}\mathbf{e}_{a-1} \tag{30.59}$$

Then the set $\{\mathbf{e}_1,...\mathbf{e}_a\}$ is linearly independent. We put $\mathscr{V}_1$ to 'be the subspace generated by $\{\mathbf{e}_1,...\mathbf{e}_a\}$. Then by definition $\dim \mathscr{V}_1 = a$, and the restriction of $\mathbf{B}$ on $\mathscr{V}_1$ is nilcyclic.

Now recall that for any subspace of a vector space we can define a factor space (cf. Section 11). As usual we denote the factor space of $\mathscr{V}$ over $\mathscr{V}_1$ by $\mathscr{V}/\mathscr{V}_1$. From the result of Exercise 11.5 and (30.59), we have

$$\dim \mathscr{V}/\mathscr{V}_1 = N - a < N - 1 \tag{30.60}$$

Thus we can apply the theorem to the factor space $\mathscr{V}/\mathscr{V}_1$. For definiteness, let us use the notation of Section 11, namely, if $\mathbf{v} \in \mathscr{V}$, then $\bar{\mathbf{v}}$ denotes the equivalence set of $\mathbf{v}$ in $\mathscr{V}/\mathscr{V}_1$. This notation means that the superimposed bar is the canonical projection from $\mathscr{V}$ to $\mathscr{V}/\mathscr{V}_1$. From (30.59) it is easy to see that $\mathscr{V}_1$ is $\mathbf{B}$-invariant. Hence if $\mathbf{u}$ and $\mathbf{v}$ belong to the same equivalence set, so do $\mathbf{Bu}$ and $\mathbf{Bv}$. Therefore we can define an endomorphism $\bar{\mathbf{B}}$ on the factor space $\mathscr{V}/\mathscr{V}_1$, by

$$\bar{\mathbf{B}}\bar{\mathbf{v}} = \overline{\mathbf{Bv}} \tag{30.61}$$

for all $\mathbf{v} \in \mathscr{V}$ or equivalently for all $\bar{\mathbf{v}} \in \mathscr{V}/\mathscr{V}_1$, Applying (30.60) repeatedly, we have also

$$\bar{\mathbf{B}}^k\bar{\mathbf{v}} = \overline{\mathbf{B}^k\mathbf{v}} \tag{30.62}$$

for all integers k. In particular, $\bar{\mathbf{B}}$ is nilpotent and

$$\overline{\mathbf{B}^a} = \mathbf{0} \tag{30.63}$$

By the induction hypothesis we can then find a direct sum decomposition of the form

$$\mathscr{V}/\mathscr{V}_1 = \mathscr{U}_1 \oplus \cdots \oplus \mathscr{U}_p \tag{30.64}$$

for the factor space $\mathscr{V}/\mathscr{V}_1$ and a corresponding direct sum decomposition

$$\bar{\mathbf{B}} = \mathbf{F}_1 + \cdots \mathbf{F}_p \tag{30.65}$$

for $\bar{\mathbf{B}}$. In particular, there are cyclic bases in the subspaces $\mathscr{U}_1,...,\mathscr{U}_p$ for the nilcyclic endomorphisms which are the restrictions of $\mathbf{F}_1,,,,,\mathbf{F}_p$ to the corresponding subspaces. For definiteness, let $\{\bar{\mathbf{f}}_1,...,\bar{\mathbf{f}}_b\}$ be a cyclic basis in $\mathscr{U}_1$, say

$$\bar{\mathbf{B}}^b\bar{\mathbf{f}}_b = \mathbf{0}, \qquad \bar{\mathbf{B}}^{b-1}\bar{\mathbf{f}}_b = \mathbf{f}_1,...,\bar{\mathbf{B}}\bar{\mathbf{f}}_b = \bar{\mathbf{f}}_{b-1} \tag{30.66}$$

From (30.63), is necessarily less than or equal to a.

From $(30.66)_1$ and (30.62) we see that $\bar{\mathbf{B}}^b\bar{\mathbf{f}}_b$ belongs to $\mathscr{V}_1$ and, thus can be expressed as a linear combination of $\{\mathbf{e}_1,...,\mathbf{e}_a\}$, say

$$\bar{\mathbf{B}}^b\bar{\mathbf{f}}_b = \alpha_1\mathbf{e}_1 + \cdots + \alpha_a\mathbf{e}_a = (\alpha_1\mathbf{B}^{a-1} + \cdots + \alpha_{a-1}\mathbf{B} + \alpha_a\mathbf{I})\mathbf{e}_a \tag{30.67}$$

Now there are two possibilities: *(i)* $\bar{\mathbf{B}}^b\bar{\mathbf{f}}_b = \mathbf{0}$ or *(ii)* $\bar{\mathbf{B}}^b\bar{\mathbf{f}}_b \neq \mathbf{0}$. In case *(i)* we define as before

$$\mathbf{B}^{b-1}\mathbf{f}_b = \mathbf{f}_1,...,\mathbf{B}\mathbf{f}_b = \mathbf{f}_{b-1} \tag{30.68}$$

Then $\{\mathbf{f}_1,...,\mathbf{f}_b\}$ is a linearly independent set in $\mathscr{V}$, and we put $\mathscr{V}_2$ to be the subspace generated by $\{\mathbf{f}_1,...,\mathbf{f}_b\}$, On the other hand, in case *(ii)* from (30.53) we see that b is strictly less than a; moreover, from (30.67) we have

$$\mathbf{0} = \mathbf{B}^a\mathbf{f}_b = (\alpha_{a-b+1}\mathbf{B}^{a-1} + \cdots + \alpha_a\mathbf{B}^{a-b})\mathbf{e}_a = \alpha_{a-b+1}\mathbf{e}_1 + \cdots + \alpha_a\mathbf{e}_b \tag{30.69}$$

which implies

$$= \alpha_{a-b+1} = \cdots = \alpha_a = 0 \tag{30.70}$$

or equivalently

$$\mathbf{B}^b\mathbf{f}_b = (\alpha_1\mathbf{B}^{a-1} + \cdots + \alpha_{a-b}\mathbf{B}^b)\mathbf{e}_a \tag{30.71}$$

Hence we can choose another vector $\mathbf{f}_b{}'$ in the same equivalence set of $\mathbf{f}_b$ by

$$\mathbf{f}_b' = \mathbf{f}_b - (\alpha_1 \mathbf{B}^{a-b-1} + \cdots + \alpha_{a-b}\mathbf{I})\mathbf{e}_a = \mathbf{f}_b - \alpha_1 \mathbf{e}_{b+1} - \cdots - \alpha_{a-b}\mathbf{e}_a \tag{30.72}$$

which now obeys the condition $\mathbf{B}^b \mathbf{f}_b' = \mathbf{0}$, and we can proceed in exactly the same way as in case *(i)*. Thus in any case every cyclic basis $\{\overline{\mathbf{f}}_1, ..., \overline{\mathbf{f}}_b\}$ for $\mathscr{U}_1$ gives rise to a cyclic set $\{\mathbf{f}_1, ..., \mathbf{f}_b\}$ in $\mathscr{V}$.

Applying this result to each one of the subspaces $\mathscr{U}_1, ..., \mathscr{U}_p$ we obtain cyclic sets $\{\mathbf{f}_1, ..., \mathbf{f}_b\}$, $\{\mathbf{g}_1, ..., \mathbf{g}_c\}$,…, and subspaces $\mathscr{V}_2, ..., \mathscr{V}_{p+1}$ generated by them in $\mathscr{V}$. Now it is clear that the union of $\{\mathbf{e}_1, ..., \mathbf{e}_a\}$, $\{\mathbf{f}_1, ..., \mathbf{f}_b\}, \{\mathbf{g}_1, ..., \mathbf{g}_c\}$,…, form a basis of $\mathscr{V}$ since from (30.59), (30.60), and (30.64) there are precisely N vectors in the union; further, if we have

$$\alpha_1 \mathbf{e}_1 + \cdots + \alpha_a \mathbf{e}_a + \beta_1 \mathbf{f}_1 + \cdots + \beta_b \mathbf{f}_b + \gamma_1 \mathbf{g}_1 + \cdots + \gamma_c \mathbf{g}_c + \cdots = \mathbf{0} \tag{30.73}$$

then taking the canonical projection to $\mathscr{V}/\mathscr{V}_1$ yields

$$\beta_1 \overline{\mathbf{f}}_1 + \cdots + \beta_b \overline{\mathbf{f}}_b + \gamma_1 \overline{\mathbf{g}}_1 + \cdots + \gamma_c \overline{\mathbf{g}}_c + \cdots = \mathbf{0}$$

which implies

$$\beta_1 = \cdots = \beta_b = \gamma_1 = \cdots = \gamma_c = \cdots = \mathbf{0}$$

and substituting this result back into (30.73) yields

$$\alpha_1 = \cdots = \alpha_a = 0$$

Thus $\mathscr{V}$ has a direct sum decomposition given by (30.54) with $M = p + 1$ and $\mathbf{B}$ has a corresponding decomposition given by (30.55) where $\mathbf{B}_1, ..., \mathbf{B}_M$ have the prescribed properties.

Now the only assertion yet to be proved is equation (30.57). This result follows from the general rule that for any nilcylic endomorphism $\mathbf{C}$ on a L-dimensional space we have

$$\dim K(\mathbf{C}^k) - \dim(\mathbf{C}^{k-1}) = \begin{cases} 1 & \text{for } 1 \le k \le L \\ 0 & \text{for } k > L \end{cases}$$

Applying this rule to the restriction of $\mathbf{B}_j$ to $\mathscr{V}$ for all $j = 1,...,M$ and using the fact that the kernel of $\mathbf{B}^k$ is equal to the direct sum of the kernel of the restriction of $(\mathbf{B}_j)^k$ for all $j = 1,...,M$, prove easily that (30.57) holds. Thus the proof is complete.

In general, we cannot expect the subspaces $\mathscr{V}_1,...,\mathscr{V}_M$ in the decomposition (30.54) to be unique. Indeed, if there are two subspaces among $\mathscr{V}_1,...,\mathscr{V}_M$ having the same dimension, say $\dim \mathscr{V}_1 = \dim \mathscr{V}_2 = a$, then we can decompose the direct sum $\mathscr{V}_1 \oplus \mathscr{V}_2$ in many other ways, e.g.,

$$\mathscr{V}_1 \oplus \mathscr{V}_2 = \overline{\mathscr{V}}_1 \oplus \overline{\mathscr{V}}_2 \tag{30.74}$$

and when we substitute (30.74) into (30.54) the new decomposition

$$\mathscr{V} = \overline{\mathscr{V}}_1 \oplus \overline{\mathscr{V}}_2 \oplus \mathscr{V}_3 \oplus \cdots \oplus \mathscr{V}_N$$

possesses exactly the same properties as the original decomposition (30.54). For instance we can define $\mathscr{V}_1$ and $\mathscr{V}_2$ to be the subspaces generated by the linearly independent cyclic set $\{\tilde{\mathbf{e}}_1,...,\tilde{\mathbf{e}}_a\}$ and $\{\tilde{\mathbf{f}}_1,...,\tilde{\mathbf{f}}_a\}$, where we choose the starting vectors $\tilde{\mathbf{e}}_a$ and $\tilde{\mathbf{f}}_a$ by

$$\tilde{\mathbf{e}}_a = \alpha \mathbf{e}_a + \beta \mathbf{f}_a, \qquad \tilde{\mathbf{f}}_a = \gamma \mathbf{e}_a + \delta \mathbf{f}_a$$

provided that the coefficient matrix on the right hand side is nonsingular.

If we apply the preceding theorem to the restriction of $\mathbf{A} - \lambda_k \mathbf{I}$ on $\mathscr{U}_k$, we see that the inequality (30.41) is the best possible one in general. Indeed, $(30.41)_2$ becomes an equality if and only if $\mathscr{U}_k$ has the decomposition

$$\mathscr{U}_k = \mathscr{U}_{k1} \oplus \cdots \oplus \mathscr{U}_{kM} \tag{30.75}$$

where the dimensions of the subspaces $\mathscr{U}_{k1},...,\mathscr{U}_{kM}$ are

$$\dim \mathcal{U}_{k1} = a_k, \quad \dim \mathcal{U}_{k2} = \cdots = \dim \mathcal{U}_{kM} = 1$$

If there are more than one subspaces among $\mathcal{U}_{k1},...,\mathcal{U}_{kM}$ having dimension greater than one, then $(30.41)_2$ is a strict inequality.

The matrix of the restriction of $\mathbf{A} - \lambda_k \mathbf{I}$ to $\mathcal{U}_k$ relative to the union of the cyclic basis for $\mathcal{U}_{k1},...,\mathcal{U}_{kM}$ has the form

$$A_k = \begin{bmatrix} A_{k1} & & & 0 \\ & A_{k2} & & \\ & & \ddots & \\ 0 & & & A_{kM} \end{bmatrix} \tag{30.76}$$

where each submatrix A_{kj} in the diagonal of A_k has the form

$$A_{kj} = \begin{bmatrix} \lambda_k & 1 & & 0 \\ & \lambda_k & 1 & \\ & & \ddots & \\ & & & 1 \\ 0 & & & \lambda_k \end{bmatrix} \tag{30.77}$$

Substituting (30.76) into (24.5) yields the *Jordan normal form* for $\mathbf{A}$:

$$M(\mathbf{A}) = \begin{bmatrix} A_{11} & & & 0 \\ & A_{12} & & \\ & & \ddots & \\ 0 & & & A_M \end{bmatrix} \tag{30.78}$$

The Jordan normal form is an important result since it gives a geometric interpretation of an arbitrary endomorphism of a vector space. In general, we say that two endomorphisms $\mathbf{A}$ and $\mathbf{A}'$ are *similar* if the matrix of $\mathbf{A}$ relative to a basis is identical to the matrix of $\mathbf{A}'$ relative to another basis. From the transformation law (22.7), we see that $\mathbf{A}$ and $\mathbf{A}'$ are similar if and only if there exists a nonsingular endomorphism $\mathbf{T}$ such that

$$\mathbf{A}' = \mathbf{T}\mathbf{A}\mathbf{T}^{-1} \tag{30.79}$$

Clearly, (30.79) defines an equivalence relation on $\mathscr{L}(\mathscr{V};\mathscr{V})$. We call the equivalence sets relative to (30.79) the *conjugate subsets* of $\mathscr{L}(\mathscr{V};\mathscr{V})$. Now for each $\mathbf{A} \in \mathscr{L}(\mathscr{V};\mathscr{V})$ the Jordan normal form of $\mathbf{A}$ is a particular matrix of $\mathbf{A}$ and is unique to within an arbitrary change of ordering of the various square blocks on the diagonal of the matrix. Hence $\mathbf{A}$ and $\mathbf{A}'$ are similar if and only if they have the same Jordan normal form. Thus the Jordan normal form characterizes the conjugate subsets of $\mathscr{L}(\mathscr{V};\mathscr{V})$

Exercises

30.1 Prove Theorem 30.5.

30.2 Prove the general case of Theorem 30.7.

30.3 Let $\mathbf{U}$ be an unitary endomorphism of an inner product space $\mathscr{V}$. Show that $\mathbf{U}$ has a diagonal form.

30.4 If $\mathscr{V}$ is a real inner product space, show that an orthogonal endomorphism $\mathbf{Q}$ in general does not have a diagonal form, but it has the spectral form

$$M(\mathbf{Q}) = \begin{bmatrix} 1 & & & & & & & & & & \\ & \ddots & & & & & & & & & \\ & & 1 & & & & & & 0 & & \\ & & & -1 & & & & & & & \\ & & & & \ddots & & & & & & \\ & & & & & \ddots & & & & & \\ & & & & & & \ddots & & & & \\ & & & & & & & -1 & & & \\ & & & & & & & & \cos\theta_1 & -\sin\theta_1 & \\ & & 0 & & & & & & \sin\theta_1 & \cos\theta_1 & \\ & & & & & & & & & \ddots & \\ & & & & & & & & & & \ddots \\ & & & & & & & & & & & \cos\theta_L & -\sin\theta_L \\ & & & & & & & & & & & \sin\theta_L & \cos\theta_L \end{bmatrix}$$

where the angles $\theta_1, ..., \theta_L$ may or may not be distinct.

30.5 Determine whether the endomorphism $\mathbf{A}$ whose matrix relative to a certain basis is

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

can have a diagonal matrix relative to another basis. Does the result depend on whether the scalar field is real or complex?

30.6 Determine the Jordan normal form for the endomorphism $\mathbf{A}$ whose matrix relative to a certain basis is

$$M(\mathbf{A}) = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Chapter 7

TENSOR ALGEBRA

The concept of a tensor is of major importance in applied mathematics. Virtually ever discipline in the physical sciences makes some use of tensors. Admittedly, one does not always need to spend a lot of time and effort to gain a computational facility with tensors as they are used in the applications. However, we take the position that a better understanding of tensors is obtained if we follow a systematic approach, using the language of finite-dimensional vector spaces. We begin with a brief discussion of linear functions on a vector space. Since in the applications the scalar field is usually the real field, from now on we shall consider real vector spaces only.

Section 31. Linear Functions, the Dual Space

Let $\mathscr{V}$ be a real vector space of dimension N. We consider the space of linear functions $\mathscr{L}(\mathscr{V};\mathscr{R})$ from $\mathscr{V}$ into the real numbers $\mathscr{R}$. By Theorem 16.1, $\mathscr{L}(\mathscr{V};\mathscr{R}) = \dim \mathscr{V} = N$. Thus $\mathscr{V}$ and $\mathscr{L}(\mathscr{V};\mathscr{R})$ are isomorphic. We call $\mathscr{L}(\mathscr{V};\mathscr{R})$ the *dual Space* of $\mathscr{V}$, and we denote it by the special notation $\mathscr{V}^*$. To distinguish elements of $\mathscr{V}$ from those of $\mathscr{V}^*$, we shall call the former elements *vectors* and the latter elements *covectors*. However, these two names are strictly relative to each other. Since $\mathscr{V}^*$ is a N-dimensional vector space by itself, we can apply any result valid for a vector space in general to $\mathscr{V}^*$ as well as to $\mathscr{V}$. In fact, we can even define a dual space $(\mathscr{V}^*)^*$ for $\mathscr{V}^*$ just as we define a dual space $\mathscr{V}^*$ for $\mathscr{V}$. In order not to introduce too many new concepts at the same time, we shall postpone the second dual space $(\mathscr{V}^*)^*$ until the next section. Hence in this section $\mathscr{V}$ shall be a given N-dimensional space and $\mathscr{V}^*$ shall denote its dual space. As usual, we denote typical elements of $\mathscr{V}$ by $\mathbf{u}, \mathbf{v}, \mathbf{w}, \ldots$. Then the typical elements of $\mathscr{V}^*$ are denoted by $\mathbf{u}^*, \mathbf{v}^*, \mathbf{w}^*, \ldots$. However, it should be noted that the asterisk here is strictly a convenient notation, not a symbol for a function from $\mathscr{V}$ to $\mathscr{V}^*$. Thus $\mathbf{u}^*$ is not related in any particular way to $\mathbf{u}$. Also, for some covectors, such as those that constitute a dual basis to be defined shortly, this mutation becomes rather cumbersome. In such cases, the notation is simply abandoned. For instance, without fear of ambiguity we denote the null covector in $\mathscr{V}^*$ by the same notation as the null vector in $\mathscr{V}$, namely $\mathbf{0}$, instead of $\mathbf{0}^*$.

If $\mathbf{v}^* \in \mathscr{V}^*$, then $\mathbf{v}^*$ is a linear function from $\mathscr{V}$ to $\mathscr{R}$, i.e.,

$$\mathbf{v}^* : \mathscr{V} \to \mathscr{R}$$

such that for any vectors $\mathbf{u}, \mathbf{v}, \in \mathscr{V}$ and scalars $\alpha, \beta, \in \mathscr{R}$

$$\mathbf{v}^*(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha\mathbf{v}^*(\mathbf{u}) + \beta\mathbf{v}^*(\mathbf{v})$$

Of course, the linear operations on the right hand side are those of $\mathscr{R}$ while those on the left hand side are linear operations in $\mathscr{V}$. For a reason the will become apparent later, it is more convenient to denote the value of $\mathbf{v}^*$ at $\mathbf{v}$ by the notation $\langle \mathbf{v}^*, \mathbf{v} \rangle$. Then the bracket $\langle\ ,\ \rangle$ operation can be viewed as a function

$$\langle\ ,\ \rangle : \mathscr{V}^* \times \mathscr{V} \to \mathscr{R}$$

It is easy to verify that this operation has the following properties:

$$\begin{aligned}
&(i) \quad \langle \alpha\mathbf{v}^* + \beta\mathbf{u}^*, \mathbf{v} \rangle = \alpha \langle \mathbf{v}^*, \mathbf{v} \rangle + \beta \langle \mathbf{u}^*, \mathbf{v} \rangle \\
&(ii) \quad \langle \mathbf{v}^*, \alpha\mathbf{u} + \beta\mathbf{v} \rangle = \alpha \langle \mathbf{v}^*, \mathbf{u} \rangle + \beta \langle \mathbf{v}^*, \mathbf{v} \rangle \\
&(iii) \text{ For any given } \mathbf{v}, \langle \mathbf{v}^*, \mathbf{v} \rangle \text{ vanishes for all } \mathbf{v} \in \mathscr{V} \text{ if and only if } \mathbf{v}^* = \mathbf{0}. \\
&(iv) \text{ Similarly, for any given } \mathbf{v}, \langle \mathbf{v}^*, \mathbf{v} \rangle \text{ vanishes for all } \mathbf{v}^* \in \mathscr{V}^* \text{ if and only if } \mathbf{v} = \mathbf{0}.
\end{aligned} \tag{31.1}$$

The first two properties define $\langle , \rangle$ to be a *bilinear* operation on $\mathscr{V}^* \times \mathscr{V}$, and the last two properties define $\langle , \rangle$ to be a *definite* operation. These properties resemble the properties of an inner product, so that we call the operation $\langle , \rangle$ the *scalar product*. As we shall see, we can define many concepts associated with the scalar product similar to corresponding concepts associated with an inner product. The first example is the concept of the *dual basis*, which is the counterpart of the concept of the reciprocal basis.

If $\{\mathbf{e}_1, \ldots \mathbf{e}_N\}$ is a basis for $\mathscr{V}$, we define it *dual basis* to be a basis $\{\mathbf{e}^1, \ldots \mathbf{e}^N\}$ for $\mathscr{V}^*$ such that

$$\langle \mathbf{e}^j, \mathbf{e}_i \rangle = \delta_i^j \tag{31.2}$$

for all $i, j = 1, \ldots, N$. The reader should compare this condition with the condition (14.1) that defines the reciprocal basis. By exactly the same argument as before we can prove the following theorem.

Theorem 31.1. The dual basis relative to a given basis exists and it unique.

Notice that we have dropped the asterisk notation for the covector $\mathbf{e}^j$ in a dual basis; the superscript alone is enough to distinguish $\{\mathbf{e}^1, \ldots \mathbf{e}^N\}$ from $\{\mathbf{e}_1, \ldots \mathbf{e}_N\}$. However, it should be kept in mind that, unlike the reciprocal basis, the dual basis is a basis for $\mathscr{V}^*$, not a basis for $\mathscr{V}$. In particular, it makes no sense to require a basis be the same as its dual basis. This means the component form of a vector $\mathbf{v} \in \mathscr{V}$ relative to a basis $\{\mathbf{e}_1, \ldots \mathbf{e}_N\}$,

$$\mathbf{v} = v^i \mathbf{e}_i \tag{31.3}$$

must never be confused with the component form of a covector $\mathbf{v}^* \in \mathscr{V}^*$ relative to the dual basis,

$$\mathbf{v}^* = v_i \mathbf{e}^i \tag{31.4}$$

In order to emphasize the difference of these two component forms, we call v^i the *contravariant components* of $\mathbf{v}$ and v_i the *covariant components* of $\mathbf{v}^*$. A vector has contravariant components only and a covector has covariant components only. The terminology for the components is not inconsistent with the same terminology defined earlier for an inner product space, since we have the following theorem.

Theorem 31.2. Given any inner product on $\mathscr{V}$, there exists a unique isomorphism

$$\mathbf{G}: \ \mathscr{V} \to \mathscr{V}^* \tag{31.5}$$

which is induced by the inner product in such a way that

$$\langle \mathbf{G}\mathbf{v}, \mathbf{w} \rangle = \mathbf{v} \cdot \mathbf{w}, \qquad \mathbf{v}, \mathbf{w} \in \mathscr{V} \tag{31.6}$$

Under this isomorphism the image of any orthonormal basis $\{\mathbf{i}_1,\ldots,\mathbf{i}_N\}$ is the dual basis $\{\mathbf{i}^1,\ldots,\mathbf{i}^N\}$, namely

$$\mathbf{G}\mathbf{i}_k = \mathbf{i}^k, \qquad k = 1,\ldots,N \tag{31.7}$$

and, more generally, if $\{\mathbf{e}_1,\ldots\mathbf{e}_N\}$ is an arbitrary basis, then the image of its reciprocal basis $\{\overline{\mathbf{e}}^1,\ldots\overline{\mathbf{e}}^N\}$, is the dual basis $\{\mathbf{e}^1,\ldots\mathbf{e}^N\}$, namely

$$\mathbf{G}\overline{\mathbf{e}}^k = \mathbf{e}^k, \qquad k = 1,\ldots,N \tag{31.8}$$

Proof. Since we now consider only real vector spaces and real inner product spaces, the right-hand side of (31.6), clearly, is a linear function of $\mathbf{w}$ for each $\mathbf{v} \in \mathscr{V}$. Thus $\mathbf{G}$ is well defined by the condition (31.6). We must show that $\mathbf{G}$ is an isomorphism. The fact that $\mathbf{G}$ is a linear transformation is obvious, since the right-hand side of (31.6) is linear in $\mathbf{v}$ for each $\mathbf{w} \in \mathscr{V}$. Also, $\mathbf{G}$ is one-to-one because, from (31.6), if $\mathbf{G}\mathbf{u} = \mathbf{G}\mathbf{v}$, then $\mathbf{u}\cdot\mathbf{w} = \mathbf{v}\cdot\mathbf{w}$ for all $\mathbf{w}$ and thus $\mathbf{u} = \mathbf{v}$. Now since we already know that $\dim\mathscr{V} = \dim\mathscr{V}^*$, any one-to-one linear transformation from $\mathscr{V}$ to $\mathscr{V}^*$ is necessarily onto and hence an isomorphism. The proof of (31.8) is obvious, since by the definition of the reciprocal basis we have

$$\overline{\mathbf{e}}^i \cdot \mathbf{e}_j = \delta^i_j, \qquad i, j = 1,\ldots,N$$

and by the definition of the dual basis we have

$$\langle \mathbf{e}^i, \mathbf{e}_j \rangle = \delta^i_j, \qquad i, j = 1,\ldots,N$$

Comparing these definitions with (31.6), we obtain

$$\langle \mathbf{G}\overline{\mathbf{e}}^i = \mathbf{e}_j \rangle = \langle \mathbf{e}^i, \mathbf{e}_j \rangle, \qquad i, j = 1,\ldots,N$$

which implies (31.8) because $\{\mathbf{e}_1,\ldots\mathbf{e}_N\}$ is a basis of $\mathscr{V}$.

Because of this theorem, if a particular inner product is assigned on $\mathscr{V}$, then we can identify $\mathscr{V}$ with $\mathscr{V}^*$ by suppressing the notation for the isomorphisms $\mathbf{G}$ and $\mathbf{G}^{-1}$. In other words, we regard a vector $\mathbf{v}$ also as a linear function on $\mathscr{V}$:

$$\langle \mathbf{v}, \mathbf{w} \rangle \equiv \mathbf{v} \cdot \mathbf{w} \tag{31.9}$$

According to this rule the reciprocal basis is identified with the duel basis and the inner product becomes the scalar product. However, since a vector space can be equipped with many inner products, unless a particular inner product is chosen, we cannot identify $\mathscr{V}$ with $\mathscr{V}^*$ in general. In this section, we shall not assign any particular inner product in $\mathscr{V}$, so $\mathscr{V}$ and $\mathscr{V}^*$ are different vector spaces.

We shall now derive some formulas which generalize the results of an inner product space to a vector space in general. First, if $\mathbf{v} \in \mathscr{V}$ and $\mathbf{v}^* \in \mathscr{V}^*$ are arbitrary, then their scalar products $\langle \mathbf{v}^*, \mathbf{v} \rangle$ can be computed in component form as follows: Choose a basis $\{\mathbf{e}_i\}$ and its dual basis $\{\mathbf{e}^i\}$ for $\mathscr{V}$ and $\mathscr{V}^*$, respectively, so that we can express $\mathbf{v}$ and $\mathbf{v}^*$ in component form (31.3) and (31.4). Then from (31.1) and (31.2) we have

$$\langle \mathbf{v}^*, \mathbf{v} \rangle = \langle v_i \mathbf{e}^i, v^j \mathbf{e}_j \rangle = v_i v^j \langle \mathbf{e}^i, \mathbf{e}_j \rangle = v_i v^j \delta^i_j = v_i v^i \tag{31.10}$$

which generalizes the formula (14.16). Applying (31.10) to $\mathbf{v}^* = \mathbf{e}^i$, we obtain

$$\langle \mathbf{e}^i, \mathbf{v} \rangle = v^i \tag{31.11}$$

which generalizes the formula $(14.15)_1$; similarly applying (31.10) to $\mathbf{v} = \mathbf{e}_j$, we obtain

$$\langle \mathbf{v}^*, \mathbf{e}_i \rangle = v_i \tag{31.12}$$

which generalizes the formula $(14.15)_2$

Next recall that for inner product spaces $\mathscr{V}$ and $\mathscr{U}$ we define the adjoint $\mathbf{A}^*$ of a linear transformation $\mathbf{A}:\mathscr{V}\to\mathscr{U}$ to be a linear transformation $\mathbf{A}^*:\mathscr{U}\to\mathscr{V}$ such that the following condition [cf.(18.1)] is satisfied:

$$\mathbf{u}\cdot\mathbf{A}\mathbf{v}=\mathbf{A}^*\mathbf{u}\cdot\mathbf{v},\qquad \mathbf{u}\in\mathscr{U},\quad \mathbf{v}\in\mathscr{V}$$

If we do not make use of any inner product, we simply replace this condition by

$$\left\langle \mathbf{u}^*,\mathbf{A}\mathbf{v}\right\rangle=\left\langle \mathbf{A}^*\mathbf{u}^*,\mathbf{v}\right\rangle,\qquad \mathbf{u}^*\in\mathscr{U}^*,\quad \mathbf{v}\in\mathscr{V} \tag{31.13}$$

then $\mathbf{A}^*$ is a linear transformation from $\mathscr{U}^*$ to $\mathscr{V}^*$,

$$\mathbf{A}^*:\mathscr{U}^*\to\mathscr{V}^*$$

and is called the *dual* of $\mathbf{A}$. By the same argument as before we can prove the following theorem.

Theorem 31.3. For every linear transformation $\mathbf{A}:\mathscr{V}\to\mathscr{U}$ there exists a unique dual $\mathbf{A}^*:\mathscr{U}^*\to\mathscr{V}^*$ satisfying the condition (31.13).

If we choose a basis $\{\mathbf{e}_1,\ldots\mathbf{e}_N\}$ for $\mathscr{V}$ and a basis $\{\mathbf{b}_1,\ldots,\mathbf{b}_M\}$ for $\mathscr{U}$ and express the linear transformation $\mathbf{A}$ by (18.2) and the linear transformation $\mathbf{A}^*$ by (18.3), where $\{\mathbf{b}^\alpha\}$ and $\{\mathbf{e}^k\}$ are now regarded as the dual bases of $\{\mathbf{b}_\alpha\}$ and $\{\mathbf{e}_k\}$, respectively, then (18.4) remains valid in the more general context, except hat we now have

$$A^*{}_k{}^\alpha=A^\alpha{}_k \tag{31.14}$$

since we no longer consider complex spaces. Of course, the formulas (18.5) and (18.6) are now replace by

$$\left\langle \mathbf{b}^\alpha,\mathbf{A}\mathbf{e}_k\right\rangle=A^\alpha{}_k \tag{31.15}$$

and

$$\left\langle \mathbf{A}^* \mathbf{b}^\alpha, \mathbf{e}_k \right\rangle = A^*{}_k{}^\alpha \tag{31.16}$$

respectively.

For an inner product space the orthogonal complement of a subspace $\mathscr{U}$ of $\mathscr{V}$ is a subspace $\mathscr{U}^\perp$ given by [cf.(13.2)]

$$\mathscr{U}^\perp = \left\{ \mathbf{v} \middle| \mathbf{u} \cdot \mathbf{v} = 0 \quad \text{for all } \mathbf{u} \in \mathscr{U} \right\}$$

By the same token, if $\mathscr{V}$ is a vector space in general, then we define the *orthogonal complement* of $\mathscr{U}$ to be the subspace $\mathscr{U}^\perp$ of $\mathscr{V}^*$ given by

$$\mathscr{U}^\perp = \left\{ \mathbf{v}^* \middle| \left\langle \mathbf{v}^*, \mathbf{u} \right\rangle = 0 \quad \text{for all } \mathbf{u} \in \mathscr{U} \right\} \tag{31.17}$$

In general if $\mathbf{v} \in \mathscr{V}$ and $\mathbf{v}^* \in \mathscr{V}^*$ are arbitrary, then $\mathbf{v}$ and $\mathbf{v}^*$ are said to be *orthogonal* to each other if $\left\langle \mathbf{v}^*, \mathbf{v} \right\rangle = 0$. We can prove the following theorem by the same argument used previously for inner product spaces.

Theorem 31.4. *If $\mathscr{U}$ is a subspace of $\mathscr{V}$, then*

$$\dim \mathscr{U} + \dim \mathscr{U}^\perp = \dim \mathscr{V} \tag{31.18}$$

However, $\mathscr{V}$ is no longer the direct sum of $\mathscr{U}$ and $\mathscr{U}^\perp$ since $\mathscr{U}^\perp$ is a subspace of $\mathscr{V}^*$, not a subspace of $\mathscr{V}$.

Using the same line of reasoning, we can generalize Theorem 18.3 to the following.

Theorem 31.5. If $\mathbf{A} : \mathscr{V} \to \mathscr{U}$ is a linear transformation, then

$$K\left(\mathbf{A}^*\right) = R\left(\mathbf{A}\right)^{\perp} \tag{31.19}$$

and

$$K\left(\mathbf{A}\right) = R\left(\mathbf{A}^*\right)^{\perp} \tag{31.20}$$

Similarly, we can generalize Theorem 18.4 to the following.

Theorem 31.6. Any linear transformation $\mathbf{A}$ and its dual $\mathbf{A}^*$ have the same rank.

Finally, formulas for transferring from one basis to another basis can be generalized from inner product spaces to vector spaces in general. If $\{\mathbf{e}_1,\ldots\mathbf{e}_N\}$ and $\{\hat{\mathbf{e}}_1,\ldots\hat{\mathbf{e}}_N\}$ are bases for $\mathscr{V}$, then as before we can express one basis in component form relative to another, as shown by (14.17) and (14.18). Now suppose that $\{\mathbf{e}^1,\ldots\mathbf{e}^N\}$ and $\{\hat{\mathbf{e}}^1,\ldots\hat{\mathbf{e}}^N\}$ are the dual bases of $\{\mathbf{e}_1,\ldots\mathbf{e}_N\}$ and $\{\hat{\mathbf{e}}_1,\ldots\hat{\mathbf{e}}_N\}$, respectively. Then it can be verified easily that

$$\hat{\mathbf{e}}^q = \hat{T}_k^q\mathbf{e}^k, \qquad \mathbf{e}^q = T_k^q\hat{\mathbf{e}}^k \tag{31.21}$$

where T_k^q and $\hat{T}_k^q$ are defined by (14.17) and (14.18), i.e.,

$$\mathbf{e}_k = \hat{T}_k^q\hat{\mathbf{e}}_q, \qquad \hat{\mathbf{e}}_k = T_k^q\mathbf{e}_q \tag{31.22}$$

From these relations if $\mathbf{v} \in \mathscr{V}$ and $\mathbf{v}^* \in \mathscr{V}^*$ have the component forms (31.3) and (31.4) relative to $\{\mathbf{e}_i\}$ and $\{\mathbf{e}^i\}$ and the component forms

$$\mathbf{v} = \hat{\upsilon}^i\hat{\mathbf{e}}_i \tag{31.23}$$

and

$$\mathbf{v}^* = \hat{\upsilon}_i\hat{\mathbf{e}}^i \tag{31.24}$$

relative to $\{\hat{\mathbf{e}}_i\}$ and $\{\hat{\mathbf{e}}^i\}$, respectively, then we have the following transformation laws:

$$\hat{\upsilon}_q = T_q^k\upsilon_k \tag{31.25}$$

$$\hat{\upsilon}^q = \hat{T}^q_k \upsilon^k \tag{31.26}$$

$$\hat{\upsilon}^k = T^k_q \hat{\upsilon}^q \tag{31.27}$$

and

$$\upsilon_k = \hat{T}^q_k \hat{\upsilon}_q \tag{31.28}$$

which generalize the formulas (14.24)-(14.27), respectively.

Exercises

31.1 If $\mathbf{A}: \mathscr{V} \to \mathscr{U}$ and $\mathbf{B}: \mathscr{U} \to \mathscr{W}$ are linear transformations, show that

$$\left(\mathbf{BA}\right)^* = \mathbf{A}^*\mathbf{B}^*$$

31.2 If $\mathbf{A}: \mathscr{V} \to \mathscr{U}$ and $\mathbf{B}: \mathscr{V} \to \mathscr{U}$ are linear transformations, show that

$$\left(\alpha\mathbf{A} + \beta\mathbf{B}\right)^* = \alpha\mathbf{A}^* + \beta\mathbf{B}^*$$

and that

$$\mathbf{A}^* = \mathbf{0} \Leftrightarrow \mathbf{A} = \mathbf{0}$$

These two conditions mean that the operation of taking the dual is an isomorphism $*: \mathscr{L}\left(\mathscr{V};\mathscr{U}\right) \to \mathscr{L}\left(\mathscr{U}^*;\mathscr{V}^*\right)$.

31.3 $\mathbf{A}: \mathscr{V} \to \mathscr{U}$ is an isomorphism, show that $\mathbf{A}^*: \mathscr{U}^* \to \mathscr{V}^*$ is also an isomorphism; moreover,

$$\left(\mathbf{A}^{-1}\right)^* = \left(\mathbf{A}^*\right)^{-1}$$

31.4 If $\mathscr{V}$ has the decomposition $\mathscr{V} = \mathscr{U}_1 \oplus \mathscr{U}_2$ and if $\mathbf{P}: \mathscr{V} \to \mathscr{V}$ is the projection of $\mathscr{V}$ on $\mathscr{U}_1$ along $\mathscr{U}_2$, show that $\mathscr{V}^*$ has the decomposition $\mathscr{V}^* = \mathscr{U}_1^\perp \oplus \mathscr{U}_2^\perp$ and that $\mathbf{P}^*: \mathscr{V}^* \to \mathscr{V}^*$ is the projection of $\mathscr{V}^*$ on $\mathscr{U}_2^\perp$ along $\mathscr{U}_1^\perp$.

31.5 If $\mathscr{U}_1$ and $\mathscr{U}_2$ are subspaces of $\mathscr{V}$, show that

$$\left(\mathscr{U}_1 + \mathscr{U}_2\right)^\perp = \mathscr{U}_1^\perp \cap \mathscr{U}_2^\perp \qquad \textit{and} \qquad \left(\mathscr{U}_1 \cap \mathscr{U}_2\right)^\perp = \mathscr{U}_1^\perp + \mathscr{U}_2^\perp$$

31.6 Show that the linear transformation $\mathbf{G}: \mathscr{V} \to \mathscr{V}^*$ defined in Theorem 31.2 obeys the condition

$$\langle \mathbf{G}\mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{G}\mathbf{w}, \mathbf{v} \rangle$$

31.7 Show that an inner product on $\mathscr{V}^*$ is induced by an inner product on $\mathscr{V}$ by the formula

$$\mathbf{v}^* \cdot \mathbf{w}^* \equiv \mathbf{G}^{-1}\mathbf{v}^* \cdot \mathbf{G}^{-1}\mathbf{w}^* \tag{31.29}$$

where $\mathbf{G}$ is the isomorphism defined in Theorem 31.2.

31.8 If $\{\mathbf{e}_i\}$ is a basis for $\mathscr{V}$ and $\{\mathbf{e}^i\}$ is its dual basis in $\mathscr{V}^*$, show that $\{\mathbf{G}\mathbf{e}_i\}$ is the reciprocal basis of $\{\mathbf{e}^i\}$ with respect to the inner product on $\mathscr{V}^*$ defined by (31.29) in the preceding exercise.

Section 32. The Second Dual space, canonical Isomorphisms

In the preceding section we defined the dual space $\mathscr{V}^*$ of any vector space $\mathscr{V}$ to be the space of linear functions $\mathscr{L}(\mathscr{V};\mathscr{R})$ from $\mathscr{V}$ to $\mathscr{R}$. By the same procedure we can define the dual space $(\mathscr{V}^*)^*$ of $\mathscr{V}^*$ by

$$(\mathscr{V}^*)^* = \mathscr{L}(\mathscr{V}^*;\mathscr{R}) = \mathscr{L}(\mathscr{L}(\mathscr{V};\mathscr{R});\mathscr{R}) \tag{32.1}$$

For simplicity let us denote this space by $\mathscr{V}^{**}$, called the *second dual space* of $\mathscr{V}$. Of course, the dimension of $\mathscr{V}^{**}$ is the same as that of $\mathscr{V}$, namely

$$\dim\mathscr{V} = \dim\mathscr{V}^* = \dim\mathscr{V}^{**} \tag{32.2}$$

Using the system of notation introduced in the preceding section, we write a typical element of $\mathscr{V}^{**}$ by $\mathbf{v}^{**}$. Then $\mathbf{v}^{**}$ is a linear function on $\mathscr{V}^*$

$$\mathbf{v}^{**} : \mathscr{V}^* \to \mathscr{R}$$

Further, for each $\mathbf{v}^* \in \mathscr{V}^*$ we denote the value of $\mathscr{V}^{**}$ at $\mathbf{v}^*$ by $\langle \mathbf{v}^{**}, \mathbf{v}^* \rangle$. The $\langle , \rangle$ operation is now a mapping from $\mathscr{V}^{**} \times \mathscr{V}^*$ to $\mathscr{R}$ and possesses the same four properties given in the preceding section.

Unlike the dual space $\mathscr{V}^*$, the second dual space $\mathscr{V}^{**}$ can always be identified as $\mathscr{V}$ without using any inner product. The isomorphism

$$\mathbf{J} : \mathscr{V} \to \mathscr{V}^{**} \tag{32.3}$$

is defined by the condition

$$\langle \mathbf{Jv}, \mathbf{v}^* \rangle = \langle \mathbf{v}^*, \mathbf{v} \rangle, \qquad \mathbf{v} \in \mathscr{V}, \quad \mathbf{v}^* \in \mathscr{V}^* \tag{32.4}$$

Clearly, $\mathbf{J}$ is well defined by (32.4) since for each $\mathbf{v} \in \mathscr{V}$ the right-hand side of (32.4) is a linear function of $\mathbf{v}^*$. To see that $\mathbf{J}$ is an isomorphism, we notice first that $\mathbf{J}$ is a linear transformation, because for each $\mathbf{v}^* \in \mathscr{V}^*$ the right-hand side is linear in $\mathbf{v}$. Now $\mathbf{J}$ also one-to-one, since if $\mathbf{Jv} = \mathbf{Ju}$, then (32.4) implies that

$$\langle \mathbf{v}^*, \mathbf{v} \rangle = \langle \mathbf{v}^*, \mathbf{u} \rangle, \quad \mathbf{v}^* \in \mathscr{V}^* \tag{32.5}$$

which then implies $\mathbf{u} = \mathbf{v}$. From (32.2), we conclude that the one-to-one linear transformation $\mathbf{J}$ is onto, and thus $\mathbf{J}$ is an isomorphism. We summarize this result in the following theorem.

Theorem 32.1. There exists a unique isomorphism $\mathbf{J}$ from $\mathscr{V}$ to $\mathscr{V}^{**}$ satisfying the condition (32.4).

Since the isomorphism $\mathbf{J}$ is defined without using any structure in addition to the vector space structure on $\mathscr{V}$, its notation can often be suppressed without any ambiguity. We shall adopt such a convention here and identify any $\mathbf{v} \in \mathscr{V}$ as a linear function on $\mathscr{V}^*$

$$\mathbf{v} : \mathscr{V}^* \to \mathscr{R}$$

by the condition that defines $\mathbf{J}$, namely

$$\langle \mathbf{v}, \mathbf{v}^* \rangle = \langle \mathbf{v}^*, \mathbf{v} \rangle, \quad \text{for all } \mathbf{v}^* \in \mathscr{V}^* \tag{32.6}$$

In doing so, we allow the same symbol $\mathbf{v}$ to represent two different objects: an element of the vector space $\mathscr{V}$ and a linear function on the vector space $\mathscr{V}^*$, and the two objects are related to each other through the condition (32.6).

To distinguish an isomorphism such as $\mathbf{J}$, whose notation may be suppressed without causing any ambiguity, from an isomorphism such as $\mathbf{G}$, defined by (31.6), whose notation may not be suppressed, because there are many isomorphisms of similar nature, we call the former isomorphism a *canonical* or *natural isomorphism*. Whether or not an isomorphism is canonical is usually determined by a convention, not by any axioms. A general rule for choosing a canonical isomorphism is that the isomorphism must be defined without using any additional structure other than the basic structure already assigned to the underlying spaces; further, by suppressing the notation of the canonical isomorphism no ambiguity is likely to arise. Hence the choice of a canonical isomorphism depends on the basic structure of the vector spaces. If we deal with inner product spaces equipped with particular inner products, the isomorphism $\mathbf{G}$ can safely be regarded as canonical, and by choosing $\mathbf{G}$ to be canonical, we can achieve much economy in writing. On the other hand, if we consider vector spaces without any pre-assigned inner product, then we cannot make all possible isomorphisms $\mathbf{G}$ canonical, otherwise the notation becomes ambiguous.

It should be noticed that not every isomorphism whose definition depends only on the basis structure of the underlying space can be made canonical. For example, the operation of taking the dual:

$$* : \mathscr{L}(\mathscr{V}; \mathscr{U}) \to \mathscr{L}(\mathscr{U}^*; \mathscr{V}^*)$$

is defined by using the vector space structure of $\mathscr{V}$ and $\mathscr{U}$ only. However, by suppressing the notation $*$, we encounter immediately much ambiguity, especially when $\mathscr{U}$ is equal to $\mathscr{V}$. Surely, we do not wish to make every endomorphism $\mathbf{A}:\mathscr{V}\to\mathscr{V}$ self-adjoint! Another example will illustrate the point even clearer. The operation of taking the opposite vector of any vector is an isomorphism

$$-:\mathscr{V}\to\mathscr{V}$$

which is defined by using the vector space structure alone. Evidently, we cannot suppress the minus sign without any ambiguity.

To test whether the isomorphism $\mathbf{J}$ can be made a canonical one without any ambiguity, we consider the effect of this choice on the notations for the dual basis and the dual of a linear transformation. Of course we wish to have $\{\mathbf{e}_i\}$, when considered as a basis for $\mathscr{V}^{**}$, to be the dual basis of $\{\mathbf{e}^i\}$, and $\mathbf{A}$, when considered as a linear transformation from $\mathscr{V}^{**}$ to $\mathscr{U}^{**}$, to be the dual of $\mathbf{A}^*$. These results are indeed correct and they are contained in the following.

Theorem 32.2. Given any basis $\{\mathbf{e}_i\}$ for $\mathscr{V}$, then the dual basis of its dual basis $\{\mathbf{e}^i\}$ is $\{\mathbf{J}\mathbf{e}_i\}$.

Proof. This result is more or less obvious. By definition, the basis $\{\mathbf{e}_i\}$ and its dual basis $\{\mathbf{e}^i\}$ are related by

$$\langle \mathbf{e}^i, \mathbf{e}_j \rangle = \delta^i_j$$

for $i, j = 1,\ldots,N$. From (32.4) we have

$$\langle \mathbf{J}\mathbf{e}_j, \mathbf{e}^i \rangle = \langle \mathbf{e}^i, \mathbf{e}_j \rangle$$

Comparing the preceding two equations, we see that

$$\langle \mathbf{J}\mathbf{e}_j, \mathbf{e}^i \rangle = \delta^i_j$$

which means that $\{\mathbf{J}\mathbf{e}_1,\ldots,\mathbf{J}\mathbf{e}_N\}$ is the dual basis of $\{\mathbf{e}^1,\ldots\mathbf{e}^N\}$.

Because of this theorem, after suppressing the notation for $\mathbf{J}$ we say that $\{\mathbf{e}_i\}$ and $\{\mathbf{e}^j\}$ are dual relative to each other. The next theorem shows that the relation holds between $\mathbf{A}$ and $\mathbf{A}^*$.

Theorem 32.3. Given any linear transformation $\mathbf{A}:\mathscr{V}\to\mathscr{U}$, the dual of its dual $\mathbf{A}^*$ is $\mathbf{J}_{\mathscr{U}}\mathbf{A}\mathbf{J}_{\mathscr{V}}^{-1}$. Here $\mathbf{J}_{\mathscr{V}}$ denotes the isomorphism from $\mathscr{V}$ to $\mathscr{V}^{**}$ defined by (32.4) and $\mathbf{J}_{\mathscr{U}}$ denotes the isomorphism from $\mathscr{U}$ to $\mathscr{U}^{**}$ defined by a similar condition.

Proof. By definition $\mathbf{A}$ and $\mathbf{A}^*$ are related by (31.13)

$$\langle \mathbf{u}^*, \mathbf{A}\mathbf{v}\rangle = \langle \mathbf{A}^*\mathbf{u}^*, \mathbf{v}\rangle$$

for all $\mathbf{u}^*\in\mathscr{U}^*, \mathbf{v}\in\mathscr{V}$. From (32.4) we have

$$\langle \mathbf{u}^*, \mathbf{A}\mathbf{v}\rangle = \langle \mathbf{J}_{\mathscr{U}}\mathbf{A}\mathbf{v}, \mathbf{u}^*\rangle, \quad \langle \mathbf{A}^*\mathbf{u}^*, \mathbf{v}\rangle = \langle \mathbf{J}_{\mathscr{V}}\mathbf{v}, \mathbf{A}^*\mathbf{u}^*\rangle$$

Comparing the preceding three equations, we see that

$$\langle \mathbf{J}_{\mathscr{U}}\mathbf{A}\mathbf{J}_{\mathscr{V}}^{-1}(\mathbf{J}_{\mathscr{V}}\mathbf{v}), \mathbf{u}^*\rangle = \langle \mathbf{J}_{\mathscr{V}}\mathbf{v}, \mathbf{A}^*\mathbf{u}^*\rangle$$

Since $\mathbf{J}_{\mathscr{V}}$ is an isomorphism, we can rewrite the last equation as

$$\langle \mathbf{J}_{\mathscr{U}}\mathbf{A}\mathbf{J}_{\mathscr{V}}^{-1}\mathbf{v}^{**}, \mathbf{u}^*\rangle = \langle \mathbf{v}^{**}, \mathbf{A}^*\mathbf{u}^*\rangle$$

Because $\mathbf{v}^{**}\in\mathscr{V}^{**}$ and $\mathbf{u}^*\in\mathscr{U}^*$ are arbitrary, it follows that $\mathbf{J}_{\mathscr{U}}\mathbf{A}\mathbf{J}_{\mathscr{V}}^{-1}$ is the dual of $\mathbf{A}^*$. So if we suppress the notations for $\mathbf{J}_{\mathscr{U}}$ and $\mathbf{J}_{\mathscr{V}}$, then $\mathbf{A}$ and $\mathbf{A}^*$ are the duals relative to each other.

A similar result exists for the operation of taking the orthogonal complement of a subspace; we have the following result.

Theorem 32.4. Given any subspace $\mathscr{U}$ of $\mathscr{V}$, the orthogonal complement of its orthogonal complement $\mathscr{U}^{\perp}$ is $\mathbf{J}(\mathscr{U})$.

We leave the proof of this theorem as an exercise. Because of this theorem we say that $\mathscr{U}$ and $\mathscr{U}^{\perp}$ are orthogonal to each other. As we shall see in the next few sections, the use of canonical isomorphisms, like the summation convention, is an important device to achieve economy in writing. We shall make use of this device whenever possible, so the reader should be prepared to allow one symbol to represent two or more different objects.

The last three theorems show clearly the advantage of making $\mathbf{J}$ a canonical isomorphism, so from now on we shall suppress the symbol for $\mathbf{J}$. In general, if an isomorphism from a vector space $\mathscr{V}$ to a vector space $\mathscr{U}$ is chosen to be canonical, then we write

$$\mathscr{V} \cong \mathscr{U} \tag{32.7}$$

In particular, we have

$$\mathscr{V} \cong \mathscr{V}^{**} \tag{32.8}$$

Exercises

32.1 Prove theorem 32.4.

32.2 Show that by making $\mathbf{J}$ a canonical isomorphism essentially we have identified $\mathscr{V}$ with $\mathscr{V}^{**}, \mathscr{V}^{****}, \ldots,$ and $\mathscr{V}^{*}$ with $\mathscr{V}^{***}, \mathscr{V}^{****}, \ldots$. So a symbol $\mathbf{v}$ or $\mathbf{v}^{*}$, in fact, represents infinitely many objects.

Section 33. Multilinear Functions, Tensors

In Section 31 we discussed the concept of a linear function and the concept of a bilinear function. These concepts can be generalized in an obvious way to *multilinear functions*. In general if $\mathscr{V}_1,\ldots,\mathscr{V}_s$, is a collection of vector spaces, then a *s-linear function* is a function

$$\mathbf{A}:\mathscr{V}_1\times\cdots\times\mathscr{V}_s\to\mathscr{R} \tag{33.1}$$

that is linear in each of its variables while the other variables are held constant. If the vector spaces $\mathscr{V}_1,\ldots,\mathscr{V}_s$, are the vector space $\mathscr{V}$ or its dual space $\mathscr{V}^*$, then $\mathbf{A}$ is called a *tensor* on $\mathscr{V}$. More specifically, a *tensor of order (p, q)* on $\mathscr{V}$, where p and q are positive integers, is a *(p+q)*-linear function

$$\underbrace{\mathscr{V}^*\times\cdots\times\mathscr{V}^*}_{p\text{ times}}\times\underbrace{\mathscr{V}\times\cdots\times\mathscr{V}}_{q\text{ times}}\to\mathscr{R} \tag{33.2}$$

We shall extend this definition to the case $p=q=0$ and define a tensor of order $(0,0)$ to be a scalar in $\mathscr{R}$. A tensor of order $(p,0)$ is a pure *contravariant* tensor of order *p* and a tensor of order $(0,q)$ is a pure *covariant* tensor of order q. In particular, a vector $\mathbf{v}\in\mathscr{V}$ is a pure contravariant tensor of order one. This terminology, of course, is defined relative to a given vector space $\mathscr{V}$ as we have explained in Section 31. If a tensor is not a pure contravariant tensor or a pure covariant tensor, then it is a *mixed* tensor, and for a mixed tensor of order (p,q), p is the *contravariant* order and q is the *covariant* order.

For definiteness, we denote the set of all tensors of order (p,q) on $\mathscr{V}$ by the symbol $\mathscr{T}_q^{\,p}(\mathscr{V})$. However, the set of pure contravariant tensors of order p shall be denoted simply by $\mathscr{T}^p(\mathscr{V})$ and the set of pure covariant tensors of order q shall be denoted simply $\mathscr{T}_q(\mathscr{V})$. Of course, tensors of order $(0,0)$ form the set $\mathscr{R}$ and

$$\mathscr{T}^1(\mathscr{V})=\mathscr{V},\qquad \mathscr{T}_1(\mathscr{V})=\mathscr{V}^* \tag{33.3}$$

Here we have made use of the identification of $\mathscr{V}$ with $\mathscr{V}^{**}$ as explained in Section 32.

We shall now give some examples of tensors.

Example 1. If $\mathbf{A}$ is an endomorphism of $\mathscr{V}$, then we define a function $\hat{\mathbf{A}}:\mathscr{V}^*\times\mathscr{V}\to\mathscr{R}$ by

$$\hat{\mathbf{A}}\left(\mathbf{v}^*, \mathbf{v}\right) \equiv \left\langle \mathbf{v}^*, \mathbf{A}\mathbf{v} \right\rangle \tag{33.4}$$

for all $\mathbf{v}^* \in \mathscr{V}^*$ and $\mathbf{v} \in \mathscr{V}$. Clearly $\hat{\mathbf{A}}$ is bilinear and thus $\hat{\mathbf{A}} \in \mathscr{T}_1^1(\mathscr{V})$. As we shall see later, it is possible to establish a canonical isomorphism from $\mathscr{L}(\mathscr{V}, \mathscr{V})$ to $\mathscr{T}_1^1(\mathscr{V})$ in such a way that the endomorphism $\mathbf{A}$ is identified with the bilinear function $\hat{\mathbf{A}}$. Then the same symbol $\mathbf{A}$ shall represent two objects, namely, an endomorphism of $\mathscr{V}$ and a bilinear function of $\mathscr{V}^* \times \mathscr{V}$. Then (33.4) becomes simply

$$\mathbf{A}\left(\mathbf{v}^*, \mathbf{v}\right) \equiv \left\langle \mathbf{v}^*, \mathbf{A}\mathbf{v} \right\rangle \tag{33.5}$$

Under this canonical isomorphism, the identity automorphism of $\mathscr{V}$ is identified with the scalar product,

$$\mathbf{I}\left(\mathbf{v}^*, \mathbf{v}\right) = \left\langle \mathbf{v}^*, \mathbf{I}\mathbf{v} \right\rangle = \left\langle \mathbf{v}^*, \mathbf{v} \right\rangle \tag{33.6}$$

Example 2. If $\mathbf{v}$ is a vector in $\mathscr{V}$ and $\mathbf{v}^*$ is a covector in $\mathscr{V}^*$, then we define a function

$$\mathbf{v} \otimes \mathbf{v}^* : \mathscr{V} \times \mathscr{V}^* \to \mathscr{R} \tag{33.7}$$

by

$$\mathbf{v} \otimes \mathbf{v}^* \left(\mathbf{u}^*, \mathbf{u}\right) \equiv \left\langle \mathbf{u}^*, \mathbf{v} \right\rangle \left\langle \mathbf{v}^*, \mathbf{u} \right\rangle \tag{33.8}$$

for all $\mathbf{u}^* \in \mathscr{V}^*, \mathbf{u} \in \mathscr{V}$. Clearly, $\mathbf{v} \otimes \mathbf{v}^*$ is a bilinear function, so $\mathbf{v} \otimes \mathbf{v}^* \in \mathscr{T}_1^1(\mathscr{V})$. If we make use of the canonical isomorphism to be established between $\mathscr{T}_1^1(\mathscr{V})$ and $\mathscr{L}(\mathscr{V}; \mathscr{V})$, the tensor $\mathbf{v} \otimes \mathbf{v}^*$ corresponds to an endomorphism of $\mathscr{V}$ such that

$$\mathbf{v} \otimes \mathbf{v}^* \left(\mathbf{u}^*, \mathbf{u}\right) = \left\langle \mathbf{u}^*, \mathbf{v} \otimes \mathbf{v}^* \mathbf{u} \right\rangle$$

or equivalently

$$\left\langle \mathbf{u}^*, \mathbf{v} \right\rangle \left\langle \mathbf{v}^*, \mathbf{u} \right\rangle = \left\langle \mathbf{u}^*, \mathbf{v} \otimes \mathbf{v}^* \mathbf{u} \right\rangle$$

for all $\mathbf{u}^* \in \mathscr{V}^*, \mathbf{u} \in \mathscr{V}$. By the bilinearity of the scalar product, the last equation can be rewritten in the form

$$\left\langle \mathbf{u}^*, \left\langle \mathbf{v}^*, \mathbf{u} \right\rangle \mathbf{v} \right\rangle = \left\langle \mathbf{u}^*, \mathbf{v} \otimes \mathbf{v}^* \mathbf{u} \right\rangle$$

Then by the definiteness of the scalar product, we have

$$\left\langle \mathbf{v}^*, \mathbf{u} \right\rangle \mathbf{v} = \mathbf{v} \otimes \mathbf{v}^* \mathbf{u}, \qquad \text{for all } \mathbf{u} \in \mathscr{V} \tag{33.9}$$

which defines $\mathbf{v} \otimes \mathbf{v}^*$ as an endomorphism of $\mathscr{V}$. The tensor or the endomorphism $\mathbf{v} \otimes \mathbf{v}^*$ is called the *tensor product* of $\mathbf{v}$ and $\mathbf{v}^*$.

Clearly, the tensor product can be defined for arbitrary number of vectors and covectors. Let $\mathbf{v}_1, \ldots, \mathbf{v}_p$ be vectors in $\mathscr{V}$ and $\mathbf{v}^1, \ldots, \mathbf{v}^q$ be covectors in $\mathscr{V}^*$. Then we define a function

$$\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_p \otimes \mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^q : \underbrace{\mathscr{V}^* \times \cdots \times \mathscr{V}^*}_{p \text{ times}} \times \underbrace{\mathscr{V} \times \cdots \times \mathscr{V}}_{q \text{ times}} \to \mathscr{R}$$

by

$$\begin{aligned} &\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_p \otimes \mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^q \left(\mathbf{u}^1, \ldots, \mathbf{u}^p, \mathbf{u}_1, \ldots, \mathbf{u}_q \right) \\ &\qquad \equiv \left\langle \mathbf{u}^1, \mathbf{v}_1 \right\rangle \cdots \left\langle \mathbf{u}^p, \mathbf{v}_p \right\rangle \left\langle \mathbf{v}^1, \mathbf{u}_1 \right\rangle \cdots \left\langle \mathbf{v}^q, \mathbf{u}_q \right\rangle \end{aligned} \tag{33.10}$$

for all $\mathbf{u}_1, \ldots, \mathbf{u}_q \in \mathscr{V}$ and $\mathbf{u}^1, \ldots, \mathbf{u}^q \in \mathscr{V}^*$. Clearly this function is $(p+q)-$linear, so that $\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_p \otimes \mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^q \in \mathscr{T}_q^{\,p}(\mathscr{V})$, is called the *tensor product* of $\mathbf{v}_1, \ldots, \mathbf{v}_p$ and $\mathbf{v}^1, \ldots, \mathbf{v}^q$.

Having seen some examples of tensors on $\mathscr{V}$, we turn now to the structure of the set $\mathscr{T}_q^{\,p}(\mathscr{V})$. We claim that $\mathscr{T}_q^{\,p}(\mathscr{V})$ has the structure of a vector space and the dimension of $\mathscr{T}_q^{\,p}(\mathscr{V})$ is equal to $N^{(p+q)}$, where N is the dimension of $\mathscr{V}$. To make $\mathscr{T}_q^{\,p}(\mathscr{V})$ a vector space, we define the operation of addition of any $\mathbf{A}, \mathbf{B} \in \mathscr{T}_q^{\,p}(\mathscr{V})$ and the scalar multiplication of $\mathbf{A}$ by $\alpha \in \mathscr{R}$ by

$$\begin{aligned} &(\mathbf{A} + \mathbf{B})\left(\mathbf{v}^1, \ldots, \mathbf{v}^p, \mathbf{v}_1, \ldots \mathbf{v}_q \right) \\ &\qquad \equiv \mathbf{A}\left(\mathbf{v}^1, \ldots, \mathbf{v}^p, \mathbf{v}_1, \ldots \mathbf{v}_q \right) + \mathbf{B}\left(\mathbf{v}^1, \ldots, \mathbf{v}^p, \mathbf{v}_1, \ldots \mathbf{v}_q \right) \end{aligned} \tag{33.11}$$

and

$$(\alpha \mathbf{A})\left(\mathbf{v}^1, \ldots, \mathbf{v}^p, \mathbf{v}_1, \ldots \mathbf{v}_q \right) \equiv \alpha \mathbf{A}\left(\mathbf{v}^1, \ldots, \mathbf{v}^p, \mathbf{v}_1, \ldots \mathbf{v}_q \right) \tag{33.12}$$

respectively, for all $\mathbf{v}^1,\ldots,\mathbf{v}^p \in \mathscr{V}^*$ and $\mathbf{v}_1,\ldots,\mathbf{v}_q \in \mathscr{V}$. We leave as an exercise to the reader the proof of the following theorem.

Theorem 33.1. $\mathscr{T}_q^{\,p}(\mathscr{V})$ is a vector space with respect to the operations of addition and scalar multiplication defined by (33.11) and (33.12). The null element of $\mathscr{T}_q^{\,p}(\mathscr{V})$,of course, is the zero tensor $\mathbf{0}$:

$$\mathbf{0}\left(\mathbf{v}^1,\ldots,\mathbf{v}^p,\ \mathbf{v}_1,\ldots,\mathbf{v}_q\right)=0 \tag{33.13}$$

for all $\mathbf{v}^1,\ldots,\mathbf{v}^p \in \mathscr{V}^*$ and $\mathbf{v}_1,\ldots,\mathbf{v}_q \in \mathscr{V}$.

Next, we determine the dimension of the vector space $\mathscr{T}_q^{\,p}(\mathscr{V})$ by introducing the concept of a *product basis*.

Theorem 33.2. Let $\{\mathbf{e}_i\}$ and $\{\mathbf{e}^i\}$ be dual bases for $\mathscr{V}$ and $\mathscr{V}^*$. Then the set of tensor products

$$\left\{\mathbf{e}_{i_1}\otimes\cdots\otimes\mathbf{e}_{i_p}\otimes\mathbf{e}^{j_1}\otimes\cdots\otimes\mathbf{e}^{i_q},\ i_1,\ldots,i_p,\ j_1,\ldots,j_q=1,\ldots,N\right\} \tag{33.14}$$

forms a basis for $\mathscr{T}_q^{\,p}(\mathscr{V})$, called the *product basis*. In particular,

$$\dim\mathscr{T}_q^{\,p}(\mathscr{V})=N^{(p+q)} \tag{33.15}$$

Proof. We shall prove that the set of tensor products (33.14) is a linearly independent generating set for $\mathscr{T}_q^{\,p}(\mathscr{V})$. To prove that the set (33.14) is linearly independent, let

$$A^{i_1,\ldots,i_p}{}_{j_1,\ldots,j_q}\mathbf{e}_{i_1}\otimes\cdots\otimes\mathbf{e}_{i_p}\otimes\mathbf{e}^{j_1}\otimes\cdots\otimes\mathbf{e}^{i_q}=\mathbf{0} \tag{33.16}$$

where the right-hand side is the zero tensor given by (33.13). Then from (33.10) we have

$$\begin{aligned}0&=A^{i_1\cdots i_p}{}_{j_1\cdots j_q}\mathbf{e}_{i_1}\otimes\cdots\otimes\mathbf{e}_{i_p}\otimes\mathbf{e}^{j_q}\otimes\cdots\otimes\mathbf{e}^{j_q}\left(\mathbf{e}^{k_1},\ldots,\mathbf{e}^{k_p},\mathbf{e}_{l_1},\ldots,\mathbf{e}_{l_q}\right)\\&=A^{i_1\cdots i_p}{}_{j_1\cdots j_q}\left\langle\mathbf{e}^{k_1},\mathbf{e}_{i_1}\right\rangle\cdots\left\langle\mathbf{e}^{k_p},\mathbf{e}_{i_p}\right\rangle\left\langle\mathbf{e}^{j_1},\mathbf{e}_{l_1}\right\rangle\cdots\left\langle\mathbf{e}^{j_q},\mathbf{e}_{l_q}\right\rangle\\&=A^{i_1\cdots i_p}{}_{j_1\cdots j_q}\delta_{i_1}^{k_1}\cdots\delta_{i_p}^{k_p}\delta_{l_1}^{j_1}\cdots\delta_{l_q}^{j_q}=A^{k_1\cdots k_p}{}_{l_1\cdots l_q}\end{aligned} \tag{33.17}$$

which shows that the set (33.14) is linearly independent. Next, we show that every tensor $\mathbf{A} \in \mathscr{T}_q^{\,p}(\mathscr{V})$ can be expressed as a linear combination of the set (33.14). We define $N^{(p+q)}$ scalars $\left\{A^{i_1 \ldots i_p}{}_{j_1 \ldots j_q}, i_1 \ldots i_p, j_1 \ldots j_q = 1, \ldots, N\right\}$ by

$$A^{i_1 \ldots i_p}{}_{j_1 \ldots j_q} = \mathbf{A}\left(\mathbf{e}^{i_1}, \ldots, \mathbf{e}^{i_p}, \mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_q}\right) \tag{33.18}$$

Now we reverse the steps of equation (33.17) and obtain

$$\begin{aligned} A^{i_1 \ldots i_p}{}_{j_1 \ldots j_q} \mathbf{e}_{i_1} \otimes \cdots \otimes \mathbf{e}_{i_p} \otimes \mathbf{e}^{j_1} \otimes \cdots \otimes \mathbf{e}^{j_q}\left(\mathbf{e}^{k_1}, \ldots, \mathbf{e}^{k_p}, \mathbf{e}_{l_1}, \ldots \mathbf{e}_{l_q}\right) \\ = \mathbf{A}\left(\mathbf{e}^{k_1}, \ldots, \mathbf{e}^{k_p}, \mathbf{e}_{l_1}, \ldots \mathbf{e}_{l_q}\right) \end{aligned} \tag{33.19}$$

for all $k_1, \ldots, k_p,\ l_1, \ldots, l_q = 1, \ldots, N$. Since $\mathbf{A}$ is multilinear and since $\{\mathbf{e}_i\}$ and $\{\mathbf{e}^i\}$ are dual bases for $\mathscr{V}$ and $\mathscr{V}^*$, the condition (33.19) implies that

$$\begin{aligned} A^{i_1 \ldots i_p}{}_{j_1 \ldots j_q} \mathbf{e}_{i_1} \otimes \cdots \otimes \mathbf{e}_{i_p} \otimes \mathbf{e}^{j_1} \otimes \cdots \otimes \mathbf{e}^{j_q}\left(\mathbf{v}^1, \ldots, \mathbf{v}^p, \mathbf{v}_1, \ldots \mathbf{v}_q\right) \\ = \mathbf{A}\left(\mathbf{v}^1, \ldots, \mathbf{v}^p, \mathbf{v}_1, \ldots \mathbf{v}_q\right) \end{aligned}$$

for all $\mathbf{v}_1, \ldots, \mathbf{v}_q \in \mathscr{V}$ and $\mathbf{v}^1, \ldots, \mathbf{v}^p \in \mathscr{V}^*$ and thus

$$\mathbf{A} = A^{i_1 \ldots i_p}{}_{j_1 \ldots j_q} \mathbf{e}_{i_1} \otimes \cdots \otimes \mathbf{e}_{i_p} \otimes \mathbf{e}^{j_1} \otimes \cdots \otimes \mathbf{e}^{j_q} \tag{33.20}$$

Now from Theorem 9.10 we conclude that the set (33.14) is a basis for $\mathscr{T}_q^{\,p}(\mathscr{V})$.

Having determined the dimension of $\mathscr{T}_q^{\,p}(\mathscr{V})$, we can now prove that $\mathscr{T}_1^{\,1}(\mathscr{V})$ is isomorphic to $\mathscr{L}(\mathscr{V};\mathscr{V})$, a result mentioned in Example 1. As before we define the operation

$$\hat{}: \mathscr{L}(\mathscr{V};\mathscr{V}) \to \mathscr{T}_1^{\,1}(\mathscr{V})$$

by the condition (33.4). Since from (16.8) and (33.15), the vector space $\mathscr{L}(\mathscr{V};\mathscr{V})$ and $\mathscr{T}_1^{\,1}(\mathscr{V})$ are of the same dimension, namely N^2, it suffices to show that the operation $\hat{}$ is one-to-one. Indeed, let $\hat{\mathbf{A}} = \mathbf{0}$. Then from (33.13) and (33.4) we have

$$\left\langle \mathbf{v}^*, \mathbf{A}\mathbf{v} \right\rangle = 0$$

for all $\mathbf{v}^* \in \mathscr{V}^*$ and all $\mathbf{v} \in \mathscr{V}$. Now since the scalar product is definite, this condition implies

$$\mathbf{A}\mathbf{v} = \mathbf{0}$$

for all $\mathbf{v} \in \mathscr{V}$, and thus $\mathbf{A} = \mathbf{0}$. Consequently the operation $\hat{}$ is an isomorphism.

As remarked in Example 1, we shall suppress the notation $\hat{}$ by regarding the isomorphism it represents as a canonical one. Therefore, we can replace the formula (33.4) by the formula (33.5).

Now returning to the space $\mathscr{T}_q^{\,p}(\mathscr{V})$ in general, we see that a corollary of Theorem 33.2 is the simple fact that the set of all tensor products of the form (33.10) is a generating set for $\mathscr{T}_q^{\,p}(\mathscr{V})$. We define a tensor that can be represented as a tensor product to be a *simple tensor*. It should be noted, however, that such a representation for a simple tensor is not unique. Indeed, from (33.8), we have

$$\mathbf{v} \otimes \mathbf{v}^* = (2\mathbf{v}) \otimes \left(\frac{1}{2} \mathbf{v}^* \right)$$

for any $\mathbf{v} \in \mathscr{V}$ and $\mathbf{v}^* \in \mathscr{V}^*$ since

$$\begin{aligned} (2\mathbf{v}) \otimes \left(\frac{1}{2} \mathbf{v}^* \right) (\mathbf{u}^*, \mathbf{u}) &= \langle \mathbf{u}^*, 2\mathbf{v} \rangle \left\langle \frac{1}{2} \mathbf{v}^*, \mathbf{u} \right\rangle \\ &= \langle \mathbf{u}^*, \mathbf{v} \rangle \langle \mathbf{v}^*, \mathbf{u} \rangle = \mathbf{v} \otimes \mathbf{v}^* (\mathbf{u}^*, \mathbf{u}) \end{aligned}$$

for all $\mathbf{u}^*$ and $\mathbf{u}$. In general, the tensor product, as defined by (33.10), can be regarded as a mapping

$$\otimes : \underbrace{\mathscr{V} \times \cdots \times \mathscr{V}}_{p \text{ times}} \times \underbrace{\mathscr{V}^* \times \cdots \times \mathscr{V}^*}_{q \text{ times}} \to \mathscr{T}_q^{\,p}(\mathscr{V}) \tag{33.21}$$

given by

$$\otimes : (\mathbf{v}_1, \cdots, \mathbf{v}_p, \mathbf{v}^1, \cdots, \mathbf{v}^q) = \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_p \otimes \mathbf{v}^1 \times \cdots \otimes \mathbf{v}^q \tag{33.22}$$

for all $\mathbf{v}_1, \ldots, \mathbf{v}_p \in \mathscr{V}$ and $\mathbf{v}^1, \ldots, \mathbf{v}^q \in \mathscr{V}^*$. It is easy to verify that the mapping $\otimes$ is multilinear in the usual sense, i.e.,

$$\begin{aligned}\otimes\left(\mathbf{v}_1,\cdots,\alpha\mathbf{v}+\beta\mathbf{u},\ldots,\mathbf{v}^q\right)&=\alpha\otimes\left(\mathbf{v}_1,\cdots,\mathbf{v},\ldots,\mathbf{v}^q\right)\\&\quad+\beta\otimes\left(\mathbf{v}_1,\cdots,\mathbf{u},\ldots,\mathbf{v}^q\right)\end{aligned}\tag{33.23}$$

where $\alpha\mathbf{v}+\beta\mathbf{u}$, $\mathbf{v}$ and $\mathbf{u}$ all take the same position in the argument of $\otimes$ but that position is arbitrary. From (33.23), we see that

$$\begin{aligned}&\left(\alpha\mathbf{v}_1\right)\otimes\mathbf{v}_2\otimes\cdots\otimes\mathbf{v}_p\otimes\mathbf{v}^1\otimes\cdots\otimes\mathbf{v}^q\\&\quad=\mathbf{v}_1\otimes\left(\alpha\mathbf{v}_2\right)\otimes\cdots\otimes\mathbf{v}_p\otimes\mathbf{v}^1\otimes\cdots\otimes\mathbf{v}^q\end{aligned}$$

We can extend the operation of tensor product from vectors and covectors to tensors in general. If $\mathbf{A}\in\mathscr{T}_q^{\,p}\left(\mathscr{V}\right)$ and $\mathbf{B}\in\mathscr{T}_s^{\,r}\left(\mathscr{V}\right)$, then we define their tensor product $\mathbf{A}\otimes\mathbf{B}$ to be a tensor of order $\left(p+r,q+s\right)$ by

$$\begin{aligned}&\mathbf{A}\otimes\mathbf{B}\left(\mathbf{v}^1,\ldots,\mathbf{v}^{p+r},\mathbf{v}_1,\ldots\mathbf{v}_{q+s}\right)\\&\quad=\mathbf{A}\left(\mathbf{v}^1,\ldots,\mathbf{v}^p,\mathbf{v}_1,\ldots\mathbf{v}_q\right)\mathbf{B}\left(\mathbf{v}^{p+1},\ldots,\mathbf{v}^{p+r},\mathbf{v}_{q+1},\ldots,\mathbf{v}_{q+s}\right)\end{aligned}\tag{33.24}$$

for all $\mathbf{v}^1,\ldots,\mathbf{v}^{p+r}\in\mathscr{V}^*$ and $\mathbf{v}_1,\ldots,\mathbf{v}_{q+s}\in\mathscr{V}$. Applying this definition to arbitrary tensors $\mathbf{A}$ and $\mathbf{B}$ yields a mapping

$$\otimes:\mathscr{T}_q^{\,p}\left(\mathscr{V}\right)\times\mathscr{T}_s^{\,r}\left(\mathscr{V}\right)\to\mathscr{T}_{q+s}^{\,p+r}\left(\mathscr{V}\right)\tag{33.25}$$

Clearly this operation can be further extended to more than two tensor spaces, say

$$\otimes:\mathscr{T}_{q_1}^{\,p_1}\left(\mathscr{V}\right)\times\cdots\times\mathscr{T}_{q_k}^{\,p_k}\left(\mathscr{V}\right)\to\mathscr{T}_{q_1+\cdots+q_k}^{\,p_1+\cdots+p_k}\left(\mathscr{V}\right)\tag{33.26}$$

in such a way that

$$\otimes\left(\mathbf{A}_1,\ldots,\mathbf{A}_k\right)=\mathbf{A}_1\otimes\cdots\otimes\mathbf{A}_k\tag{33.27}$$

where $\mathbf{A}_i\in\mathscr{T}_{q_i}^{\,p_i}\left(\mathscr{V}\right)$, $i=1,\ldots,k$. It is easy to verify that this tensor product operation is also multilinear in the sense generalizing (33.23) to tensors. In component form

$$\left(\mathbf{A}\otimes\mathbf{B}\right)^{i_1\ldots i_p}{}_{j_1\ldots j_{q+s}}=\mathbf{A}^{i_1\ldots i_p}{}_{j_1\ldots j_q}\mathbf{B}^{i_{p+1}\ldots i_{p+r}}{}_{j_{q+1}\ldots j_{q+s}}\tag{33.28}$$

which can be generalized obviously for (33.27) also. From (33.28), or from (33.24), we see that the tensor product is not commutative, but it is associative and distributive.

Relative to a product basis the component form of a tensor $\mathbf{A} \in \mathscr{T}_q^{\,p}(\mathscr{V})$ is given by (33.20) where the components of $\mathbf{A}$ can be obtained by (33.18). If we transfer the basis $\{\mathbf{e}_i\}$ to $\{\hat{\mathbf{e}}_i\}$ as shown by (31.21) and (31.22), then the components of $\mathbf{A}$ as well as the product basis relative to $\{\mathbf{e}_i\}$ must be transformed also. The following theorem gives the transformation laws.

Theorem 33.3. Under the transformation of bases (31.21) and (31.22) the product basis (33.14) for $\mathscr{T}_q^{\,p}(\mathscr{V})$ transforms according to the rule

$$\begin{aligned}&\hat{\mathbf{e}}_{i_1} \otimes \cdots \otimes \hat{\mathbf{e}}_{i_p} \otimes \hat{\mathbf{e}}^{j_1} \otimes \cdots \otimes \hat{\mathbf{e}}^{j_q} \\ &\quad = T_{i_1}^{k_1} \cdots T_{i_p}^{k_p} \hat{T}_{l_1}^{j_1} \cdots \hat{T}_{l_q}^{j_q} \mathbf{e}_{k_1} \otimes \cdots \otimes \mathbf{e}_{k_p} \otimes \mathbf{e}^{l_1} \otimes \cdots \otimes \mathbf{e}^{l_q}\end{aligned} \tag{33.29}$$

and the components of any tensor $\mathbf{A} \in \mathscr{T}_q^{\,p}(\mathscr{V})$ transform according to the rule.

$$\hat{A}^{i_1 \ldots i_p}{}_{j_1 \ldots j_q} = \hat{T}_{k_1}^{i_1} \cdots \hat{T}_{k_p}^{i_p} T_{j_1}^{l_1} \cdots T_{j_q}^{l_q} A^{k_1 \ldots k_p}{}_{l_1 \ldots l_q} \tag{33.30}$$

The proof of these rules involves no more than the multilinearity of the tensor product $\otimes$ and the tensor $\mathbf{A}$. Many classical treatises on tensors use the transformation rule such as (33.30) to define a tensor. The next theorem connects this alternate definition with the one we used.

Theorem 33.4. Given any two sets of $N^{(p+q)}$ scalars $\left\{A^{i_1 \ldots i_p}{}_{j_1 \ldots j_q}\right\}$ and $\left\{\hat{A}^{i_1 \ldots i_p}{}_{j_1 \ldots j_q}\right\}$ related by the transformation rule (33.30), there exists a tensor $\mathbf{A} \in \mathscr{T}_q^{\,p}(\mathscr{V})$ whose components relative to the product bases of $\{\mathbf{e}_i\}$ to $\{\hat{\mathbf{e}}_i\}$ are $\left\{A^{i_1 \ldots i_p}{}_{j_1 \ldots j_q}\right\}$ and $\left\{\hat{A}^{i_1 \ldots i_p}{}_{j_1 \ldots j_q}\right\}$ provided that the bases are related by (31.21) and (31.22).

This theorem is obvious, since we can define the tensor $\mathbf{A}$ by (33.20); then the transformation rule (33.30) shows that the components of $\mathbf{A}$ relative to the product basis of $\{\mathbf{e}_i\}$ are $\left\{A^{i_1 \ldots i_p}{}_{j_1 \ldots j_q}\right\}$. Thus a tensor $\mathbf{A}$ corresponds to an equivalence set of components, with the transformation rule serving as the equivalence relation.

As an illustration of the preceding theorem, let us examine whether or not there exists a tensor whose components relative to the product basis of any basis are the values of the generalized Kronecker delta introduced in Section 20. The answer turns out to be yes, since we have the identify

$$\delta^{i_1\cdots i_r}_{j_1\cdots j_r} \equiv \hat{T}^{i_1}_{k_1}\cdots\hat{T}^{i_r}_{k_r}T^{l_1}_{j_1}\cdots T^{l_r}_{j_r}\,\delta^{k_1\cdots k_r}_{l_1\cdots l_r} \tag{33.31}$$

which follows from the fact that $\left[T^i_j\right]$ and $\left[\hat{T}^i_j\right]$ are the inverse of each other. We leave the proof of this identity as an exercise for the reader. From Theorem 33.4 and the identity (33.31), we see that there exist a tensor $\mathbf{K}$, of order (r,r) such that

$$\mathbf{K}_r = \frac{1}{r!}\delta^{i_1\cdots i_r}_{j_1\cdots j_r}\mathbf{e}_{i_1}\otimes\cdots\otimes\mathbf{e}_{i_r}\otimes\mathbf{e}^{j_1}\otimes\cdots\otimes\mathbf{e}^{j_r} \tag{33.32}$$

relative to any basis $\{\mathbf{e}_i\}$. This tensor plays an important role in the next chapter.

By the same line of reasoning, we may ask whether or not there exists a tensor whose components relative to the product basis of any basis are the values of the $\varepsilon-$symbols also introduced in Section 20. The answer turns out to be no, since we have the identities

$$\varepsilon_{i_1\ldots i_N} = \det\left[\hat{T}^i_j\right]T^{j_1}_{i_1}\cdots T^{j_N}_{i_N}\varepsilon_{j_1\ldots j_N} \tag{33.33}$$

and

$$\varepsilon^{i_1\ldots i_N} = \det\left[T^l_j\right]\hat{T}^{i_1}_{j_1}\cdots\hat{T}^{i_N}_{j_N}\varepsilon^{j_1\ldots j_N} \tag{33.34}$$

which also follows from the fact that $\left[T^i_j\right]$ and $\left[\hat{T}^i_j\right]$ are the inverses of each other [cf. (21.8)]. Since these identities do not agree with the transformation rule (33.30), we can conclude that there exists no tensor whose components relative to the product basis of any basis are always the values of the $\varepsilon-$symbols. In other words, if the values of the $\varepsilon-$symbols are the components of a tensor relative to the product basis of one particular basis $\{\mathbf{e}_i\}$, then the components of the same tensor relative to the product basis of another basis generally are not the values of the $\varepsilon-$symbols, unless the transformation matrices $\left[T^i_j\right]$ and $\left[\hat{T}^i_j\right]$ have unit determinant.

In the classical treatises on tensors, the transformation rules (33.33) and (33.34), or more generally

$$A^{i_1\cdots i_p}{}_{j_1\cdots j_q} = \varepsilon\det\left[T^i_j\right]^w\hat{T}^{i_1}_{k_1}\cdots\hat{T}^{i_p}_{k_p}T^{l_1}_{j_1}\cdots T^{l_q}_{j_q}A^{k_1\ldots k_p}{}_{l_1\ldots l_q} \tag{33.35}$$

are used to define *relative tensors*. The exponent w and the coefficient ε on the right-hand side of (33.35) are called the *weight* and the *parity* of the relative tensor, respectively. A relative tensor is called *polar* if its parity has the value +1 in all transformations, while a relative tensor is

called *axial* if ε is equal to the sign of the determinant of the transformation matrix $\left[T_j^i\right]$. In particular, (33.33) shows that $\left\{\varepsilon_{i_1\ldots i_N}\right\}$ are the components of an axial covariant tensor of order N and weight -1, while (33.34) shows that $\left\{\varepsilon^{i_1\ldots i_N}\right\}$ are the components of an axial contravariant tensor of order N and weight $+1$. We shall see some more examples of relative tensors in the next chapter.

Exercises

33.1 Prove Theorem 33.1.

33.2 Prove equation (33.31).

33.3 Under the canonical isomorphism of $\mathscr{L}(\mathscr{V};\mathscr{V})$ with $\mathscr{T}_1^1(\mathscr{V})$, show that an endomorphism $\mathbf{A}:\mathscr{V}\to\mathscr{V}$ and its dual $\mathbf{A}^*:\mathscr{V}^*\to\mathscr{V}^*$ correspond to the same tensor.

33.4 Define an isomorphism from $\mathscr{L}\left(\mathscr{L}(\mathscr{V};\mathscr{V});\mathscr{L}(\mathscr{V};\mathscr{V})\right)$ to $\mathscr{T}_2^2(\mathscr{V})$ independent of any basis.

33.5 If $\mathbf{A}\in\mathscr{T}_2(\mathscr{V})$, show that the determinant of the component matrix $\left[A_{ij}\right]$ of $\mathbf{A}$ defines a polar scalar of weight two, i.e., the determinant obeys the transformation rule

$$\det\left[\hat{A}_{ij}\right]=\left(\det\left[T_l^k\right]\right)^2\det\left[A_{ij}\right]$$

33.6 Define isomorphisms from $\mathscr{L}(\mathscr{V};\mathscr{V}^*)$ to $\mathscr{T}_2(\mathscr{V})$, $\mathscr{L}(\mathscr{V}^*;\mathscr{V})$ to $\mathscr{T}^2(\mathscr{V})$, and from $\mathscr{L}(\mathscr{V}^*;\mathscr{V}^*)$ to $\mathscr{T}_1^1(\mathscr{V})$ independent of any basis.

33.7 Given a relation of the form

$$A_{i_1\ldots i_p klm}B(k,1,m)=C_{i_1\ldots i_p}$$

where $\left\{A_{i_1\ldots i_p klm}\right\}$ and $\left\{C_{i_1\ldots i_p}\right\}$ are components of tensors with respect to any basis, show that the set $\left\{B(k,l,m)\right\}$, likewise, can be regarded as components of a tensor, belonging to $\mathscr{T}^3(\mathscr{V})$ in this case. This result is known as the *quotient theorem* in classical treatises on tensors.

33.8 If $\mathbf{A}\in\mathscr{T}_{p+q}(\mathscr{V})$, define a tensor $\mathbf{T}_{pq}\mathbf{A}\in\mathscr{T}_{p+q}(\mathscr{V})$ by

$$\mathbf{T}_{pq}\mathbf{A}(\mathbf{u}_1,\ldots,\mathbf{u}_q,\mathbf{v}_1,\ldots,\mathbf{v}_p)\equiv\mathbf{A}(\mathbf{v}_1,\ldots,\mathbf{v}_p,\mathbf{u}_1,\ldots,\mathbf{u}_q) \tag{33.36}$$

for all $\mathbf{u}_1,\ldots,\mathbf{u}_q,\mathbf{v}_1,\ldots,\mathbf{v}_p \in \mathscr{V}$. Show that the operation

$$\mathbf{T}_{pq}:\ \mathscr{T}_{p+q}(\mathscr{V}) \to \mathscr{T}_{p+q}(\mathscr{V}) \tag{33.37}$$

is an automorphism of $\mathscr{T}_{p+q}(\mathscr{V})$. We call $\mathbf{T}_{pq}$ the *generalized transpose operation*. What is the relation between the components of $\mathbf{A}$ and $\mathbf{T}_{pq}\mathbf{A}$?

Section 34. Contractions

In this section we shall consider the operation of contracting a tensor of order (p,q) to obtain a tensor of order $(p-1,q-1)$, where p,q are greater than or equal to one. To define this important operation, we prove first a useful property of the tensor space $\mathscr{T}_q^{\,p}(\mathscr{V})$. In the preceding section we defined a tensor of order (p,q) to be a multilinear function

$$\mathbf{A}:\underbrace{\mathscr{V}^*\times\cdots\times\mathscr{V}^*}_{p\text{ times}}\times\underbrace{\mathscr{V}\times\cdots\times\mathscr{V}}_{q\text{ times}}\to\mathscr{R}$$

Clearly, this concept can be generalized to a multilinear transformation from the (p,q)-fold Cartesian product of $\mathscr{V}$ and $\mathscr{V}^*$ to an arbitrary vector space $\mathscr{U}$, namely

$$\mathbf{Z}:\underbrace{\mathscr{V}\times\cdots\times\mathscr{V}}_{p\text{ times}}\times\underbrace{\mathscr{V}^*\times\cdots\times\mathscr{V}^*}_{q\text{ times}}\to\mathscr{U} \tag{34.1}$$

The condition that $\mathbf{Z}$ be a multilinear transformation is similar to that for a multilinear function, namely, $\mathbf{Z}$ is linear in each one of its variables while its other variables are held constant, e.g., the tensor product $\otimes$ given by (33.21) is a multilinear transformation. The next theorem shows that, in some sense, any multilinear transformation $\mathbf{Z}$ of the form (34.1) can be factored through the tensor product $\otimes$ given by (33.21). This fact is known as the *universal factorization property* of the tensor product.

Theorem 34.1. If $\mathbf{Z}$ is an arbitrary multilinear transformation of the form (34.1), then there exists a unique linear transformation

$$\mathbf{C}:\mathscr{T}_q^{\,p}(\mathscr{V})\to\mathscr{U} \tag{34.2}$$

such that

$$\mathbf{Z}\left(\mathbf{v}_1\ldots,\mathbf{v}_p,\mathbf{v}^1,\ldots,\mathbf{v}^q\right)=\mathbf{C}\left(\mathbf{v}_1\otimes\cdots\otimes\mathbf{v}_p\otimes\mathbf{v}^1\otimes\cdots\otimes\mathbf{v}^q\right) \tag{34.3}$$

for all $\mathbf{v}_1\ldots,\mathbf{v}_p\in\mathscr{V}$ and $\mathbf{v}^1,\ldots,\mathbf{v}^q\in\mathscr{V}^*$.

Proof. Since the simple tensors form a generating set of $\mathscr{T}_q^{\,p}(\mathscr{V})$, if the linear transformation $\mathbf{C}$ satisfying (34.3) exists, then it must be unique. To prove the existence of $\mathbf{C}$, we choose a basis $\{\mathbf{e}_i\}$ for $\mathscr{V}$ and define the product basis (33.14) for $\mathscr{T}_q^{\,p}(\mathscr{V})$ as before. Then we define

$$\mathbf{C}\left(\mathbf{e}_{i_1}\otimes\cdots\otimes\mathbf{e}_{i_p}\otimes\mathbf{e}^{j_1}\otimes\cdots\otimes\mathbf{e}^{j_q}\right)\equiv\mathbf{Z}\left(\mathbf{e}_{i_1},\ldots,\mathbf{e}_{i_p},\mathbf{e}^{j_1},\ldots,\mathbf{e}^{j_q}\right) \tag{34.4}$$

for all $i_1,\ldots,i_p, j_1,\ldots, j_q = 1,\ldots,N$ and extend $\mathbf{C}$ to all tensors in $\mathscr{T}_q^{\,p}(\mathscr{V})$ by linearity. Now it is clear that the linear transformation $\mathbf{C}$ defined in this way satisfies the condition (34.3), since both $\mathbf{C}$ and $\mathbf{Z}$ are multilinear in the vectors $\mathbf{v}_1,\ldots,\mathbf{v}_p$ and the covectors $\mathbf{v}^1,\ldots,\mathbf{v}^q$, and they agree on the dual bases $\{\mathbf{e}_i\}$ and $\{\mathbf{e}^i\}$ for $\mathscr{V}$ and $\mathscr{V}^*$, as shown in (34.4). Hence they agree on all $\left(\mathbf{v}_1,\ldots,\mathbf{v}_p,\mathbf{v}^1,\ldots,\mathbf{v}^q\right)$.

If we use the symbol $\otimes$ to denote the multilinear transformation (33.21), then the condition (34.3) can be rewritten as

$$\mathbf{Z}=\mathbf{C}\circ\otimes \tag{34.5}$$

where the operation $\circ$ on the right-hand side of (34.5) denotes the composition as defined in Section 3. Equation (34.5) expresses the meaning of the universal factorization property. In the modern treatises on tensors, this property is often used to define the tensor product and the tensor spaces. Our approach to the concept of a tensor is a compromise between this abstract modern concept and the classical concept based on transformation rules; the preceding theorem and Theorems 33.3 and 33.4 connect our concept with the other two.

Having proved the universal factorization property of the tensor product, we can now define the operation of contraction. Recall that if $\mathbf{v}\in\mathscr{V}$ and $\mathbf{v}^*\in\mathscr{V}^*$, then the scalar product $\left\langle\mathbf{v},\mathbf{v}^*\right\rangle$ is a scalar. Here we have used the canonical isomorphism given by (32.6). Of course, the operation

$$\langle\ ,\ \rangle\colon \mathscr{V}\times\mathscr{V}^*\to\mathscr{R}$$

is a bilinear function. By Theorem 34.1 $\langle\ ,\ \rangle$ can be factored through the tensor product

$$\otimes\colon \mathscr{V}\times\mathscr{V}^*\to\mathscr{T}_1^{\,1}(\mathscr{V})$$

i.e., there exists a linear map

$$\mathbf{C}\colon \mathscr{T}_1^{\,1}(\mathscr{V})\to\mathscr{R} \tag{34.6}$$

such that

$$\langle\ ,\ \rangle=\mathbf{C}\circ\otimes$$

or, equivalently,

$$\left\langle \mathbf{v}, \mathbf{v}^* \right\rangle = \mathbf{C}\left(\mathbf{v} \otimes \mathbf{v}^* \right) \tag{34.7}$$

for all $\mathbf{v} \in \mathscr{V}$ and $\mathbf{v}^* \in \mathscr{V}^*$. This linear function $\mathbf{C}$ is the simplest kind of contraction operation. It transforms the tensor space $\mathscr{T}_1^{\,1}(\mathscr{V})$ to the tensor space $\mathscr{T}_{1-1}^{1-1}(\mathscr{V}) = \mathscr{T}_0^{\,0}(\mathscr{V}) = \mathscr{R}$.

In general if $\mathbf{A} \in \mathscr{T}_q^{\,p}(\mathscr{V})$ is an arbitrary simple tensor, say

$$\mathbf{A} = \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_p \otimes \mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^q \tag{34.8}$$

then for each pair of integers (i, j), where $1 \le i \le p, 1 \le j \le q$, we seek a unique linear transformation

$$\mathbf{C}_j^{\,i} : \mathscr{T}_q^{\,p}(\mathscr{V}) \to \mathscr{T}_{q-1}^{\,p-1}(\mathscr{V}) \tag{34.9}$$

such that

$$\mathbf{C}_j^{\,i}\mathbf{A} = \left\langle \mathbf{v}^j, \mathbf{v}_i \right\rangle \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_{i-1} \otimes \mathbf{v}_{i+1} \otimes \cdots \otimes \mathbf{v}_p \otimes \mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^{j-1} \otimes \mathbf{v}^{j+1} \otimes \cdots \otimes \mathbf{v}^q \tag{34.10}$$

for all simple tensors $\mathbf{A}$. A more compact notation for the tensor product on the right-hand side is

$$\mathbf{v}_1 \otimes \cdots \otimes \breve{\mathbf{v}}_i \cdots \otimes \mathbf{v}_p \otimes \cdots \otimes \mathbf{v}^1 \otimes \cdots \breve{\mathbf{v}}^j \cdots \otimes \mathbf{v}^q \tag{34.11}$$

Since the representation of a simple tensor by a tensor product is not unique, the existence of such a linear transformation $\mathbf{C}_j^{\,i}$ is by no means obvious. However, we can prove that $\mathbf{C}_j^{\,i}$ does exist and is uniquely determined by the condition (34.10). Indeed, using the universal factorization property, we can prove the following theorem.

Theorem 34.2. A unique contraction operation $\mathbf{C}_j^{\,i}$ satisfying the condition (34.10) exists.

Proof. We define a multilinear transformation of the form (34.1) with $\mathscr{U} = \mathscr{T}_{q-1}^{\,p-1}(\mathscr{V})$ by

$$\begin{aligned} &\mathbf{Z}\left(\mathbf{v}_1, \ldots, \mathbf{v}_p, \mathbf{v}^1, \ldots \mathbf{v}^q \right) \\ &\quad = \left\langle \mathbf{v}^j, \mathbf{v}_i \right\rangle \mathbf{v}_1 \otimes \cdots \breve{\mathbf{v}}_i \cdots \otimes \mathbf{v}_p \otimes \mathbf{v}^1 \otimes \cdots \breve{\mathbf{v}}^j \cdots \otimes \mathbf{v}^q \end{aligned} \tag{34.12}$$

for all $\mathbf{v}_1,\ldots,\mathbf{v}_p \in \mathscr{V}$ and $\mathbf{v}^1,\ldots\mathbf{v}^q \in \mathscr{V}^*$. Then by Theorem 34.1 there exists a unique linear transformation $\mathbf{C}_j^i$ of the form (34.9) such that (34.10) holds, and thus the proof is complete.

Next we express the contraction operation in component form.

Theorem 34.3. If $\mathbf{A} \in \mathscr{T}_q^{\,p}(\mathscr{V})$ is an arbitrary tensor of order (p,q), then in component form relative to any dual bases $\{\mathbf{e}_i\}$ and $\{\mathbf{e}^i\}$ we have

$$\left(\mathbf{C}_j^i\mathbf{A}\right)^{k_1\ldots\breve{k}_i\ldots k_p}{}_{l_1\ldots\breve{l}_j\ldots l_q} = A^{k_1\ldots k_{i-1}tk_{i+1}\ldots k_p}{}_{l_1\ldots l_{j-1}tl_{j+1}\ldots l_q} \tag{34.13}$$

Proof. Since the contraction operation is linear applying it to the component form (33.20), we get

$$\mathbf{C}_j^i\mathbf{A} = A^{k_1\ldots k_p}{}_{l_1\ldots l_q}\left\langle \mathbf{e}^{l_j}, \mathbf{e}_{k_i}\right\rangle \mathbf{e}_{k_1}\otimes\cdots\breve{\mathbf{e}}_{k_i}\cdots\otimes\mathbf{e}_{k_p}\otimes\mathbf{e}^{l_1}\otimes\cdots\breve{\mathbf{e}}^{l_j}\cdots\otimes\mathbf{e}^{l_q} \tag{34.14}$$

which means nothing but the formula (34.13) because we have $\left\langle \mathbf{e}^{l_j}, \mathbf{e}_{k_i}\right\rangle = \delta_{k_i}^{l_j}$.

In particular, for the special case $\mathbf{C}$ given by (34.6) we have

$$\mathbf{C}(\mathbf{A}) = A_t^t \tag{34.15}$$

for all $\mathbf{A} \in \mathscr{T}_1^{\,1}(\mathscr{V})$. If we now make use of the canonical isomorphism

$$\mathscr{T}_1^{\,1}(\mathscr{V}) \cong \mathscr{L}(\mathscr{V};\mathscr{V})$$

we see that $\mathbf{C}$ coincides with the trace operation defined by (19.8).

Using the contraction operation, we can define a scalar product for tensors. If $\mathbf{A} \in \mathscr{T}_q^{\,p}(\mathscr{V})$ and $\mathbf{B} \in \mathscr{T}_p^{\,q}(\mathscr{V})$, the tensor product $\mathbf{A}\otimes\mathbf{B}$ is a tensor in $\mathscr{T}_{p+q}^{\,p+q}(\mathscr{V})$ and is defined by (33.24). We apply the contraction

$$\mathbf{C} = \underbrace{\mathbf{C}_1^1\circ\cdots\circ\mathbf{C}_1^1}_{q \text{ times}}\circ\underbrace{\mathbf{C}_{q+1}^1\circ\cdots\circ\mathbf{C}_{q+1}^1}_{p \text{ times}} \tag{34.16}$$

to $\mathbf{A}\otimes\mathbf{B}$, then the result is a scalar $\langle\mathbf{A},\mathbf{B}\rangle$, namely

$$\langle \mathbf{A}, \mathbf{B} \rangle \equiv \mathbf{C}(\mathbf{A} \otimes \mathbf{B}) \tag{34.17}$$

called the scalar product of $\mathbf{A}$ and $\mathbf{B}$. It is a simple matter to see that $\langle\ ,\ \rangle$ is a bilinear and definite function in the sense similar to those properties of the scalar product of $\mathbf{v}$ and $\mathbf{v}^*$ explained in Section 31. We can use the bilinear function

$$\langle\ ,\ \rangle : \mathscr{T}_q^{\,p}(\mathscr{V}) \times \mathscr{T}_p^{\,q}(\mathscr{V}) \to \mathscr{R} \tag{34.18}$$

to identify the space $\mathscr{T}_q^{\,p}(\mathscr{V})$ with the dual space $\mathscr{T}_p^{\,q}(\mathscr{V})^*$ of the space $\mathscr{T}_p^{\,q}(\mathscr{V})$ or equivalently, we can define the dual space $\mathscr{T}_p^{\,q}(\mathscr{V})^*$ abstractly as usual and then introduce a canonical isomorphism from $\mathscr{T}_q^{\,p}(\mathscr{V})$ to $\mathscr{T}_p^{\,q}(\mathscr{V})^*$ through (34.17). Thus we write

$$\mathscr{T}_q^{\,p}(\mathscr{V}) \cong \mathscr{T}_p^{\,q}(\mathscr{V})^* \tag{34.19}$$

Of course we shall also identify the second dual space $\mathscr{T}_p^{\,q}(\mathscr{V})^{**}$ with $\mathscr{T}_p^{\,q}(\mathscr{V})$ as explained in general in Section 32. Hence, we have

$$\mathscr{T}_q^{\,p}(\mathscr{V})^* \cong \mathscr{T}_p^{\,q}(\mathscr{V})^{**} \tag{34.20}$$

which follows also by interchanging p and q in (34.19). From (34.13) and (34.16), the scalar product $\langle \mathbf{A}, \mathbf{B} \rangle$ is given by the component form

$$\langle \mathbf{A}, \mathbf{B} \rangle = A^{i_1 \cdots i_p}{}_{j_1 \cdots j_q} B^{j_1 \cdots j_p}{}_{i_1 \cdots i_p} \tag{34.21}$$

relative to any dual bases $\{\mathbf{e}_i\}$ and $\{\mathbf{e}^i\}$ for $\mathscr{V}$ and $\mathscr{V}^*$.

Exercises

34.1 Show that

$$\langle \mathbf{A}, \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_q \otimes \mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^p \rangle = \mathbf{A}(\mathbf{v}^1, \cdots, \mathbf{v}^p, \mathbf{v}_1, \cdots, \mathbf{v}_q)$$

for all $\mathbf{A} \in \mathscr{T}_q^{\,p}(\mathscr{V})$, $\mathbf{v}_1, \cdots, \mathbf{v}_q \in \mathscr{V}$ and $\mathbf{v}^1, \cdots, \mathbf{v}^p \in \mathscr{V}^*$.

34.2 Give another proof of the quotient theorem in Exercise 33.7 by showing that the operation

$$\mathbf{J}\colon \mathscr{T}_s^{\,r}(\mathscr{V}) \to \mathscr{L}\left(\mathscr{T}_q^{\,p}(\mathscr{V}); \mathscr{T}_{q-r}^{\,p-s}(\mathscr{V})\right)$$

defined by

$$(\mathbf{JA})\mathbf{B} \equiv \mathbf{C}_1^1 \underbrace{\circ \cdots \circ \mathbf{C}_1^1}_{s \text{ times}} \circ \underbrace{\mathbf{C}_{s+1}^1 \circ \cdots \circ \mathbf{C}_{s+1}^1}_{r \text{ times}} (\mathbf{A} \otimes \mathbf{B})$$

for all $\mathbf{A} \in \mathscr{T}_s^{\,r}(\mathscr{V})$ and $\mathbf{B} \in \mathscr{T}_q^{\,p}(\mathscr{V})$ is an isomorphism. Since there are many such isomorphisms by choosing contractions of pairs of indices differently, we do not make any one of them a canonical isomorphism in general unless stated explicitly.

34.3 Use the universal factorization property and prove the existence and the uniqueness of the generalized transpose operation $\mathbf{T}_{pq}$ defined by (33.36) in Exercise 33.8. Hint: Require $\mathbf{T}_{pq}$ to be an automorphism of $\mathscr{T}_{p+q}(\mathscr{V})$ such that

$$\mathbf{T}_{pq}\left(\mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^p \otimes \mathbf{u}^1 \otimes \cdots \otimes \mathbf{u}^q\right) = \mathbf{u}^1 \otimes \cdots \otimes \mathbf{u}^q \otimes \mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^p \tag{34.22}$$

for all simple tensors $\mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^p \otimes \mathbf{u}^1 \otimes \cdots \otimes \mathbf{u}^q \in \mathscr{T}_{p+q}(\mathscr{V})$.

34.4 $\left\{\mathbf{e}_{i_1} \otimes \cdots \otimes \mathbf{e}_{i_p} \otimes \mathbf{e}^{j_1} \otimes \cdots \otimes \mathbf{e}^{j_q}\right\}$ is the product basis for $\mathscr{T}_q^{\,p}(\mathscr{V})$ induced by the dual bases $\{\mathbf{e}_i\}$ and $\{\mathbf{e}^i\}$; construct its dual basis in $\mathscr{T}_p^{\,q}(\mathscr{V})$ with respect to the scalar product defined by (34.17).

Section 35. Tensors on Inner Product Spaces

The main result of this section is that if $\mathscr{V}$ is equipped with a particular inner product, then we can identify the tensor spaces of the same total order by means of various canonical isomorphisms, a special case of these being the isomorphism $\mathbf{G}$ from $\mathscr{V}$ to $\mathscr{V}^*$, which we have introduced in Section 31 [cf. (31.5)].

Recall that the isomorphism $\mathbf{G}: \mathscr{V} \to \mathscr{V}^*$ is defined by the condition [cf. (31.6)]

$$\langle \mathbf{Gu}, \mathbf{v} \rangle = \langle \mathbf{Gv}, \mathbf{u} \rangle = (\mathbf{u} \cdot \mathbf{v}), \quad \mathbf{u}, \mathbf{v} \in \mathscr{V} \tag{35.1}$$

In general, if $\mathbf{A}$ is a simple, pure, contravariant tensor of order p, say

$$\mathbf{A} = \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_p$$

then we define the *pure covariant representation* of $\mathbf{A}$ to be the tensor

$$\mathbf{G}^p\mathbf{A} = \mathbf{Gv}_1 \otimes \cdots \otimes \mathbf{Gv}_p \in \mathscr{T}_p(\mathscr{V}) \tag{35.2}$$

By linearity, $\mathbf{G}^p$ can be extended to all of $\mathscr{T}^p(\mathscr{V})$. Equation (35.2) means that

$$\mathbf{G}^p\mathbf{A}(\mathbf{u}_1,\ldots,\mathbf{u}_p) = \langle \mathbf{Gv}_1, \mathbf{u} \rangle \cdots \langle \mathbf{Gv}_p, \mathbf{u}_p \rangle = (\mathbf{v}_1 \cdot \mathbf{u}_1) \cdots (\mathbf{v}_p \cdot \mathbf{u}_p) \tag{35.3}$$

for all $\mathbf{u}_1,\ldots,\mathbf{u}_p \in \mathscr{V}$. Equation(35.3) generalizes the condition (35.1) from $\mathbf{v}$ to $\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_p$. Of course, because the tensor product is not one-to-one, we have to show that the covariant tensor $\mathbf{G}^p\mathbf{A}$ does not depend on the representation of $\mathbf{A}$. This fact follows directly from the universal factorization of the tensor product. Indeed, if we define a p-linear map

$$\mathbf{Z}: \underbrace{\mathscr{V} \times \cdots \times \mathscr{V}}_{p \text{ times}} \to \mathscr{T}_p(\mathscr{V})$$

by

$$\mathbf{Z}(\mathbf{v}_1,\ldots,\mathbf{v}_p) \equiv \mathbf{Gv}_1 \otimes \cdots \otimes \mathbf{Gv}_p \tag{35.4}$$

then $\mathbf{Z}$ can be factored through the tensor product $\otimes$, i.e., there exists a unique linear transformation $\mathbf{G}^p$ from $\mathscr{T}^p(\mathscr{V})$ to $\mathscr{T}_p(\mathscr{V})$,

$$\mathbf{G}^p : \mathscr{T}^p(\mathscr{V}) \to \mathscr{T}_p(\mathscr{V}) \tag{35.5}$$

such that

$$\mathbf{Z} = \mathbf{G}^p \circ \otimes$$

or, equivalently,

$$\mathbf{Z}(\mathbf{v}_1, \ldots, \mathbf{v}_p) \equiv \mathbf{G}^p(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_p) = \mathbf{G}^p \mathbf{A} \tag{35.6}$$

Comparing (35.6) with (35.4) we see that $\mathbf{G}^p$ obeys the condition (35.2), and thus $\mathbf{G}^p \mathbf{A}$ is well defined by (35.3).

It is easy to verify that $\mathbf{G}^p$ is an isomorphism. Indeed $(\mathbf{G}^p)^{-1}$ is the unique linear transformation from $\mathscr{T}_p(\mathscr{V})$ to $\mathscr{T}^p(\mathscr{V})$ such that

$$(\mathbf{G}^p)^{-1}(\mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^p) = \mathbf{G}^{-1}\mathbf{v}^1 \otimes \cdots \otimes \mathbf{G}^{-1}\mathbf{v}^p \tag{35.7}$$

for all $\mathbf{v}^1, \ldots, \mathbf{v}^p \in \mathscr{V}^*$. Thus $\mathbf{G}^p$ makes $\mathscr{T}^p(\mathscr{V})$ and $\mathscr{T}_p(\mathscr{V})$ isomorphic, just as $\mathbf{G}$ makes $\mathscr{V}$ and $\mathscr{V}^*$ isomorphic. In fact, $\mathbf{G} = \mathbf{G}^1$.

Clearly, we can extend the preceding argument to mixed tensor spaces on $\mathscr{V}$ also. If $\mathbf{A}$ is a mixed simple tensor in $\mathscr{T}_q^{\ p}(\mathscr{V})$, say

$$\mathbf{A} = \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_p \otimes \mathbf{u}^1 \otimes \cdots \otimes \mathbf{u}^q \tag{35.8}$$

then we define the *pure covariant representation* of $\mathbf{A}$ to be the tensor

$$\mathbf{G}_q^p \mathbf{A} = \mathbf{G}\mathbf{v}_1 \otimes \cdots \otimes \mathbf{G}\mathbf{v}_p \otimes \mathbf{u}_1 \cdots \otimes \mathbf{u}^q \tag{35.9}$$

and $\mathbf{G}_q^p$ is an isomorphism from $\mathscr{T}_q^{\ p}(\mathscr{V})$ to $\mathscr{T}_{p+q}(\mathscr{V})$,

$$\mathbf{G}_q^p : \mathscr{T}_q^{\ p}(\mathscr{V}) \to \mathscr{T}_{p+q}(\mathscr{V}) \tag{35.10}$$

Indeed, its inverse is characterized by

$$\left(\mathbf{G}_q^p\right)^{-1}\left(\mathbf{v}^1\otimes\cdots\otimes\mathbf{v}^p\otimes\mathbf{u}^1\otimes\cdots\otimes\mathbf{u}^q\right)=\mathbf{G}^{-1}\mathbf{v}^1\otimes\cdots\otimes\mathbf{G}^{-1}\mathbf{v}^p\otimes\mathbf{u}^1\otimes\cdots\otimes\mathbf{u}^q \tag{35.11}$$

for all $\mathbf{v}^1,\ldots,\mathbf{v}^p,\mathbf{u}^1,\ldots,\mathbf{u}^q\in\mathscr{V}^*$. Clearly, by suitable compositions of the operations $\mathbf{G}_q^p$ and their inverses, we can define isomorphisms between tensor spaces of the same total order. For example, if p_1 and q_1 are another pair of integers such that

$$p_1+q_1=p+q$$

then $\mathscr{T}_q^p(\mathscr{V})$ is isomorphic to $\mathscr{T}_{q_1}^{p_1}(\mathscr{V})$ by the isomorphism

$$\left(\mathbf{G}_{q_1}^{p_1}\right)^{-1}\circ\mathbf{G}_q^p:\mathscr{T}_q^p(\mathscr{V})\to\mathscr{T}_{q_1}^{p_1}(\mathscr{V}) \tag{35.12}$$

In particular, if $q_1=0, p_1=p+q$, then for any $\mathbf{A}\in\mathscr{T}_q^p(\mathscr{V})$ the tensor $\left(\mathbf{G}^{p+q}\right)^{-1}\mathbf{G}_q^p\mathbf{A}\in\mathscr{T}^{p+q}(\mathscr{V})$ is called the *pure contravariant* representation of $\mathbf{A}$. For example, if $\mathbf{A}$ is given by (35.8), then its pure contravariant representation is given by

$$\left(\mathbf{G}^{p+q}\right)^{-1}\mathbf{G}_q^p\mathbf{A}=\mathbf{v}_1\otimes\cdots\otimes\mathbf{v}_p\otimes\mathbf{G}^{-1}\mathbf{u}^1\otimes\cdots\otimes\mathbf{G}^{-1}\mathbf{u}^q$$

We shall now express the isomorphism $\mathbf{G}_q^p$ in component form. If $\mathbf{A}\in\mathscr{T}_q^p(\mathscr{V})$ has the component representation

$$\mathbf{A}=A^{i_1\ldots i_p}{}_{j_1\ldots j_q}\mathbf{e}_{i_1}\otimes\cdots\otimes\mathbf{e}_{i_p}\otimes\mathbf{e}^{j_1}\otimes\cdots\otimes\mathbf{e}^{i_q} \tag{35.13}$$

then from (35.9) we obtain

$$\mathbf{G}_q^p\mathbf{A}=A^{i_1\ldots i_p}{}_{j_1\ldots j_q}\mathbf{G}\mathbf{e}_{i_1}\otimes\cdots\otimes\mathbf{G}\mathbf{e}_{i_p}\otimes\mathbf{e}^{j_1}\otimes\cdots\otimes\mathbf{e}^{j_q} \tag{35.14}$$

This result can be written as

$$\mathbf{G}_q^p\mathbf{A}=A^{i_1\ldots i_p}{}_{j_1\ldots j_q}\overline{\mathbf{e}}_{i_1}\otimes\cdots\otimes\overline{\mathbf{e}}_{i_p}\otimes\mathbf{e}^{j_1}\otimes\cdots\otimes\mathbf{e}^{j_q} \tag{35.15}$$

where $\{\overline{\mathbf{e}}_i\}$ is a basis in $\mathscr{V}^*$ reciprocal to $\{\mathbf{e}^i\}$, i.e.,

$$\overline{\mathbf{e}}_i\cdot\mathbf{e}^j=\delta_i^j \tag{35.16}$$

since it follows from (19.1), (12.6), and (35.1) that we have

$$\overline{\mathbf{e}}_i = \mathbf{G}\mathbf{e}_i = e_{ji}\mathbf{e}^j \tag{35.17}$$

where

$$e_{ji} = \mathbf{e}_j \cdot \mathbf{e}_i \tag{35.18}$$

Substituting (35.17) into (35.15), we obtain

$$\mathbf{G}_q^p\mathbf{A} = A_{i_1\cdots i_p j_1\cdots j_q}\mathbf{e}^{i_1}\otimes\cdots\otimes\mathbf{e}^{i_p}\otimes\mathbf{e}^{j_1}\otimes\cdots\otimes\mathbf{e}^{j_q} \tag{35.19}$$

where

$$A_{i_1\cdots i_p j_1\cdots j_q} = e_{i_1k_1}\cdots e_{i_pk_p}A^{k_1\cdots k_p}{}_{j_1\cdots j_q} \tag{35.20}$$

This equation illustrates the component form of the isomorphism $\mathbf{G}_q^p$. It has the effect of *lowering* the first *p* superscripts on the components $\left\{A^{i_1\cdots i_p}{}_{j_1\cdots j_q}\right\}$. Thus $\mathbf{A}\in\mathscr{T}_q^{\,p}(\mathscr{V})$ and $\mathbf{G}_q^p\mathbf{A}\in\mathscr{T}_{p+q}(\mathscr{V})$ have the same components if the bases in (35.13) and (35.15) are used. On the other hand, if the usual product basis for $\mathscr{T}_q^{\,p}(\mathscr{V})$ and $\mathscr{T}_{p+q}(\mathscr{V})$ are used, as in (35.13) and (35.19), the components of $\mathbf{A}$ and $\mathbf{G}_q^p\mathbf{A}$ are related by (35.20).

Since $\mathbf{G}_q^p$ is an isomorphism, its inverse exists and is a linear transformation from $\mathscr{T}_{p+q}(\mathscr{V})$ and $\mathscr{T}_q^{\,p}(\mathscr{V})$. If $\mathbf{A}\in\mathscr{T}_{p+q}(\mathscr{V})$ has the component representation

$$\mathbf{A} = A_{i_1\cdots i_p j_1\cdots j_q}\mathbf{e}^{i_1}\otimes\cdots\otimes\mathbf{e}^{i_p}\otimes\mathbf{e}^{j_1}\otimes\cdots\otimes\mathbf{e}^{j_q} \tag{35.21}$$

then from (35.11)

$$\left(\mathbf{G}_q^p\right)^{-1}\mathbf{A} = A_{i_1\cdots i_p j_1\cdots j_q}\mathbf{G}^{-1}\mathbf{e}^{i_1}\otimes\cdots\otimes\mathbf{G}^{-1}\mathbf{e}^{i_p}\otimes\mathbf{e}^{j_1}\otimes\cdots\otimes\mathbf{e}^{i_q} \tag{35.22}$$

By the same argument as before, this formula can be rewritten as

$$\left(\mathbf{G}_q^p\right)^{-1}\mathbf{A} = A_{i_1\cdots i_p j_1\cdots j_q}\overline{\mathbf{e}}^{i_1}\otimes\cdots\otimes\overline{\mathbf{e}}^{i_p}\otimes\mathbf{e}^{j_1}\otimes\cdots\otimes\mathbf{e}^{j_q} \tag{35.23}$$

where $\{\overline{\mathbf{e}}^i\}$ is a basis of $\mathscr{V}$ reciprocal to $\{\mathbf{e}_i\}$, or equivalently as

$$\left(\mathbf{G}_q^p\right)^{-1}\mathbf{A} = A^{i_1\cdots i_p}{}_{j_1\cdots j_q}\mathbf{e}_{i_1}\otimes\cdots\otimes\mathbf{e}_{i_p}\otimes\mathbf{e}^{j_1}\otimes\cdots\otimes\mathbf{e}^{j_q} \tag{35.24}$$

where

$$A^{i_1\cdots i_p}{}_{j_1\cdots j_q} = e^{i_1k_1}\cdots e^{i_pk_p}A_{k_1\cdots k_p j_1\cdots j_q} \tag{35.25}$$

Here of course $\left[e^{ij}\right]$ is the inverse matrix of $\left[e_{ij}\right]$ and is given by

$$e^{ij} = \mathbf{e}^i\cdot\mathbf{e}^j \tag{35.26}$$

since

$$\overline{\mathbf{e}}^i = \mathbf{G}^{-1}\mathbf{e}^i = e^{ij}\mathbf{e}_j \tag{35.27}$$

Equation (35.25) illustrates the component form of the isomorphism $\left(\mathbf{G}_q^p\right)^{-1}$. It has the effect of raising the first p subscripts on the components $\left\{A_{i_1\cdots i_p j_1\cdots j_q}\right\}$.

Combining (35.20) and (35.25), we see that the component form of the isomorphism $\left(\mathbf{G}_{q_1}^{p_1}\right)^{-1}\circ\mathbf{G}_q^p$ in (35.12) is given by raising the first $p_1 - p$ subscripts of $\mathbf{A}\in\mathscr{T}_q^p(\mathscr{V})$ if $p_1 > p$, or by lowering the last $p - p_1$ superscripts of $\mathbf{A}$ if $p > p_1$. Thus if $\mathbf{A}$ has the component form (35.13), then $\left(\mathbf{G}_{q_1}^{p_1}\right)^{-1}\mathbf{G}_q^p\mathbf{A}$ has the component form

$$\left(\mathbf{G}_{q_1}^{p_1}\right)^{-1}\mathbf{G}_q^p\mathbf{A} = A^{i_1\cdots i_{p_1}}{}_{j_1\cdots j_{q_1}}\mathbf{e}_{i_1}\otimes\cdots\otimes\mathbf{e}_{i_{p_1}}\otimes\mathbf{e}^{j_1}\otimes\cdots\otimes\mathbf{e}^{j_{q_1}} \tag{35.28}$$

where

$$A^{i_1\cdots i_{p_1}}{}_{j_1\cdots j_{q_1}} = A^{i_1\cdots i_p}{}_{k_1\cdots k_{p_1-p}j_1\cdots j_{q_1}}\,e^{i_{p+1}k_1}\cdots e^{i_{p_1}k_{p_1-p}} \quad\text{if}\quad p_1 > p \tag{35.29}$$

and

$$A^{i_1\cdots i_{p_1}}{}_{j_1\cdots j_{q_1}} = A^{i_1\cdots i_{p_1}\cdots k_{p-p_1}}{}_{j_{p-p_1+1}\cdots j_{q_1}}\,e_{k_1j_1}\cdots e_{k_{p-p_1}j_{p-p_1}} \quad\text{if}\quad p > p_1 \tag{35.30}$$

For example, if $\mathbf{A} \in \mathscr{T}_2^1(\mathscr{V})$ has the representation

$$\mathbf{A} = A^i{}_{jk}\mathbf{e}_i \otimes \mathbf{e}^j \otimes \mathbf{e}^k$$

then the covariant representation of $\mathbf{A}$ is

$$\mathbf{G}_2^1\mathbf{A} = A^i{}_{jk}\overline{\mathbf{e}}_i \otimes \mathbf{e}^j \otimes \mathbf{e}^k = A^i{}_{jk}e_{il}\mathbf{e}^l \otimes \mathbf{e}^j \otimes \mathbf{e}^k = A_{ijk}\mathbf{e}^l \otimes \mathbf{e}^j \otimes \mathbf{e}^k$$

the contravariant representation of $\mathbf{A}$ is

$$\left(\mathbf{G}^3\right)^{-1}\mathbf{G}_2^1\mathbf{A} = A^i{}_{jk}\mathbf{e}_i \otimes \overline{\mathbf{e}}^j \otimes \overline{\mathbf{e}}^k = A^i{}_{jk}e^{jl}e^{km}\mathbf{e}_i \otimes \mathbf{e}_l \otimes \mathbf{e}_m \equiv A^{ilm}\mathbf{e}_i \otimes \mathbf{e}_l \otimes \mathbf{e}_m$$

and the representation of $\mathbf{A}$ in $\mathscr{T}_1^2(\mathscr{V})$ is

$$\left(\mathbf{G}_1^2\right)^{-1}\mathbf{G}_2^1\mathbf{A} = A^i{}_{jk}\mathbf{e}_i \otimes \overline{\mathbf{e}}^j \otimes \mathbf{e}^k = A^i{}_{jk}e^{jl}\mathbf{e}_i \otimes \mathbf{e}_l \otimes \mathbf{e}^k \equiv A^{il}{}_k\mathbf{e}_i \otimes \mathbf{e}_l \otimes \mathbf{e}^k$$

etc. These formulas follow from (35.20), (35.24), and (35.29).

Of course, we can apply the operations of lowering and raising to indices at any position, e.g., we can define a "component" $A^{i_1}{}_{j_2\ \cdots\ j_b\cdots}^{\ i_3\cdots i_a\cdots}$ for $\mathbf{A} \in \mathscr{T}_r(\mathscr{V})$ by

$$A^{i_1}{}_{j_2\ \cdots\ j_b\cdots}^{\ i_3\cdots i_a\cdots} \equiv A_{j_1 j_2 \cdots j_r}e^{i_1 j_1}e^{i_3 j_3}\cdots e^{i_a j_a}\cdots \tag{35.31}$$

However, for simplicity, we have not yet assigned any symbol to such representations whose "components" have an irregular arrangement of superscripts and subscripts. In our notation for the component representation of a tensor $\mathbf{A} \in \mathscr{T}_q^p(\mathscr{V})$ the contravariant superscripts always come first, so that the components of $\mathbf{A}$ are written as $A^{i_1\cdots i_p}{}_{j_1\cdots j_q}$ as shown in (35.13), not as $A_{j_1\cdots j_q}{}^{i_1\cdots i_p}$ or as any other rearrangement of positions of the superscripts $i_1\ldots i_p$ and the subscripts $j_1\ldots j_q$, such as the one defined by (35.31). In order to indicate precisely the position of the contravariant indices and the covariant indices in the irregular component form, we may use, for example, the notation

$$\underset{(1)}{\mathscr{V}} \otimes \underset{(2)}{\mathscr{V}^*} \otimes \underset{(3)}{\mathscr{V}} \otimes \cdots \otimes \underset{(a)}{\mathscr{V}} \otimes \cdots \otimes \underset{(b)}{\mathscr{V}^*} \otimes \cdots \tag{35.32}$$

for the tensor space whose elements have components of the form on the left-hand side of (35.31), where the order of $\mathscr{V}$ and $\mathscr{V}^*$ in (35.32) are the same as those of the contravariant and

the covariant indices, respectively, in the component form. In particular, the simple notation $\mathscr{T}_q^{\,p}(\mathscr{V})$ now corresponds to

$$\underset{(1)}{\mathscr{V}}\otimes\cdots\otimes\underset{(p)}{\mathscr{V}}\otimes\underset{(p+1)}{\mathscr{V}^*}\otimes\cdots\otimes\underset{(p+q)}{\mathscr{V}^*} \tag{35.33}$$

Since the irregular tensor spaces, such as the one in (35.32), are not convenient to use, we shall avoid them as much as possible, and we shall not bother to generalize the notation $\mathbf{G}_q^p$ to isomorphisms from the irregular tensor spaces to their corresponding pure covariant representations.

So far, we have generalized the isomorphism $\mathbf{G}$ for vectors to the isomorphisms $\mathbf{G}_q^p$ for tensors of type (p,q) in general. Equation (35.1) for $\mathbf{G}$, however, contains more information than just the fact that $\mathbf{G}$ is an isomorphism. If we read that equation reversely, we see that $\mathbf{G}$ can be used to compute the inner product on $\mathscr{V}$. Indeed, we can rewrite that equation as

$$\mathbf{v}\cdot\mathbf{u}=\langle\mathbf{G}\mathbf{v},\mathbf{u}\rangle\equiv\langle\mathbf{G}^{-1}\mathbf{v},\mathbf{u}\rangle \tag{35.34}$$

This idea can be generalized easily to tensors. For example, if $\mathbf{A}$ and $\mathbf{B}$ are tensors in $\mathscr{T}^r(\mathscr{V})$, then we can define an inner product $\mathbf{A}\cdot\mathbf{B}$ by

$$\mathbf{A}\cdot\mathbf{B}\equiv\langle\mathbf{G}^r\mathbf{A},\mathbf{B}\rangle \tag{35.35}$$

We leave the proof to the reader that (35.35) actually defines an inner product $\mathscr{T}^r(\mathscr{V})$. By use of (34.21) and (35.20), it is possible to write (35.35) in the component form

$$\mathbf{A}\cdot\mathbf{B}=A_{j_1\ldots j_r}e^{i_1j_1}\cdots e^{i_rj_r}B_{i_1\ldots i_r} \tag{35.36}$$

or, equivalently, in the forms

$$\mathbf{A}\cdot\mathbf{B}=\begin{cases}A^{i_1\ldots i_r}B_{i_1\ldots i_r}\\ A^{i_1\ldots i_{r-1}}{}_{j_r}e^{i_rj_r}B_{i_1\ldots i_r}\\ A^{i_1\ldots i_r}B^{j_1}{}_{i_2\ldots i_r}e_{i_1j_1},\qquad\text{etc.}\end{cases} \tag{35.37}$$

Equation (35.37) suggests definitions of inner products for other tensor spaces besides $\mathscr{T}^r(\mathscr{V})$. If $\mathbf{A}$ and $\mathbf{B}$ are in $\mathscr{T}_q^{\,p}(\mathscr{V})$, we define $\mathbf{A}\cdot\mathbf{B}$ by

$$\mathbf{A}\cdot\mathbf{B}\equiv\left(\mathbf{G}^{p+q}\right)^{-1}\mathbf{G}_q^p\mathbf{A}\cdot\left(\mathbf{G}^{p+q}\right)^{-1}\mathbf{G}_q^p\mathbf{B}=\left\langle\mathbf{G}_q^p\mathbf{A},\left(\mathbf{G}^{p+q}\right)^{-1}\mathbf{G}_q^p\mathbf{B}\right\rangle \tag{35.38}$$

Again, we leave it as an exercise to the reader to establish that (35.38) does define an inner product on $\mathscr{T}_q^{\,p}(\mathscr{V})$; moreover, the inner product can be written in component form by (35.37) (35.37) also.

In section 32, we pointed out that if we agree to use a particular inner product, then the isomorphism $\mathbf{G}$ can be regarded as canonical. The formulas of this section clearly indicate the desirability of this procedure if for no other reason than notational simplicity. Thus, from this point on, we shall identify the dual space $\mathscr{V}^*$ with $\mathscr{V}$, i.e.,

$$\mathscr{V}\cong\mathscr{V}^*$$

by suppressing the symbol $\mathbf{G}$. Then in view of (35.2) and (35.7), we shall suppress the symbol $\mathbf{G}_q^p$ also. Thus, we shall identify all tensor spaces of the same total order, i.e., we write

$$\mathscr{T}^r(\mathscr{V})\cong\mathscr{T}_1^{\,r\text{-}1}(\mathscr{V})\cong\cdots\cong\mathscr{T}_r(\mathscr{V}) \tag{35.39}$$

By this procedure we can replace scalar products throughout our formulas by inner products according to the formulas (31.9) and (35.38).

The identification (35.39) means that, as long as the total order of a tensor is given, it is no longer necessary to specify separately the contravariant order and the covariant order. These separate orders will arise only when we select a particular component representation of the tensor. For example, if $\mathbf{A}$ is of total order r, then we can express $\mathbf{A}$ by the following different component forms:

$$\begin{aligned}\mathbf{A}&=A^{i_1\cdots i_r}\mathbf{e}_{i_1}\otimes\cdots\otimes\mathbf{e}_{i_r}\\&=A^{i_1\cdots i_{r-1}}{}_{j_r}\mathbf{e}_{i_1}\otimes\cdots\otimes\mathbf{e}_{i_{r-1}}\otimes\mathbf{e}^{j_r}\\&\;\;\vdots\\&=A_{j_1\cdots j_r}\mathbf{e}^{j_1}\otimes\cdots\otimes\mathbf{e}^{j_r}\end{aligned} \tag{35.40}$$

where the placement of the indices indicates that the first form is the pure contravariant representation, the last form is the pure covariant representation, while the intermediate forms are various mixed tensor representations, all having the same total order r. Of course, the various representations in (35.40) are related by the formulas (35.20), (35.25), (35.29), and (35.30). In fact, if (35.31) is used, we can even represent $\mathbf{A}$ by an irregular component form such as

$$\mathbf{A}=A^{i_1\ i_3\cdots i_a\ \cdots}_{\ \ j_2\ \cdots\ j_b\cdots}\mathbf{e}_{i_1}\otimes\mathbf{e}^{j_2}\otimes\mathbf{e}_{i_3}\otimes\cdots\otimes\mathbf{e}_{i_a}\otimes\cdots\otimes\mathbf{e}^{j_b}\otimes\cdots \tag{35.41}$$

provided that the total order is unchanged.

The contraction operator defined in Section 34 can now be rewritten in a form more convenient for tensors defined on an inner product space. If $\mathbf{A}$ is a simple tensor of (total) order $r, r \geq 2$, with the representation

$$\mathbf{A} = \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_r$$

then $\mathbf{C}_{ij}\mathbf{A}$, where $1 \leq i < j \leq r$, is a simple tensor of (total) order $r-2$ defined by (34.10),

$$\mathbf{C}_{ij}\mathbf{A} \equiv (\mathbf{v}_i \cdot \mathbf{v}_j)\mathbf{v}_1 \otimes \cdots \breve{\mathbf{v}}_i \cdots \breve{\mathbf{v}}_j \cdots \otimes \mathbf{v}_r \tag{35.42}$$

By linearity, $\mathbf{C}_{ij}$ can be extended to all tensors of order r. If $\mathbf{A} \in \mathscr{T}_r(\mathscr{V})$ has the representation

$$\mathbf{A} = A_{k_1 \ldots k_r} \mathbf{e}^{k_1} \otimes \cdots \otimes \mathbf{e}^{k_r}$$

then by (35.42) and (35.26)

$$\begin{aligned} \mathbf{C}_{ij}\mathbf{A} &= A_{k_1 \ldots k_r} \mathbf{C}_{ij} \left(\mathbf{e}^{k_1} \otimes \cdots \otimes \mathbf{e}^{k_r} \right) \\ &= e^{k_i k_j} A_{k_1 \ldots k_r} \mathbf{e}^{k_1} \otimes \cdots \breve{\mathbf{e}}^{k_i} \cdots \breve{\mathbf{e}}^{k_j} \cdots \otimes \mathbf{e}^{k_r} \\ &= A_{k_1 \ldots \overset{k_j}{\underset{i}{\ldots}} \ldots \overset{k_j}{\underset{j}{\ldots}} k_r} \mathbf{e}^{k_1} \otimes \cdots \breve{\mathbf{e}}_{k_i} \cdots \breve{\mathbf{e}}^{k_j} \cdots \otimes \mathbf{e}^{k_r} \end{aligned} \tag{35.43}$$

It follows from the definition (35.42) that the complete contraction operator $\mathbf{C}$ [cf. (35.16)]

$$\mathbf{C} : \mathscr{T}_{2r}(\mathscr{V}) \to \mathscr{R}$$

can be written

$$\mathbf{C} = \mathbf{C}_{12} \circ \mathbf{C}_{13} \circ \cdots \circ \mathbf{C}_{1(r+1)} \tag{35.44}$$

Also, if $\mathbf{A}$ and $\mathbf{B}$ are in $\mathscr{T}_r(\mathscr{V})$, it is easily establish that [cf. (34.17)]

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{C}(\mathbf{A} \otimes \mathbf{B}) \tag{35.45}$$

In closing this chapter, it is convenient for later use to record certain formulas here. The identity automorphism $\mathbf{I}$ of $\mathscr{V}$ corresponds to a tensor of order 2. Its pure covariant representation is simply the inner product

$$\mathbf{I}(\mathbf{u},\mathbf{v}) = \mathbf{u}\cdot\mathbf{v} \tag{35.46}$$

The tensor $\mathbf{I}$ can be represented in any of the following forms:

$$\mathbf{I} = e_{ij}\mathbf{e}^i \otimes \mathbf{e}^j = e^{ij}\mathbf{e}_i \otimes \mathbf{e}_j = \delta^i_j \mathbf{e}_i \otimes \mathbf{e}^j = \delta^j_i \mathbf{e}^i \otimes \mathbf{e}_j \tag{35.47}$$

There is but one contraction for $\mathbf{I}$, namely

$$\mathbf{C}_{12}\mathbf{I} = \operatorname{tr}\mathbf{I} = \delta^i_i = N \tag{35.48}$$

In view of (35.46), the identity tensor is also called the *metric tensor* of the inner product space.

Exercises

35.1 Show that (35.45) defines an inner product on $\mathscr{T}_r(\mathscr{V})$

35.2 Show that the formula (35.38) can be rewritten as

$$\mathbf{A}\cdot\mathbf{B} = \left\langle \left(\mathbf{G}^q_p\right)^{-1}\mathbf{T}_{pq}\mathbf{G}^p_q\mathbf{A},\mathbf{B}\right\rangle \tag{35.49}$$

where $\mathbf{T}_{pq}$ is the generalized transpose operator defined in Exercises 33.8 and 34.3. In particular, if $\mathbf{A}$ and $\mathbf{B}$ are second order tensors, then (35.49) reduces to the more familiar formula:

$$\mathbf{A}\cdot\mathbf{B} = \operatorname{tr}\left(\mathbf{A}^T\mathbf{B}\right) \tag{35.50}$$

35.3 Show that the linear transformation

$$\mathbf{G}^p : \mathscr{T}^p(\mathscr{V}) \to \mathscr{T}_p(\mathscr{T})$$

defined by (35.2) can also be characterized by

$$\left(\mathbf{G}^p\mathbf{A}\right)\left(\mathbf{u}_1,\ldots,\mathbf{u}_p\right) \equiv \mathbf{A}\left(\mathbf{G}\mathbf{u}_1,\ldots,\mathbf{G}\mathbf{u}_p\right) \tag{35.51}$$

for all $\mathbf{A} \in \mathscr{T}^p(V)$ and all $\mathbf{u}_1, \ldots, \mathbf{u}_p \in \mathscr{V}$.

35.4 Show that if $\mathbf{A}$ is a second-order tensor, then

$$\mathbf{A} = \mathbf{C}_{23}(\mathbf{I} \otimes \mathbf{A})$$

What is the component form of this identity?

Chapter 8

EXTERIOR ALGEBRA

The purpose of this chapter is to formulate enough machinery to define the determinant of an endomorphism in a component free fashion, to introduce the concept of an orientation for a vector space, and to establish a certain isomorphism which generalizes the classical operation of vector product to a *N*-dimensional vector space. For simplicity, we shall assume throughout this chapter that the vector space, and thus its associated tensor spaces, are equipped with an inner product. Further, for definiteness, we shall use only the pure covariant representation; transformations into other representations shall be explained in Section 42.

Section 36 Skew-Symmetric Tensors and Symmetric Tensors

If $\mathbf{A} \in \mathscr{T}_r(\mathscr{V})$ and σ is a given permutation of $\{1,...,r\}$, then we can define a new tensor $\mathbf{T}_\sigma \mathbf{A} \in \mathscr{T}_r(\mathscr{V})$ by the formula

$$\mathbf{T}_\sigma \mathbf{A}(\mathbf{v}_1,..,\mathbf{v}_r) = \mathbf{A}(\mathbf{v}_{\sigma(1)},..,\mathbf{v}_{\sigma(r)}) \tag{36.1}$$

for all $\mathbf{v}_1,..,\mathbf{v}_r \in \mathscr{V}$. For example, the generalized transpose operation $\mathbf{T}_{pq}$ defined by (33.36) in Exercise 33.8 is a special case of $\mathbf{T}_\sigma$ with σ given by

$$\sigma = \begin{pmatrix} 1 & \cdot & \cdot & \cdot & q & q+1 & \cdot & \cdot & \cdot & q+p \\ p+1 & \cdot & \cdot & \cdot & p+q & 1 & \cdot & \cdot & \cdot & p \end{pmatrix}$$

Naturally, we call $\mathbf{T}_\sigma \mathbf{A}$ the σ*-transpose* of $\mathbf{A}$ for any σ in general. We have obtained the components of the transpose $\mathbf{T}_{pq}$ in Exercise 33.8. For an arbitrary permutation σ the components of $\mathbf{T}_\sigma \mathbf{A}$ are related to those of $\mathbf{A}$ by

$$(\mathbf{T}_\sigma \mathbf{A})_{i_1 \ldots i_r} = A_{i_{\sigma(1)} \ldots i_{\sigma(r)}} \tag{36.2}$$

Using the same argument as that of Exercise 34.3, we can characterize $\mathbf{T}_\sigma$ by the condition that

$$\mathbf{T}_\sigma(\mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^r) = \mathbf{v}^{\sigma^{-1}(1)} \otimes \cdots \otimes \mathbf{v}^{\sigma^{-1}(r)} \tag{36.3}$$

for all simple tensors $\mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^r \in \mathscr{T}_r(\mathscr{V})$

If $\mathbf{T}_\sigma\mathbf{A} = \mathbf{A}$ for all permutations σ $\mathbf{A}$ is said to be *(comp1etely) symmetric*. On the other hand if $\mathbf{T}_\sigma\mathbf{A} = \varepsilon_\sigma\mathbf{A}$ for all permutations σ, where ε_σ denotes the parity of σ defined in Section 20, then $\mathbf{A}$ is said to be *(complete1y) skew-symmetric*. For example, the identity tensor $\mathbf{I}$ given by (35.30) is a symmetric second-order tensor, while the tensor $\mathbf{u}\otimes\mathbf{v} - \mathbf{v}\otimes\mathbf{u}$, for any vectors $\mathbf{u},\mathbf{v} \in \mathscr{V}$, is clearly a skew-symmetric second-order tensor. We shall denote by $\hat{\mathscr{T}}_r(\mathscr{V})$ the set of all skew-symmetric tensors in $\mathscr{T}_r(\mathscr{V})$. We leave it to the reader to establish the fact that $\hat{\mathscr{T}}_r(\mathscr{V})$ is a subspace of $\mathscr{T}_r(\mathscr{V})$. Elements of $\hat{\mathscr{T}}_r(\mathscr{V})$ are often called *r-vectors or r-forms*.

Theorem 36.1. An element $\mathbf{A} \in \hat{\mathscr{T}}_r(\mathscr{V})$ assumes the value $\mathbf{0}$ if any two of its variables coincide.

Proof: We wish to establish that

$$\mathbf{A}(\mathbf{v}_1,...,\mathbf{v},...,\mathbf{v},...,\mathbf{v}_r) = \mathbf{0} \tag{36.4}$$

This result is a special case of the formula

$$\mathbf{A}(\mathbf{v}_1,...,\mathbf{v}_s,...,\mathbf{v}_t,...,\mathbf{v}_r) = -\mathbf{A}(\mathbf{v}_1,...,\mathbf{v}_t,...,\mathbf{v}_s,...,\mathbf{v}_r) = \mathbf{0} \tag{36.5}$$

which follows by the fact that $\varepsilon_\sigma = -1$ for the permutation which switches the pair of indices (s,t)) while leaving the remaining indices unchanged. If we take $\mathbf{v} = \mathbf{v}_s = \mathbf{v}_t$ in (36.5), then (36.4) follows.

Corollary. An element $\mathbf{A} \in \hat{\mathscr{T}}_r(\mathscr{V})$ assumes the value zero if it is evaluated on a linearly dependent set of vectors.

This corollary generalizes the result of the preceding theorem but is itself also a direct consequence of that theorem, for if $\{\mathbf{v}_1,...,\mathbf{v}_r\}$ is a linearly dependent set, then at least one of the vectors can be expressed as is a linear combination of the remaining ones. From the *r*-linearity of $\mathbf{A}$, $\mathbf{A}(\mathbf{v}_1,...,\mathbf{v}_r)$ can then be written as the linear combination of quantities which are all equal to zero because which are the values of $\mathbf{A}$ at arguments having at least two equal variables.

Corollary. If . r is greater than N, the dimension of $\mathscr{V}$, then $\hat{\mathscr{T}}_r(\mathscr{V}) = \{\mathbf{0}\}$.

This corollary follows from the last corollary and the fact that every set of more than N vectors in a N- dimensional space is linearly dependent.

Exercises

36.1 Show that the set of symmetric tensors of order r forms a subspace of $\mathscr{T}_r(\mathscr{V})$.

36.2 Show that

$$\mathbf{T}_{\sigma\tau} = \mathbf{T}_\sigma \mathbf{T}_\tau$$

for all permutations σ and τ. Also, show that $\mathbf{T}_\sigma$ is the identity automorphism of $\mathscr{T}_r(\mathscr{V})$ if and only if σ is the identity permutation.

Section 37 The Skew-Symmetric Operator

In this section we shall construct a projection from $\mathscr{T}_r(\mathscr{V})$ into $\mathscr{T}_r(\mathscr{V})$. This projection is called the *skew-symmetric operator.* If $\mathbf{A} \in \mathscr{T}_r(\mathscr{V})$, we define the *skew-symmetric projection* $\mathbf{K}_r\mathbf{A}$ of $\mathbf{A}$ by

$$\mathbf{K}_r\mathbf{A} = \frac{1}{r!}\sum_{\sigma} \varepsilon_\sigma \mathbf{T}_\sigma \mathbf{A} \tag{37.1}$$

where the summation is taken over all permutations σ of $\{1,...,r\}$. The endomorphism

$$\mathbf{K}_r : \mathscr{T}_r(\mathscr{V}) \to \mathscr{T}_r(\mathscr{V}) \tag{37.2}$$

defined in this way is called the *skew-symmetric operator*.

Before showing that $\mathbf{K}_r$ has the desired properties, we give one example first. For simplicity, let us choose $r = 2$. Then there are only two permutations of $\{1,2\}$, namely

$$\sigma = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}, \quad \sigma = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \tag{37.3}$$

and their parities are

$$\varepsilon_\sigma = 1 \quad \text{and} \quad \varepsilon_\sigma = -1 \tag{37.4}$$

Substituting (37.3) and (37.4) into (37.1), we get

$$(\mathbf{K}_2\mathbf{A})(\mathbf{v}_1, \mathbf{v}_2) = \frac{1}{2}\big(\mathbf{A}(\mathbf{v}_1, \mathbf{v}_2) + \mathbf{A}(\mathbf{v}_2, \mathbf{v}_1)\big) \tag{37.5}$$

for all $\mathbf{A} \in \mathscr{T}_2(\mathscr{V})$ and $\mathbf{v}_1, \mathbf{v}_2 \in \mathscr{V}$. In particular, if $\mathbf{A}$ is skew-symmetric, namely

$$\mathbf{A}(\mathbf{v}_2, \mathbf{v}_1) = -\mathbf{A}(\mathbf{v}_1, \mathbf{v}_2) \tag{37.6}$$

then (37.5) reduces to

$$(\mathbf{K}_2\mathbf{A})(\mathbf{v}_1, \mathbf{v}_2) = \mathbf{A}(\mathbf{v}_1, \mathbf{v}_2)$$

or, equivalently,

$$\mathbf{K}_2\mathbf{A} = \mathbf{A}, \qquad \mathbf{A} \in \hat{\mathscr{T}}_2(\mathscr{V}) \tag{37.7}$$

Since from (37.5), $\mathbf{K}_2\mathbf{A} \in \hat{\mathscr{T}}_2(\mathscr{V})$ for any $\mathbf{A} \in \mathscr{T}_2(\mathscr{V})$????, by (37.7) we then have

$$\mathbf{K}_2(\mathbf{K}_2\mathbf{A}) = \mathbf{K}_2\mathbf{A}$$

or, equivalently

$$\mathbf{K}_2^2 = \mathbf{K}_2 \tag{37.8}$$

which means that $\mathbf{K}_2$ is a projection. Hence, for second order tensors, $\mathbf{K}_2$ has the desired properties. We shall now prove the same for tensors in general.

Theorem 37.1. Let $\mathbf{K}_r : \mathscr{T}_r(\mathscr{V}) \to \mathscr{T}_r(\mathscr{V})$ be defined by (37.1). Then the range of $\mathbf{K}_r$ is $\hat{\mathscr{T}}_r(\mathscr{V})$, namely

$$R(\mathbf{K}_r) = \hat{\mathscr{T}}_r(\mathscr{V}) \tag{37.9}$$

Moreover, the restriction of $\mathbf{K}_r$ on $\hat{\mathscr{T}}_r(\mathscr{V})$ is the identity automorphism of $\hat{\mathscr{T}}_r(\mathscr{V})$, i.e.,

$$\mathbf{K}_r\mathbf{A} = \mathbf{A}, \qquad \mathbf{A} \in \hat{\mathscr{T}}_r(\mathscr{V}) \tag{37.10}$$

Proof. We prove the equation (37.10) first. If $\mathbf{A} \in \hat{\mathscr{T}}_r(\mathscr{V})$, then by definition we have

$$\mathbf{T}_\sigma\mathbf{A} = \varepsilon_\sigma\mathbf{A} \tag{37.11}$$

for all permutations σ. Substituting (37.11) into (37.1), we get

$$\mathbf{K}_r\mathbf{A} = \frac{1}{r!}\sum_\sigma \varepsilon_\sigma^2\mathbf{A} = \frac{1}{r!}(r!)\mathbf{A} = \mathbf{A} \tag{37.12}$$

Here we have used the familiar fact that there are a total of $r!$ permutations for r numbers $\{1,...,r\}$.

Having proved (37.10), we can conclude immediately that

$$R(\mathbf{K}_r) \supset \hat{\mathscr{T}}_r(\mathscr{V}) \tag{37.13}$$

since $\hat{\mathscr{T}}_r(\mathscr{V})$ is a subspace of $\mathscr{T}_r(\mathscr{V})$. Hence, to complete the proof it suffices to show that

$$R(\mathbf{K}_r) \subset \hat{\mathscr{T}}_r(\mathscr{V})$$

This condition means that

$$\mathbf{T}_\tau(\mathbf{K}_r\mathbf{A}) = \varepsilon_\tau \mathbf{K}_r\mathbf{A} \tag{37.14}$$

for all $\mathbf{A} \in \mathscr{T}_r(\mathscr{V})$. From (37.1), $\mathbf{T}_\tau(\mathbf{K}_r\mathbf{A})$ is given by

$$\mathbf{T}_\tau(\mathbf{K}_r\mathbf{A}) = \frac{1}{r!}\sum_\sigma \varepsilon_\sigma \mathbf{T}_\tau\left(\mathbf{T}_\sigma\mathbf{A}\right)$$

From the resu1t of Exercise 36.2, we can rewrite the preceding equation as

$$\mathbf{T}_\tau(\mathbf{K}_r\mathbf{A}) = \frac{1}{r!}\sum_\sigma \varepsilon_\sigma \mathbf{T}_{\tau\sigma}\mathbf{A} \tag{37.15}$$

Since the set of all permutations form a group, and since

$$\varepsilon_\tau\varepsilon_\sigma = \varepsilon_{\tau\sigma}$$

the right-band side of (37.15) is equal to

$$\frac{1}{r!}\sum_\sigma \varepsilon_\tau\varepsilon_{\tau\sigma}\mathbf{T}_{\tau\sigma}\mathbf{A}$$

or, equivalently,

$$\varepsilon_\tau \frac{1}{r!}\sum_\sigma \varepsilon_{\tau\sigma}\mathbf{T}_{\tau\sigma}\mathbf{A}$$

which is simply another way of writing $\varepsilon_\tau\mathbf{K}_r\mathbf{A}$, so (37.14) is proved.

By exactly the same argument leading to (37.14) we can prove also that

$$\mathbf{K}_r(\mathbf{T}_\tau\mathbf{A}) = \varepsilon_\tau\mathbf{K}_r\mathbf{A} \tag{37.16}$$

Hence $\mathbf{K}_r$ and $\mathbf{T}_\tau$ commute for all permutations τ. Also, (37.9) and (37.10) now imply that $\mathbf{K}_r^2 = \mathbf{K}_r$, for all r. Hence we have shown that $\mathbf{K}_r$ is a projection from $\mathscr{T}_r(\mathscr{V})$ to $\hat{\mathscr{T}}_r(\mathscr{V})$. From a result for projections in general [cf. equation (17.13)], $\mathscr{T}_r(\mathscr{V})$ can be decomposed into the direct sum

$$\mathscr{T}_r(\mathscr{V}) = \hat{\mathscr{T}}_r(\mathscr{V}) \oplus K(\mathbf{K}_r) \tag{37.17}$$

where $K(\mathbf{K}_r)$ is the: kernel of $\mathbf{K}_r$ and is characterized by the following theorem.

Theorem 37.2. The kernel $K(\mathbf{K}_r)$ is generated by the set of simple tensors $\mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^r$ having at least one pair of equal vectors among the vectors $\mathbf{v}^1, ..., \mathbf{v}^r$.

Proof. : Since $\mathscr{T}_r(\mathscr{V})$ is generated by simple tensors, it suffices to show that the difference

$$\mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^r - \mathbf{K}_r\left(\mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^r\right) \tag{37.18}$$

can be expressed as a linear combination of simple tensors having the prescribed property. From (36.3) and (37.1), the difference (37.18) can be written as

$$\frac{1}{r!}\sum_{\sigma}\left(\mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^r - \varepsilon_\sigma \mathbf{v}^{\sigma(1)} \otimes \cdots \otimes \mathbf{v}^{\sigma(r)}\right) \tag{37.19}$$

We claim that each sum of the form

$$\mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^r - \varepsilon_\sigma \mathbf{v}^{\sigma(1)} \otimes \cdots \otimes \mathbf{v}^{\sigma(r)} \tag{37.20}$$

can be expressed as a sum of simple tensors each having at least two equal vectors among the r vectors forming the tensor product. This fact is more or less obvious since in Section 20 we have mentioned that every permutation σ can be decomposed into a product of permutations each switching only one pair of indices, say

$$\sigma = \sigma_k \sigma_{k-1} \cdots \sigma_2 \sigma_1 \tag{37.21}$$

where the number k is even or odd corresponding to σ being an even or odd permutation, respectively. Using the decomposition (37.21), we can rewrite (37.20) in the form

$$\begin{aligned}
&+\left(\mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^r + \mathbf{v}^{\sigma_1(1)} \otimes \cdots \otimes \mathbf{v}^{\sigma_1(r)}\right)\\
&-\left(\mathbf{v}^{\sigma_1(1)} \otimes \cdots \otimes \mathbf{v}^{\sigma_1(r)} + \mathbf{v}^{\sigma_2\sigma_1(1)} \otimes \cdots \otimes \mathbf{v}^{\sigma_2\sigma_1(r)}\right)\\
&+\cdots\\
&-\cdots\\
&-\varepsilon_\sigma\left(\mathbf{v}^{\sigma_{k-1}\cdots\sigma_1(1)} \otimes \cdots \otimes \mathbf{v}^{\sigma_{k-1}\cdots\sigma_1(r)} + \mathbf{v}^{\sigma(1)} \otimes \cdots \otimes \mathbf{v}^{\sigma(r)}\right)
\end{aligned} \tag{37.22}$$

where all the intermediate terms $\mathbf{v}^{\sigma_j\cdots\sigma_1(1)} \otimes \cdots \otimes \mathbf{v}^{\sigma_j\cdots\sigma_1(r)}$, $j = 1, ..., k-1$ cancel in the sum. Since each of $\sigma_1, ..., \sigma_k$ switches only one pair of indices, a typical term in the sum (37.22) has the form

$$\mathbf{v}^a \otimes \cdots \mathbf{v}^s \cdots \mathbf{v}^t \cdots \otimes \mathbf{v}^b + \mathbf{v}^a \otimes \cdots \mathbf{v}^t \cdots \mathbf{v}^s \cdots \otimes \mathbf{v}^b$$

which can be combined into three simple tensors:

$$\mathbf{v}^a \otimes \cdots (\mathbf{u}) \cdots (\mathbf{u}) \cdots \otimes \mathbf{v}^b$$

where $\mathbf{u} = \mathbf{v}^s, \mathbf{v}^t$, and $\mathbf{v}^s + \mathbf{v}^t$. Consequently, (37.22) is a sum of simple tensors having the property prescribed by the theorem. Of course, from (37.16) all those simple tensors belong to the kernel $K(\mathbf{K}_r)$. The proof is complete.

In view of the decomposition (37.17), the subspace $\hat{\mathscr{T}}_r(\mathscr{V})$ is isomorphic to the factor space $\mathscr{T}_r(\mathscr{V})/K(\mathbf{K}_r)$. In fact, some authors use this structure to define the space $\hat{\mathscr{T}}_r(\mathscr{V})$ abstractly without making $\hat{\mathscr{T}}_r(\mathscr{V})$ a subspace of $\mathscr{T}_r(\mathscr{V})$. The preceding theorem shows that this abstract definition of $\hat{\mathscr{T}}_r(\mathscr{V})$ is equivalent to ours.

The next theorem gives a useful property of the skew-symmetric operator $\mathbf{K}_r$.

Theorem 37.3. If $\mathbf{A} \in \mathscr{T}_p(\mathscr{V})$ and $\mathbf{B} \in \mathscr{T}_q(\mathscr{V})$, then

$$\begin{aligned} \mathbf{K}_{p+q}(\mathbf{A} \otimes \mathbf{B}) &= \mathbf{K}_{p+q}(\mathbf{A} \otimes \mathbf{K}_q\mathbf{B}) \\ &= \mathbf{K}_{p+q}(\mathbf{K}_p\mathbf{A} \otimes \mathbf{B}) \\ &= \mathbf{K}_{p+q}(\mathbf{K}_p\mathbf{A} \otimes \mathbf{K}_q\mathbf{B}) \end{aligned} \tag{37.23}$$

Proof. Let τ be an arbitrary permutation of $\{1,...,q\}$. We define

$$\sigma = \begin{pmatrix} 1 & 2 & \cdot & \cdot & \cdot & p & p+2 & \cdot & \cdot & \cdot & p+q \\ 1 & 2 & \cdot & \cdot & \cdot & p & p+\tau(1) & \cdot & \cdot & \cdot & p+\tau(q) \end{pmatrix}$$

Then $\varepsilon_\sigma = \varepsilon_\tau$ and

$$\mathbf{A} \otimes \mathbf{T}_\tau\mathbf{B} = \mathbf{T}_\sigma(\mathbf{A} \otimes \mathbf{B})$$

Hence from (37.14) we have

$$\mathbf{K}_{p+q}\left(\mathbf{A}\otimes\mathbf{T}_\tau\mathbf{B}\right)=\mathbf{K}_{p+q}\left(\mathbf{T}_\sigma\left(\mathbf{A}\otimes\mathbf{B}\right)\right)=\varepsilon_\sigma\mathbf{K}_{p+q}\left(\mathbf{A}\otimes\mathbf{B}\right)$$
$$=\varepsilon_\tau\mathbf{K}_{p+q}\left(\mathbf{A}\otimes\mathbf{B}\right)$$

or, equivalently,

$$\mathbf{K}_{p+q}\left(\mathbf{A}\otimes\varepsilon_\tau\mathbf{T}_\tau\mathbf{B}\right)=\mathbf{K}_{p+q}\left(\mathbf{A}\otimes\mathbf{B}\right) \tag{37.24}$$

Summing (37.24) over all τ, we obtain $(37.23)_1$, A similar argument implies (37.23).

In closing this section, we state without proof an expression for the skew-symmetric operator in terms of the components of its tensor argument. The formula is

$$\mathbf{K}_r\mathbf{A}=\frac{1}{r!}\delta^{i_1\ldots i_r}_{j_1\ldots j_r}A_{i_1\ldots i_r}\mathbf{e}^{j_1}\otimes\cdots\otimes\mathbf{e}^{j_r} \tag{37.25}$$

We leave the proof of this formula as an exercise to the reader. A classical notation for the components of $\mathbf{K}_r\mathbf{A}$ is $A_{[j_1\ldots j_r]}$, so from (37.25)

$$A_{[j_1\ldots j_r]}=\frac{1}{r!}\delta^{i_1\ldots i_r}_{j_1\ldots j_r}A_{i_1\ldots i_r} \tag{37.26}$$

Naturally, we call $\mathbf{K}_r\mathbf{A}$ the *skew-symmetric part* of $\mathbf{A}$. The formula (37.25) and the quotient theorem mentioned in Exercise 33.7 imply that the component formula (33.32) defines a tensor $\mathbf{K}_r$ of order $2r$, a fact proved in Section 33 by means of the transformation law.

Exercises

37.1 Let $\mathbf{A}\in\mathscr{T}_r(\mathscr{V})$ and define an endomorphism $\mathbf{S}_r$ of $\mathscr{T}_r(\mathscr{V})$ by

$$\mathbf{S}_r\mathbf{A}\equiv\frac{1}{r!}\sum_\sigma\mathbf{T}_\sigma\mathbf{A}$$

where the summation is taken over all permutations σ of $\{1,\ldots,r\}$ as in (37.1). Naturally $\mathbf{S}_r$ is called the *symmetric operator*. Show that it is a projection from $\mathscr{T}_r(\mathscr{V})$ into the subspace consisting of completely symmetric tensors of order r. What is the kerne1 $K(\mathbf{S}_r)$? A classical notation for the components of $\mathbf{S}_r\mathbf{A}$ is $A_{(j_1\ldots j_r)}$.

37.2 Prove the formula (37.25).

Section 38. The Wedge Product

In Section 33 we defined the concept of the tensor product $\otimes$, first for vectors, then generalized to tensors. We pointed out that the tensor product has the important universal factorization property. In this section we shall define a similar operation, called the *wedge product* (or the *exterior product*), which we shall denote by the symbol $\wedge$. We shall define first the wedge product of any set of vectors.

If $\left(\mathbf{v}^1,...,\mathbf{v}^r\right)$ is any r-triple of vectors, then their tensor product $\mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^r$ is a simple tensor in $\mathscr{T}_r(\mathscr{V})$ [cf. equation (33.10)]. In the preceding section, we have introduced the skew-symmetric operator $\mathbf{K}_r$, which is a projection from $\mathscr{T}_r(\mathscr{V})$ onto $\hat{\mathscr{T}}_r(\mathscr{V})$. We now define

$$\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^r \equiv \wedge\left(\mathbf{v}^1,...,\mathbf{v}^r\right) \equiv r!\mathbf{K}_r\left(\mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^r\right) \tag{38.1}$$

for any vectors $\mathbf{v}^1,...,\mathbf{v}^r$. For example, if $r = 2$, from (37.5) for the special case that $\mathbf{A} = \mathbf{v}^1 \otimes \mathbf{v}^2$ we have

$$\mathbf{v}^1 \wedge \mathbf{v}^2 = \mathbf{v}^1 \otimes \mathbf{v}^2 - \mathbf{v}^2 \otimes \mathbf{v}^1 \tag{38.2}$$

In general, from (37.1) and (36.3) we have

$$\begin{aligned} \mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^r &= \sum_\sigma \varepsilon_\sigma \mathbf{T}_\sigma\left(\mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^r\right) \\ &= \sum_\sigma \varepsilon_\sigma\left(\mathbf{v}^{\sigma^{-1}(1)} \otimes \cdots \otimes \mathbf{v}^{\sigma^{-1}(r)}\right) \\ &= \sum_\sigma \varepsilon_\sigma\left(\mathbf{v}^{\sigma(1)} \otimes \cdots \otimes \mathbf{v}^{\sigma(r)}\right) \end{aligned} \tag{38.3}$$

where in deriving (38.3), we have used the fact that

$$\varepsilon_\sigma = \varepsilon_{\sigma^{-1}} \tag{38.4}$$

and the fact that the summation is taken over all permutations σ of $\{1,...,r\}$.

We can regard $(38.1)_2$ as the definition of the operation

$$\wedge : \underbrace{\mathscr{V} \times \cdots \times \mathscr{V}}_{r \text{ times}} \to \hat{\mathscr{T}}_p(\mathscr{V}) \tag{38.5}$$

which is called the operation of *wedge product*. From $(38.1)_3$ it is clear that this operation, like the tensor product, is multilinear; further, $\wedge$ is a (completely) skew-symmetric operation in the sense that

$$\wedge(\mathbf{v}^{\sigma(1)}, ..., \mathbf{v}^{\sigma(r)}) = \varepsilon_\sigma \wedge (\mathbf{v}^1, ..., \mathbf{v}^r) \tag{38.6}$$

for all $\mathbf{v}^1, ..., \mathbf{v}^r \in \mathscr{V}$ and all permutations σ of $\{1, ..., r\}$. This fact follows directly from the skew symmetry of $\mathbf{K}_r$ [cf. (37.16)]. Next we show that the wedge product has also a universal factorization property which is the condition asserted by the following.

Theorem 38.1. If $\mathbf{W}$ is an arbitrary completely skew-symmetric multilinear transformation

$$\mathbf{W} : \underbrace{\mathscr{V} \times \cdots \times \mathscr{V}}_{r \text{ times}} \to \mathscr{U} \tag{38.7}$$

where $\mathscr{U}$ is an arbitrary vector space, then there exists a unique linear transformation

$$\mathbf{D} : \hat{\mathscr{T}}_r(\mathscr{V}) \to \mathscr{U} \tag{38.8}$$

such that

$$\mathbf{W}\left(\mathbf{v}^1, ..., \mathbf{v}^r\right) = \mathbf{D}\left(\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^r\right) \tag{38.9}$$

for all $\mathbf{v}^1, ..., \mathbf{v}^r \in \mathscr{V}$. In operator form, (38.9) means that

$$\mathbf{W} = \mathbf{D} \circ \wedge \tag{38.10}$$

Proof We can use the universal factorization property of the tensor product (cf. Theorem 34.1) to decompose $\mathbf{W}$ by

$$\mathbf{W} = \mathbf{C} \circ \otimes \tag{38.11}$$

where $\mathbf{C}$ is a linear transformation from $\mathscr{T}_r(\mathscr{V})$ to $\mathscr{U}$,

$$\mathbf{C} : \mathscr{T}_r(\mathscr{V}) \to \mathscr{U} \tag{38.12}$$

Now in view of the fact that $\mathbf{W}$ is skew-symmetric, we see that the particular linear transformation $\mathbf{C}$ has the property

$$\mathbf{C}\left(\mathbf{v}^{\sigma(1)} \otimes \cdots \otimes \mathbf{v}^{\sigma(r)}\right) = \varepsilon_\sigma \mathbf{C}\left(\mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^r\right) \tag{38.13}$$

for all simple tensors $\mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^r \in \mathscr{T}_r(\mathscr{V})$ and all permutations σ of $\{1,...,r\}$. Consequently, if we multiply (38.13) by ε_σ and sum the result over all σ, then from the linearity of $\mathbf{C}$ and the definition (38.3) we have

$$\begin{aligned} \mathbf{C}\left(\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^r\right) &= r!\mathbf{C}\left(\mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^r\right) \\ &= r!\mathbf{W}\left(\mathbf{v}^1,...,\mathbf{v}^r\right) \end{aligned} \tag{38.14}$$

for all $\mathbf{v}^1,...,\mathbf{v}^r \in \mathscr{V}$. Thus the desired linear transformation $\mathbf{D}$ is simply given by

$$\mathbf{D} = \frac{1}{r!}\mathbf{C}_{\hat{\mathscr{T}}_r(\mathscr{V})} \tag{38.15}$$

where the symbol on the right-hand denotes the restriction of $\mathbf{C}$ on $\hat{\mathscr{T}}_r(\mathscr{V})$ as usual.

Uniqueness of $\mathbf{D}$ can be proved in exactly the same way as in the proof of Theorem 34.1. Here we need the fact that the tensors of the form $\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^r \in \hat{\mathscr{T}}_r(\mathscr{V})$, which may be called simply *skew-symmetric tensors* for an obvious reason, generate the space $\hat{\mathscr{T}}_r(\mathscr{V})$. This fact is a direct consequence of the following results:

(i) $\mathscr{T}_r(\mathscr{V})$ is generated by the set of all simple tensors $\mathbf{v}^1 \otimes \cdots \otimes \mathbf{v}^r$.

(ii) $\mathbf{K}_r$ is a linear trans formation from $\mathscr{T}_r(\mathscr{V})$ *onto* $\hat{\mathscr{T}}_r(\mathscr{V})$.

(iii) Equation (38.1) which defines the simple skew-symmetric tensors $\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^r$.

Knowing that the simple skew-symmetric tensors $\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^r$ form a generating set of $\hat{\mathscr{T}}_r(\mathscr{V})$, we can conclude immediately that the linear transformation $\mathbf{D}$ is unique, since its values on the generating set are uniquely determined by $\mathbf{W}$ through the basic condition (38.9).

As remarked before, the universal factorization property can be used to define the tensor product abstractly. The preceding theorem shows that the same applies to the wedge product. In fact, by following this abstract approach, one can define the vector space $\hat{\mathscr{T}}_r(\mathscr{V})$ entirely independent of the vector space $\mathscr{T}_r(\mathscr{V})$ and the operation $\wedge$ entirely independent of the operation $\otimes$.

Having defined the wedge product for vectors, we can generalize the operation easily to skew-symmetric tensors. If $\mathbf{A} \in \hat{\mathscr{T}}_p(\mathscr{V})$ and $\mathbf{B} \in \hat{\mathscr{T}}_q(\mathscr{V})$, then we define

$$\mathbf{A} \wedge \mathbf{B} = \binom{r}{p} \mathbf{K}_r(\mathbf{A} \otimes \mathbf{B}) \tag{38.16}$$

where, as before,

$$r = p + q \tag{38.17}$$

and

$$\binom{r}{p} = \frac{r(r-1)\cdots(r-p+1)}{p!} = \frac{r!}{p!q!} \tag{38.18}$$

We can regard (38.16) as the definition of the wedge product from $\hat{\mathscr{T}}_p(\mathscr{V}) \times \hat{\mathscr{T}}_q(\mathscr{V}) \to \hat{\mathscr{T}}_r(\mathscr{V})$,

$$\wedge : \hat{\mathscr{T}}_p(\mathscr{V}) \times \hat{\mathscr{T}}_q(\mathscr{V}) \to \hat{\mathscr{T}}_r(\mathscr{V}) \tag{38.19}$$

Clearly, this operation is bilinear and is characterized by the condition that

$$(\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^p) \wedge (\mathbf{u}^1 \wedge \cdots \wedge \mathbf{u}^q) = \mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^p \wedge \mathbf{u}^1 \wedge \cdots \wedge \mathbf{u}^q \tag{38.20}$$

for all simple skew-symmetric tensors $\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^p \in \hat{\mathscr{T}}_p(\mathscr{V})$ and $\mathbf{u}^1 \wedge \cdots \wedge \mathbf{u}^q \in \hat{\mathscr{T}}_q(\mathscr{V})$. We leave the proof of this simple fact as an exercise to the reader. In component form, the wedge product $\mathbf{A} \wedge \mathbf{B}$ is given by

$$\mathbf{A} \wedge \mathbf{B} = \frac{1}{p!q!} \delta^{i_1 \cdots i_r}_{j_1 \cdots j_r} A_{i_1 \cdots i_p} B_{i_{p+1} \cdots i_r} \mathbf{e}^{j_1} \otimes \cdots \otimes \mathbf{e}^{j_r} \tag{38.21}$$

relative to the product basis of any basis $\{\mathbf{e}^j\}$ for $\mathscr{V}$.

In view of (38.20), we can generalize the wedge product further to an arbitrary member of skew-symmetric tensors. For example, if $\mathbf{A} \in \hat{\mathscr{T}}_a(\mathscr{V})$, $\mathbf{B} \in \hat{\mathscr{T}}_b(\mathscr{V})$ and $\mathbf{C} \in \hat{\mathscr{T}}_c(\mathscr{V})$, then we have

$$(\mathbf{A} \wedge \mathbf{B}) \wedge \mathbf{C} = \mathbf{A} \wedge (\mathbf{B} \wedge \mathbf{C}) = \mathbf{A} \wedge \mathbf{B} \wedge \mathbf{C} \tag{38.22}$$

Moreover, $\mathbf{A} \wedge \mathbf{B} \wedge \mathbf{C}$ is also given by

$$\mathbf{A} \wedge \mathbf{B} \wedge \mathbf{C} = \frac{(a+b+c)!}{a!b!c!} \mathbf{K}_{(a+b+c)} (\mathbf{A} \otimes \mathbf{B} \otimes \mathbf{C}) \tag{38.23}$$

Further, the wedge product is multilinear in its arguments and is characterized by the associative law such as

$$\begin{aligned} &(\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^a) \wedge (\mathbf{u}^1 \wedge \cdots \wedge \mathbf{u}^b) \wedge (\mathbf{w}^1 \wedge \cdots \wedge \mathbf{w}^c) \\ &\qquad = \mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^a \wedge \mathbf{u}^1 \wedge \cdots \wedge \mathbf{u}^b \wedge \mathbf{w}^1 \wedge \cdots \wedge \mathbf{w}^c \end{aligned} \tag{38.24}$$

for all simple tensors involved.

From (38.24), the skew symmetry of the wedge product for vectors [cf. (38.6)] can be generalized to the condition such as

$$\mathbf{B} \wedge \mathbf{A} = (-1)^{pq} \mathbf{A} \wedge \mathbf{B} \tag{38.25}$$

for all $\mathbf{A} \in \hat{\mathscr{T}}_p(\mathscr{V})$ and $\mathbf{B} \in \hat{\mathscr{T}}_q(\mathscr{V})$. In particular, if $\mathbf{A}$ is of odd order, then

$$\mathbf{A} \wedge \mathbf{A} = \mathbf{0} \tag{38.26}$$

since from (38.25) if the order p of $\mathbf{A}$ is odd, then

$$\mathbf{A} \wedge \mathbf{A} = (-1)^{p^2} \mathbf{A} \wedge \mathbf{A} = -\mathbf{A} \wedge \mathbf{A} \tag{38.27}$$

A special case of (38.26) is the elementary result that

$$\mathbf{v} \wedge \mathbf{v} = \mathbf{0} \tag{38.28}$$

which is also obvious from (38.2).

Exercises

38.1 Verify (38.6), (38.20), $(38.22)_1$, and (38.23).

38.2 Show that (38.23) can be rewritten as

$$\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^r = \delta^{1\ldots r}_{i_1 \ldots i_r} \mathbf{v}^{i_1} \otimes \cdots \otimes \mathbf{v}^{i_r} \tag{38.29}$$

where the repeated indices are summed from 1 to r.

38.3 Show that

$$\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^r(\mathbf{u}_1, ..., \mathbf{u}_r) = \det \begin{bmatrix} \mathbf{v}^1 \cdot \mathbf{u}_1 & \cdot & \cdot & \cdot & \mathbf{v}^1 \cdot \mathbf{u}_r \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \mathbf{v}^r \cdot \mathbf{u}_1 & \cdot & \cdot & \cdot & \mathbf{v}^r \cdot \mathbf{u}_r \end{bmatrix} \tag{38.30}$$

for all vectors involved.

38.4 Show that

$$\mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r}(\mathbf{e}_{j_1}, ..., \mathbf{e}_{j_2}) = \delta^{i_1 \ldots i_r}_{j_1 \ldots j_r} \tag{38.31}$$

for any reciprocal bases $\{\mathbf{e}^i\}$ and $\{\mathbf{e}_i\}$.

Section 39. Product Bases and Strict Components

In the preceding section we have remarked that the space $\hat{\mathscr{T}}_r(\mathscr{V})$ is generated by the set of all simple skew-symmetric tensors of the form $\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^r$. This generating set is linearly dependent, of course. In this section we shall determine a linear1y independent generating set and thus a basis for $\hat{\mathscr{T}}_r(\mathscr{V})$ consisting entirely of simple skew-symmetric tensors . Naturally, we call such a basis a basis a *product basis* for $\hat{\mathscr{T}}_r(\mathscr{V})$.

Let $\{\mathbf{e}^i\}$ be a basis for $\mathscr{V}$ as usual. Then the simple tensors $\{\mathbf{e}^{i_1} \otimes \cdots \otimes \mathbf{e}^{i_r}\}$ form a product basis for $\mathscr{T}_r(\mathscr{V})$. Since $\hat{\mathscr{T}}_r(\mathscr{V})$ is a subspace of $\mathscr{T}_r(\mathscr{V})$, every element $\mathbf{A} \in \hat{\mathscr{T}}_r(\mathscr{V})$ has the representation

$$\mathbf{A} = A_{i_1 \ldots i_r} \mathbf{e}^{i_1} \otimes \cdots \otimes \mathbf{e}^{i_r} \tag{39.1}$$

where the repeated indices are summed from 1 to N, the dimension of $\mathscr{T}_r(\mathscr{V})$. Now, since $\mathbf{A} \in \hat{\mathscr{T}}_r(\mathscr{V})$, it is invariant under the skew-symmetric operator $\mathbf{K}_r$, namely

$$\mathbf{A} = \mathbf{K}_r \mathbf{A} = A_{i_1 \ldots i_r} \mathbf{K}_r \left(\mathbf{e}^{i_1} \otimes \cdots \otimes \mathbf{e}^{i_r} \right) \tag{39.2}$$

Then from (38.1) we can rewrite the representation (39.1) as

$$\mathbf{A} = \frac{1}{r!} A_{i_1 \ldots i_r} \mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r} \tag{39.3}$$

Thus we have shown that the set of simple skew-symmetric tensors $\{\mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r}\}$ already forms a generating set for $\hat{\mathscr{T}}_r(\mathscr{V})$.

The generating set $\{\mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r}\}$ is still not linearly independent, however. This fact is easily seen, since the wedge product is skew-symmetric, as shown by (38.6). Indeed, if i_1 and i_2 are equal, then $\mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r}$ must vanish. In general if σ is any permutation of $\{1,...,r\}$, then $\mathbf{e}^{i_{\sigma(1)}} \wedge \cdots \wedge \mathbf{e}^{i_{\sigma(r)}}$ is linearly related to $\mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r}$ by

$$\mathbf{e}^{i_{\sigma(1)}} \wedge \cdots \wedge \mathbf{e}^{i_{\sigma(r)}} = \varepsilon_\sigma \mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r} \tag{39.4}$$

Hence if we eliminate the redundant elements of the generating set $\{\mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r}\}$ by restricting the range of the indices $(i_1,...,i_r)$ in such a way that

$$i_1 < \cdots < i_r \tag{39.5}$$

the resulting subset $\left\{\mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r}, i_1 < \cdots < i_r\right\}$ remains a generating set of $\hat{\mathscr{T}}_r(\mathscr{V})$. The next theorem shows that this subset is linearly independent and thus a basis for $\hat{\mathscr{T}}_r(\mathscr{V})$, called the *product basis*.

Theorem 39.1. The set $\left\{\mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r}, i_1 < \cdots < i_r\right\}$ is linearly independent.

Proof. Suppose that the set $\left\{\mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r}, i_1 < \cdots < i_r\right\}$ obeys the homogeneous linear equation

$$\sum_{i_1 < \cdots < i_r} C_{i_1 \ldots i_r} \mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r} \tag{39.6}$$

where $\left\{C_{i_1 \ldots i_r}, i_1 < \cdots < i_r\right\}$ are scalars, and where the summation is taken over all indices $i_1, ..., i_r$ from 1 to N subject to the condition (39.5). Then we must show that $C_{i_1 \ldots i_r}$ vanishes completely. From (38.31), if we evaluate the tensor (39.6) at the argument $\left(\mathbf{e}_{i_1}, ..., \mathbf{e}_{i_r}\right)$, where $\left\{\mathbf{e}_j\right\}$ denotes the reciprocal basis of $\left\{\mathbf{e}^i\right\}$ as usual, we get

$$\sum_{i_1 < \cdots < i_r} C_{i_1 \ldots i_r} \delta^{i_1 \ldots i_r}_{j_1 \ldots j_r} = 0 \tag{39.7}$$

for all $j_1, ..., j_r$ ranging from 1 to N. In particular, if we choose $j_1 < \cdots < j_r$, then the summation reduces to only one term, namely $C_{j_1 \ldots j_r}$ and the equation yields

$$C_{j_1 \ldots j_r} = 0$$

which is the desired result.

It is easy to see that there are only $\begin{pmatrix} N \\ r \end{pmatrix}$ number of elements in the product basis

$$\left\{\mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r}, i_1 < \cdots < i_r\right\} \tag{39.8}$$

Thus we have

$$\dim \hat{\mathscr{T}}_r(\mathscr{V}) = \begin{pmatrix} N \\ r \end{pmatrix} = \frac{N!}{r!(N-r)!} \tag{39.9}$$

In particular, we recover the coro11ary of Theorem 36.1, that is

$$\dim \hat{\mathscr{T}}_r(\mathscr{V}) = 0$$

if $r > N$.

Returning now to the representation (39.3) for an arbitrary skew-symmetric tensor $\mathbf{A} \in \hat{\mathscr{T}}_r(\mathscr{V})$, we can rewrite that representation as

$$\mathbf{A} = \sum_{i_1 < \cdots < i_r} A_{i_1 \ldots i_r} \mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r} \tag{39.10}$$

The reason that we can replace the full summation in (39.3) by the restricted summation is because for each increasing *r*-tuple $(i_1,...,i_r)$ there are precisely $r!$ permutations of $\{i_1,...,i_r\}$. Further, their corresponding $r!$ terms in the full summation (39.3) are all equal to one another, since both $A_{i_1 \ldots i_r}$ and $\mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r}$ are completely skew-symmetric in the indices $(i_1,...,i_r)$, so for any permutation σ of $\{i_1,...,i_r\}$ we have

$$A_{i_{\sigma(1)} \ldots i_{\sigma(r)}} \mathbf{e}^{i_{\sigma(1)}} \wedge \cdots \wedge \mathbf{e}^{i_{\sigma(r)}} = \varepsilon_\sigma^2 A_{i_1 \ldots i_r} \mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r} = A_{i_1 \ldots i_r} \mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r}$$

In view of (39.10), we see that the scalars

$$\left\{ A_{i_1 \ldots i_r}, i_1 < \cdots < i_r \right\}$$

are the components of $\mathbf{A}$ relative to the product basis (39.8). For definiteness, we call these scalars the *strict components* of $\mathbf{A}$.

As an illustration of this concept, let us compute the strict components of the simple skew-symmetric tensor $\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^r \in \hat{\mathscr{T}}_r(\mathscr{V})$. As usual we represent the vector $\mathbf{v}^i$ in component form relative to $\{\mathbf{e}^j\}$ by

$$\mathbf{v}^i = v^i_j \mathbf{e}^j$$

for all $i = 1,...,r$. Using the skew-symmetry of the wedge product, we have

$$\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^r = v^1_{j_1} \cdots v^r_{j_r} \mathbf{e}^{j_1} \wedge \cdots \wedge \mathbf{e}^{j_r} = \sum_{i_1 < \cdots < i_r} v^1_{j_1} \cdots v^r_{j_r} \delta^{j_1 \ldots j_r}_{i_1 \ldots i_r} \mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r} \tag{39.11}$$

which means that the strict components of $\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^r$ are

$$\left\{ v^1_{j_1} \cdots v^r_{j_r} \delta^{j_1 \ldots j_r}_{i_1 \ldots i_r}, i_1 < \cdots < i_r \right\}$$

Next, we consider the transformation rule for the strict components in general.

Theorem 39.2. Under a change of basis from $\{\mathbf{e}^j\}$ to $\{\hat{\mathbf{e}}^j\}$, the strict components of a tensor $\mathbf{A} \in \hat{\mathscr{T}}_r(\mathscr{V})$ obey the transformation rule

$$\hat{A}_{j_1 \cdots j_r} = \sum_{i_1 < \cdots < i_r} T^{i_1 \cdots i_r}_{j_1 \cdots j_r} A_{i_1 \cdots i_r} \tag{39.12}$$

where $T^{i_1 \cdots i_r}_{j_1 \cdots j_r}$ is an $r \times r$ minor of the transformation matrix $\left[T^i_j\right]$ as defined by (21.21). Of course, T^i_j is given by

$$T^i_j = \mathbf{e}^i \cdot \mathbf{e}_j$$

as usual [cf. (31.21)], where $\{\hat{\mathbf{e}}_j\}$ is the reciprocal basis of $\{\hat{\mathbf{e}}^j\}$.

Proof. Since the strict components are nothing but the ordinary tensor components restricted to the subset of indices in increasing order [cf. (39.5)], we can compute their values in the usual way by

$$\hat{A}_{j_1 \cdots j_r} = \mathbf{A}(\hat{\mathbf{e}}_{j_1}, ..., \hat{\mathbf{e}}_{j_r}) \tag{39.13}$$

Substituting the strict component representation (39.10) into (39.13) and making use of the formula (38.30) we obtain

$$\begin{aligned} \hat{A}_{j_1 \cdots j_r} &= \sum_{i_1 < \cdots < i_r} A_{i_1 \cdots i_r} \mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r} \left(\hat{\mathbf{e}}_{j_1}, ..., \hat{\mathbf{e}}_{j_r}\right) \\ &= \sum_{i_1 < \cdots < i_r} A_{i_1 \cdots i_r} \det \begin{bmatrix} \mathbf{e}^{i_1} \cdot \hat{\mathbf{e}}_{j_1} & \cdot & \cdot & \cdot & \mathbf{e}^{i_1} \cdot \hat{\mathbf{e}}_{j_r} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \mathbf{e}^{i_r} \cdot \hat{\mathbf{e}}_{j_1} & \cdot & \cdot & \cdot & \mathbf{e}^{i_r} \cdot \hat{\mathbf{e}}_{j_r} \end{bmatrix} \end{aligned}$$

which is the desired result.

From (21.25), we can write the transformation rule (39.12) in the form

$$\hat{A}_{j_1 \cdots j_r} = \det\left[T^a_b\right] \sum_{i_1 < \cdots < i_r} \left(\operatorname{cof} \hat{T}^{j_1 \cdots j_r}_{i_1 \cdots i_r}\right) A_{i_1 \cdots i_r} \tag{39.14}$$

As al1 illustration of (39.12), take $r = N$. Then (39.9) implies

$$\dim \hat{\mathscr{T}}_N(\mathscr{V}) = 1 \tag{39.15}$$

And (39.12) reduces to

$$\hat{A}_{12\ldots N} = \det\left[T^i_j\right] A_{12\ldots N} \tag{39.16}$$

Comparing (39.16) with (33.15), we see that the strict component of an N-vector transforms according to the rule of an axial scalar of weight 1 [cf. (33.35)]. N-vectors are also called *densities* or *density tensors*.

Next take $r = N-1$. Then (39.9) implies that

$$\dim \hat{\mathscr{T}}_{N-1}(\mathscr{V}) = N \tag{39.17}$$

In this case the transformation rule (39.12) can be rewritten as

$$\hat{A}^k = \det\left[T^i_j\right] \hat{T}^k_l A^l \tag{39.18}$$

where the quantities $\hat{A}^k$ and A^l are defined by

$$\hat{A}^k \equiv (-1)^{N-k} A_{12\ldots\check{k}\ldots N} \tag{39.19}$$

and similarly

$$A^l \equiv (-1)^{N-l} A_{12\ldots\check{l}\ldots N} \tag{39.20}$$

Here the symbol $\check{}$ over k or l means k or l are deleted from the list of indices as before. To prove (39.18), we make use of the alternative form (39.14), obtaining

$$\hat{A}_{12\ldots\check{k}\ldots N} = \det\left[T^a_b\right] \sum_l \left(\operatorname{cof} \hat{T}^{12\ldots\check{k}\ldots N}_{12\ldots\check{l}\ldots N}\right) A_{12\ldots\check{l}\ldots N}$$

Multiplying this equation by $(-1)^{N-k}$ and using the definitions (39.19) and (39.20), we get

$$\hat{A}^k = \det\left[T^a_b\right] \sum_l (-1)^{k+l} \left(\operatorname{cof} \hat{T}^{12\ldots\check{k}\ldots N}_{12\ldots\check{l}\ldots N}\right) A^l$$

But now from (21.23), it is easy to see that the cofactor of the $(N-1)(N-1)$ minor $\hat{T}^{12\ldots\check{k}\ldots N}_{12\ldots\check{l}\ldots N}$ in the matrix $\left[\hat{T}^a_b\right]$ is simply $(-1)^{k+l}$ times the element $\hat{T}^k_l$. Thus (39.18) is proved.

From (39.19) or (39.20) we see that the quantities $\hat{A}^k$ and A^l, like the strict components $\hat{A}_{12\ldots\bar{k}\ldots N}$ and $A_{12\ldots\bar{l}\ldots N}$ characterize the tensor $\mathbf{A}$ completely. In view of the transformation rule (39.18), we see that the quantity A^l transforms according to the rule of the component of an axial vector of weight 1 [cf. (33.35)]. $(N-1)$ vectors are often called (*axial*) *vector densities*.

The operation given by (39.19) and (39.20) can be generalized to $\hat{\mathscr{T}}_{N-r}(\mathscr{V})$ in general. If $A_{i_1\ldots i_{N-r}}$ is a strict component of $\mathbf{A}\in\hat{\mathscr{T}}_{N-r}(\mathscr{V})$, then we define a quantity $A^{j_1\cdots j_r}$ by

$$\begin{aligned} A^{j_1\cdots j_r} &= \sum_{i_1<\cdots<i_{N-r}} \varepsilon^{i_1\ldots i_{N-r}j_1\cdots j_r} A_{i_1\ldots i_{N-r}} \\ &= \frac{1}{(N-r)!}\varepsilon^{i_1\ldots i_{N-r}j_1\cdots j_r} A_{i_1\ldots i_{N-r}} \end{aligned} \tag{39.21}$$

When $r=1$, (39.21) reduces to (39.20). Recall that $\varepsilon^{i_1\cdots i_N}$ transforms according to the rule of an axial contravariant tensor of order *N* and weight 1. Moreover, since $A^{j_1\cdots j_r}$ is skew-symmetric in $(j_1,\ldots,j_r)$, we can write the transformation rule as

$$\hat{A}^{j_1\cdots j_r} = \det\left[T_b^a\right]\sum_{i_1<\cdots<i_r}\hat{T}^{j_1\cdots j_r}_{i_1\ldots i_r}A^{i_1\ldots i_r} \tag{39.22}$$

where $\hat{T}^{j_1\cdots j_r}_{i_1\ldots i_r}$ is an $r\times r$ minor of the transformation matrix $\left[T_b^a\right]$. When $r=1$, (39.22) reduces to (39.18).

As an illustration of (39.20), we take $N=3$; then (39.20) yields

$$A^1 = A_{23}, \qquad A^2 = -A_{13}, \qquad A^3 = A_{12} \tag{39.23}$$

As we shall see, this operation is closely related to the classical operation of *cross product* (or vector product) which assigns an axial vector to a skew- symmetric two-vector on a three-dimensional space.

Before closing this section, we note here a convenient condition for determining whether or not a given set of vectors is linearly dependent by examining their wedge product.

Theorem 39.3. A set of r vectors $\left\{\mathbf{v}^1,\ldots,\mathbf{v}^r\right\}$ is linearly dependent if and only if their wedge product vanishes, i.e.,

$$\mathbf{v}^1\wedge\cdots\wedge\mathbf{v}^r = \mathbf{0} \tag{39.24}$$

Proof. Necessity is obvious, since if $\left\{\mathbf{v}^1,\ldots,\mathbf{v}^r\right\}$ is linearly dependent, then at least one of the vectors can be expressed as a linear combination of the other vectors, say

$$\mathbf{v}^r = \alpha_1 \mathbf{v}^1 + \cdots + \alpha_{r-1} \mathbf{v}^{r-1} \tag{39.25}$$

Then (39.24) follows because

$$\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^r = \sum_{j=1}^{r-1} \alpha_j \mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^{r-1} \wedge \mathbf{v}^j = \mathbf{0}$$

as required by the skew symmetry of the wedge product. Converse1y, if $\{\mathbf{v}^1, ..., \mathbf{v}^r\}$ is linearly independent, then it can be extended to a basis $\{\mathbf{v}^1, ..., \mathbf{v}^N\}$ (cf. Theorem 9.8). From Theorem 39.1, the simple skew-symmetric tensor $\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^N$ forms a basis for the one-dimensional space $\hat{\mathscr{T}}_N(\mathscr{V})$ and thus is nonzero, so that this factor $\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^r$ is also nonzero.

Corollary. A set of N covectors $\{\mathbf{v}^1, ..., \mathbf{v}^N\}$ is a basis for $\mathscr{V}$ if and only if $\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^N \neq \mathbf{0}$.

Exercises

39.1 Show that the product basis of $\hat{\mathscr{T}}_r(\mathscr{V})$ satisfies the following transformation rule:

$$\mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r} = \sum_{j_1 < \cdots < j_r} T^{i_1 \ldots i_r}_{j_1 \ldots j_r} \hat{\mathbf{e}}^{j_1} \wedge \cdots \wedge \hat{\mathbf{e}}^{j_r} \tag{39.26}$$

re1ative to any change of basis from $\{\hat{\mathbf{e}}^i\}$ to $\{\mathbf{e}^j\}$ with transformation matrix $[T^i_j]$.

39.2 For the skew-symmetric tensor space $\hat{\mathscr{T}}_r(\mathscr{V})$ it is customary to define the inner produce

$$* : \hat{\mathscr{T}}_r(\mathscr{V}) \times \hat{\mathscr{T}}_r(\mathscr{V}) \to \mathscr{R} \tag{39.27}$$

by requiring the product basis of an orthonormal basis be orthonormal with respect to $*$. In other words, relative to an orthonormal basis $\{\mathbf{i}^1, ..., \mathbf{i}^N\}$ the inner produce $\mathbf{A} * \mathbf{B}$ of any $\mathbf{A}, \mathbf{B} \in \hat{\mathscr{T}}_r(\mathscr{V})$ is given by

$$\mathbf{A} * \mathbf{B} = \sum_{i_1 < \cdots < i_r} A_{i_1 \ldots i_r} B_{i_1 \ldots i_r} \tag{39.28}$$

Show that this inner product is related to the ordinary inner product $\cdot$ on $\hat{\mathscr{T}}_r(\mathscr{V})$, regarded as a subspace of $\mathscr{T}_r(\mathscr{V})$, by

$$\mathbf{A} * \mathbf{B} = \frac{1}{r!}\mathbf{A} \cdot \mathbf{B} \tag{39.29}$$

for all $\mathbf{A}, \mathbf{B} \in \hat{\mathscr{T}}_r(\mathscr{V})$.

39.3 Relative to the inner product $*$, show that

$$\left(\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^r\right) * \left(\mathbf{u}^1 \wedge \cdots \wedge \mathbf{u}^r\right) = \det \begin{bmatrix} \mathbf{v}^1 \cdot \mathbf{u}^1 & \cdot & \cdot & \cdot & \mathbf{v}^1 \cdot \mathbf{u}^r \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \mathbf{v}^r \cdot \mathbf{u}^1 & \cdot & \cdot & \cdot & \mathbf{v}^r \cdot \mathbf{u}^r \end{bmatrix} \tag{39.30}$$

for all simple skew- symmetric tensors $\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^r$ and $\mathbf{u}^1 \wedge \cdots \wedge \mathbf{u}^r$ in $\hat{\mathscr{T}}_r(\mathscr{V})$.

39.4 Relative to the product basis of any basis $\left\{\mathbf{e}^i\right\}$, show that the inner product $\mathbf{A} * \mathbf{B}$ has the representation

$$\mathbf{A} * \mathbf{B} = \sum_{\substack{i_1 < \cdots < i_r \\ j_1 < \cdots < j_r}} \det \begin{bmatrix} e^{i_1 j_1} & \cdot & \cdot & \cdot & e^{i_1 j_r} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ e^{i_r j_1} & \cdot & \cdot & \cdot & e^{i_r j_r} \end{bmatrix} A_{i_1 \ldots i_r} B_{j_1 \ldots j_r} \tag{39.31}$$

where e^{ij} is given by

$$e^{ij} = \mathbf{e}^i \cdot \mathbf{e}^j \tag{39.32}$$

as before [cf. (14.8)]. In particular, if $\left\{\mathbf{e}^i\right\}$ is orthonormal, then $e^{ij} = \delta^{ij}$ and (39.31) reduces to (39.28).

39.5 Prove the transformation rule (39.22).

Section 40. Determinant and Orientation

In the preceding section we have shown that the strict component of an *N*-vector over an *N*-dimensional space $\mathscr{V}$ transforms according to the rule of an axial scalar of weight 1. For brevity, we call an *N*-vector simply a *density* or a *density tensor*. From (39.15), the space $\hat{\mathscr{T}}_N(\mathscr{V})$ of densities on $\mathscr{V}$ is one-dimensional, and for any basis $\{\mathbf{e}^i\}$ a product basis for $\hat{\mathscr{T}}_N(\mathscr{V})$ is $\mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N$. If $\mathbf{D}$ is a density, then it has the representation

$$\mathbf{D} = D_{12\ldots N}\mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N \tag{40.1}$$

where the strict component $D_{12\ldots N}$ is given by

$$D_{12\ldots N} = \mathbf{D}(\mathbf{e}_1, \ldots, \mathbf{e}_N) \tag{40.2}$$

$\{\mathbf{e}_i\}$ being the reciprocal basis of $\{\mathbf{e}^j\}$ as usual. The representation (40.1) shows that every density is a simple skew-symmetric tensor, e.g., we can represent $\mathbf{D}$ by

$$\mathbf{D} = \left(D_{12\ldots N}\mathbf{e}^1\right) \wedge \cdots \wedge \mathbf{e}^N \tag{40.3}$$

This representation is not unique, of course.

Now if $\mathbf{A}$ is an endomorphism of $\mathscr{V}$, we can define a linear map

$$f : \hat{\mathscr{T}}_N(\mathscr{V}) \to \hat{\mathscr{T}}_N(\mathscr{V}) \tag{40.4}$$

by the condition

$$f\left(\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^N\right) = \left(\mathbf{A}\mathbf{v}^1\right) \wedge \cdots \wedge \left(\mathbf{A}\mathbf{v}^N\right) \tag{40.5}$$

for all $\mathbf{v}^1, \ldots, \mathbf{v}^N \in \mathscr{V}$. We can prove the existence and uniqueness of the linear map f by the universal factorization property of $\hat{\mathscr{T}}_N(\mathscr{V})$ as shown by Theorem 38.1. Indeed, we define first the skew-symmetric multilinear map

$$F : \mathscr{V} \times \cdots \times \mathscr{V} \to \hat{\mathscr{T}}_N(\mathscr{V}) \tag{40.6}$$

by

$$F\left(\mathbf{v}^1, \ldots, \mathbf{v}^N\right) = \left(\mathbf{A}\mathbf{v}^1\right) \wedge \cdots \wedge \left(\mathbf{A}\mathbf{v}^N\right) \tag{40.7}$$

where the skew symmetry and the multilinearity of F are obvious. Then from (38.10) we can define a unique linear map f of the form (40.4) such that

$$F = f \circ \wedge \tag{40.8}$$

which means precisely the condition (40.5).

Since $\hat{\mathscr{T}}_N(\mathscr{V})$ is one-dimensional, the linear map f must have the representation

$$f\left(\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^N\right) = \alpha \mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^N \tag{40.9}$$

where α is a scalar uniquely determined by f and hence $\mathbf{A}$. We claim that

$$\alpha = \det \mathbf{A} \tag{40.10}$$

Thus the determinant of $\mathbf{A}$ can be defined free of any basis by the condition

$$\left(\mathbf{A}\mathbf{v}^1\right) \wedge \cdots \wedge \left(\mathbf{A}\mathbf{v}^N\right) = (\det \mathbf{A})\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^N \tag{40.11}$$

for all $\mathbf{v}^1, ..., \mathbf{v}^N \in \mathscr{V}$.

To prove (40.10), or, equivalently, (40.11), we choose reciprocal basis $\{\mathbf{e}_i\}$ and $\{\mathbf{e}^i\}$ for $\mathscr{V}$ and recall that the determinant of $\mathbf{A}$ is given by the determinant of the component matrix of $\mathbf{A}$. Let $\mathbf{A}$ be represented by the component forms

$$\mathbf{A}\mathbf{e}_i = A^j{}_i \mathbf{e}_j, \qquad \mathbf{A}\mathbf{e}^i = A_j{}^i \mathbf{e}^j \tag{40.12}$$

where $(40.12)_2$ is meaningful because $\mathscr{V}$ is an inner product space. Then, by definition, we have

$$\det \mathbf{A} = \det\left[A^j{}_i\right] = \det\left[A_j{}^i\right] \tag{40.13}$$

Substituting $(40.12)_2$ into (40.5), we get

$$f\left(\mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N\right) = A_{j_1}{}^1 \cdots A_{j_N}{}^N \mathbf{e}^{j_1} \wedge \cdots \wedge \mathbf{e}^{j_N} \tag{40.14}$$

The skew symmetry of the wedge product implies that

$$\mathbf{e}^{j_1} \wedge \cdots \wedge \mathbf{e}^{j_N} = \varepsilon^{j_1 \ldots j_N} \mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N \tag{40.15}$$

where the ε symbol is defined by (20.3). Combining (40.14) and (40.15), and comparing the resu1t with (40.9), we see that

$$\alpha \mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N = A_{j_1}{}^1 \cdots A_{j_N}{}^N \varepsilon^{j_1,,j_N} \mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N \tag{40.16}$$

or equivalently

$$\alpha = A_{j_1}{}^1 \cdots A_{j_N}{}^N \varepsilon^{j_1,,j_N} \tag{40.17}$$

since $\mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N$ is a basis for $\hat{\mathscr{T}}_N(\mathscr{V})$. But by definition the right-hand side of (40.17) is equal to the determinant of the matrix $\left[A_j{}^i \right]$ and thus (40.10) is proved.

As an illustration of (40.11), we see that

$$\det \mathbf{I} = 1 \tag{40.18}$$

since we have

$$\left(\mathbf{I}\mathbf{v}^1\right) \wedge \cdots \wedge \left(\mathbf{I}\mathbf{v}^N\right) = \mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^N = (\det \mathbf{I})\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^N$$

More generally, we have

$$\det(\alpha \mathbf{I}) = \alpha^N \tag{40.19}$$

Since

$$\left(\alpha \mathbf{I}\mathbf{v}^1\right) \wedge \cdots \wedge \left(\alpha \mathbf{I}\mathbf{v}^N\right) = \left(\alpha \mathbf{v}^1\right) \wedge \cdots \wedge \left(\alpha \mathbf{v}^N\right) = \alpha^N \mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^N$$

It is now also obvious that

$$\det(\mathbf{AB}) = (\det \mathbf{A})(\det \mathbf{B}) \tag{40.20}$$

since we have

$$\begin{aligned} \left(\mathbf{A}\mathbf{B}\mathbf{v}^1\right) \wedge \cdots \wedge \left(\mathbf{A}\mathbf{B}\mathbf{v}^N\right) &= (\det \mathbf{A})\left(\mathbf{B}\mathbf{v}^1\right) \wedge \cdots \wedge \left(\mathbf{B}\mathbf{v}^N\right) \\ &= (\det \mathbf{A})(\det \mathbf{B})\mathbf{v}^1 \wedge \cdots \wedge \mathbf{v}^N \end{aligned}$$

for all $\mathbf{v}^1, ..., \mathbf{v}^N \in \mathscr{V}$. Combining (40.19) and (40.20), we recover also the formula

$$\det(\alpha \mathbf{A}) = \alpha^N (\det \mathbf{A}) \tag{40.21}$$

Another application of (40.11) yields the result that $\mathbf{A}$ is non-singular if and only if $\det \mathbf{A}$ is non-zero. To see this result, notice first the simple fact that $\mathbf{A}$ is non-singular if and only if $\mathbf{A}$ transforms any basis $\{\mathbf{e}^j\}$ of $\mathscr{V}$ into a basis $\{\mathbf{A}\mathbf{e}^j\}$ of $\mathscr{V}$. From the corollary of Theorem 39.3, $\{\mathbf{A}\mathbf{e}^j\}$ is a basis of $\mathscr{V}$ if and only if $(\mathbf{A}\mathbf{e}^1)\wedge\cdots\wedge(\mathbf{A}\mathbf{e}^N)\neq\mathbf{0}$. Then from (40.11), $(\mathbf{A}\mathbf{e}^1)\wedge\cdots\wedge(\mathbf{A}\mathbf{e}^N)\neq\mathbf{0}$ if and only if $\det\mathbf{A}\neq 0$. Thus $\det\mathbf{A}\neq 0$ is necessary and sufficient for $\mathbf{A}$ to be non-singular.

The determinant, of course), is just one of the invariants of $\mathbf{A}$. In Chapter 6, Section 26, we have introduced the set of fundamental invariants $\{\mu_1,...,\mu_N\}$ for $\mathbf{A}$ by the equation

$$\det(\mathbf{A}+t\mathbf{I}) = t^N + \mu_1 t^{N-1} + \cdots + \mu_{N-1}t + \mu_N \tag{40.22}$$

Since we have shown that the determinant of any endomorphism can be characterized by the condition (40.11), if we apply (40.11) to the endomorphism $\mathbf{A}+t\mathbf{I}$, the result is

$$(\mathbf{A}\mathbf{v}^1 + t\mathbf{v}^1)\wedge\cdots\wedge(\mathbf{A}\mathbf{v}^N + t\mathbf{v}^N) = (t^N + \mu_1 t^{N-1} + \cdots + \mu_{N-1}t + \mu_N)\mathbf{v}^1\wedge\cdots\wedge\mathbf{v}^N$$

Comparing the coefficient of t^{N-k} on the two sides of this equation, we obtain

$$\mu_K \mathbf{v}^1\wedge\cdots\wedge\mathbf{v}^N = \sum_{i_1<\cdots<i_K} \mathbf{v}^1\wedge\cdots\wedge\mathbf{A}\mathbf{v}^{i_1}\wedge\cdots\wedge\mathbf{A}\mathbf{v}^{i_2}\wedge\cdots\wedge\mathbf{A}\mathbf{v}^{i_K}\wedge\cdots\wedge\mathbf{v}^N \tag{40.23}$$

where there are precisely $\binom{N}{K}$ terms in the summation on the right hand side of (40.23), each containing K factors of $\mathbf{A}$ acting on the covectors $\mathbf{v}^{i_1},...,\mathbf{v}^{i_K}$. In particular, taking $K=N$, we recover the result

$$\mu_N \mathbf{v}^1\wedge\cdots\wedge\mathbf{v}^N = (\mathbf{A}\mathbf{v}^1)\wedge\cdots\wedge(\mathbf{A}\mathbf{v}^N)$$

which is the same as (40.11) since

$$\mu_N = \det\mathbf{A}$$

Likewise, taking $K=1$, we obtain

$$\mu_1 \mathbf{v}^1\wedge\cdots\wedge\mathbf{v}^N = \sum_{i=1}^{N} \mathbf{v}^1\wedge\cdots\wedge\mathbf{A}\mathbf{v}^i\wedge\cdots\wedge\mathbf{v}^N \tag{40.24}$$

which characterizes the trace of $\mathbf{A}$ free of any basis, since

$$\mu_1 = \operatorname{tr} \mathbf{A}$$

Of course, we can prove the existence arid the uniqueness of μ_K satisfying the condition (40.23) by the universal factorization property as before.

Since the space $\hat{\mathscr{T}}_N(\mathscr{V})$ is one-dimensional, it is divided by $\mathbf{0}$ into two nonzero segments. Two nonzero densities $\mathbf{D}_1$ and $\mathbf{D}_2$ are in the same segment if they differ by a positive scalar factor, say $\mathbf{D}_2 = \lambda \mathbf{D}_1$, where $\lambda > 0$. Conversely, if $\mathbf{D}_1$ and $\mathbf{D}_2$ differ by a negative scalar factor, then they belong to different segments. If $\{\mathbf{e}^i\}$ and $\{\hat{\mathbf{e}}^i\}$ are two basis for $\mathscr{V}$, we define their corresponding product basis $\mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N$ and $\hat{\mathbf{e}}^1 \wedge \cdots \wedge \hat{\mathbf{e}}^N$ for $\hat{\mathscr{T}}_N(\mathscr{V})$ as usual; then we say that $\{\mathbf{e}^i\}$ and $\{\hat{\mathbf{e}}^i\}$ have the *same orientation* and that the change of basis is a *proper transformation* if $\mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N$ and $\hat{\mathbf{e}}^1 \wedge \cdots \wedge \hat{\mathbf{e}}^N$ belong to the same segment of $\hat{\mathscr{T}}_N(\mathscr{V})$. Conversely, *opposite orientation* and *improper transformation* are defined by the condition that $\mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N$ and $\hat{\mathbf{e}}^1 \wedge \cdots \wedge \hat{\mathbf{e}}^N$ belong to different segments of $\hat{\mathscr{T}}_N(\mathscr{V})$. From (39.26) for the case $r = N$, the product basis $\mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N$ and $\hat{\mathbf{e}}^1 \wedge \cdots \wedge \hat{\mathbf{e}}^N$ are related by

$$\mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N = \det\left[T_j^{\ i}\right] \hat{\mathbf{e}}^1 \wedge \cdots \wedge \hat{\mathbf{e}}^N \tag{40.25}$$

Consequently, the change of basis is proper or improper if and only if $\det\left[T_j^{\ i}\right]$ is positive or negative, respectively.

It is conventional to designate one segment of $\hat{\mathscr{T}}_N(\mathscr{V})$ *positive* and the other one *negative*. Whenever such a designation has been made, we say that $\hat{\mathscr{T}}_N(\mathscr{V})$ and, thus, $\mathscr{V}$, are *oriented*. For an oriented space $\mathscr{V}$ a basis $\{\mathbf{e}^i\}$ is *positively oriented* or *right-handed* if its product basis $\mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N$ belongs to the positive segment of $\hat{\mathscr{T}}_N(\mathscr{V})$; otherwise, the basis is *negatively oriented* or *left-handed*. It is customary to restrict the choice of basis for an oriented space to positively oriented bases only. Under such a restriction, the parity ε [cf, (33.35)] of all relative tensors always has the value $+1$, since the transformation is restricted to be proper. Hence, in this case it is .not necessary to distinguish relative tensors into axial ones and polar ones.

As remarked in Exercise 39.2, it is conventional to use the inner product $*$ defined by (39.29) for skew-symmetric tensors. For the space of densities $\hat{\mathscr{T}}_N(\mathscr{V})$ the inner product is given by [cf. (39.31)]

$$\mathbf{A} * \mathbf{B} = \det\left[e^{ij}\right] A_{12\ldots N} B_{12\ldots N} \tag{40.26}$$

where $A_{12\ldots N}$ and $B_{12\ldots N}$ are the strict component of $\mathbf{A}$ and $\mathbf{B}$ as defined by (40.2). Clearly, there exists an unique unit density with respect to $*$ in each segment of $\hat{\mathscr{T}}_N(\mathscr{V})$. If $\hat{\mathscr{T}}_N(\mathscr{V})$ is oriented, the unit density in the positive segment is usually denoted by $\mathbf{E}$, then the unit density in the negative segment is $-\mathbf{E}$. In general, if $\mathbf{D}$ is any density in $\hat{\mathscr{T}}_N(\mathscr{V})$, then it has the representation

$$\mathbf{D} = d\mathbf{E} \tag{40.27}$$

In accordance with the usual practice, the scalar d in (40.27), which is the component of $\mathbf{D}$ relative to the positive unit density $\mathbf{E}$, is also called a *density* or more specifically a *density scalar*, as opposed to the term *density tensor* for $\mathbf{D}$. This convention is consistent with the common practice of identifying a scalar α with an element $\alpha 1$ in $\mathscr{R}$, where $\mathscr{R}$ is, of course, an oriented one-dimensional space with positive unit element 1.

If $\{\mathbf{e}^i\}$ is a basis in an oriented space $\mathscr{V}$, then we define the *density* e^* of $\{\mathbf{e}^i\}$ to be the component of the product basis $\mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N$ relative to $\mathbf{E}$, namely

$$\mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N = e^*\mathbf{E} \tag{40.28}$$

In other words, e^* is the density of the product basis of $\{\mathbf{e}^i\}$. Clearly, $\{\mathbf{e}^i\}$ is positively oriented or negatively oriented depending on whether its density e^* is positive or negative, respectively. We say that a basis $\{\mathbf{e}^i\}$ is *unimodular* if its density has unit absolute value. i.e., $|e^*| = 1$. All orthonormal bases, right-handed or left-handed, are always unimodular. A unimodular basis in general need not be orthonormal, however.

From (40.28) and (40.26), we have

$$e^{*2} = e^*\mathbf{E} * e^*\mathbf{E} = \det\left[e^{ij}\right] \tag{40.29}$$

for any basis $\{\mathbf{e}^i\}$. Hence the *absolute density* $|e^*|$ of $\{\mathbf{e}^i\}$ is given by

$$|e^*| = \left(\det\left[e^{ij}\right]\right)^{1/2} \tag{40.30}$$

Substituting (40.30) into (40.28), we have

$$\mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N = \varepsilon\left(\det\left[e^{ij}\right]^{1/2}\right)\mathbf{E} \tag{40.31}$$

where ε is $+1$ if $\{\mathbf{e}^i\}$ is positively oriented and it is -1 if $\{\mathbf{e}^i\}$ is negatively oriented.

From (40.25) and (40.27) the density of a basis transforms according to the rule

$$e^* = \det\left[T^i{}_j\right]\hat{e}^* \tag{40.32}$$

under a change of basis from $\{\mathbf{e}^j\}$ to $\{\hat{\mathbf{e}}^j\}$. Comparing (40.32) with (33.35), we see that the density of a basis is an axial relative scalar of weight -1.

An interesting property of a unit density ($\mathbf{E}$ or $-\mathbf{E}$) is given by the following theorem.

Theorem 40.1. If $\mathbf{U}$ is a unit density in $\hat{\mathscr{T}}_N(\mathscr{V})$, then

$$\left[\mathbf{U}*\left(\mathbf{v}^1\wedge\cdots\wedge\mathbf{v}^N\right)\right]\left[\mathbf{U}*\left(\mathbf{u}^1\wedge\cdots\wedge\mathbf{u}^N\right)\right] = \det\left[\mathbf{v}^i\cdot\mathbf{u}^j\right] \tag{40.33}$$

for all $\mathbf{v}^1,...,\mathbf{v}^N$ and $\mathbf{u}^1,...,\mathbf{u}^N$ in $\mathscr{V}$.

Proof. Clearly we can represent $\mathbf{U}$ as the product basis of an orthonormal basis, say $\{\mathbf{i}_k\}$, with

$$\mathbf{U} = \mathbf{i}_1\wedge\cdots\wedge\mathbf{i}_N \tag{40.34}$$

From (40.34) and (39.30), we then have

$$\mathbf{U}*\left(\mathbf{v}^1\wedge\cdots\wedge\mathbf{v}^N\right) = \det\left[\mathbf{i}_j\cdot\mathbf{v}^i\right] = \det\left[v^i{}_j\right]$$

and

$$\mathbf{U}*\left(\mathbf{u}^1\wedge\cdots\wedge\mathbf{u}^N\right) = \det\left[\mathbf{i}_j\cdot\mathbf{u}^i\right] = \det\left[u^i{}_j\right]$$

where $v^i{}_j$ and $u^i{}_j$ are the jth components of $\mathbf{v}^i$ and $\mathbf{u}^i$ relative to $\{\mathbf{i}_k\}$, respectively. Using the product rule and the transpose rule of the determinant, we then obtain

$$\begin{aligned}\left[\mathbf{U}*\left(\mathbf{v}^1\wedge\cdots\wedge\mathbf{v}^N\right)\right]\left[\mathbf{U}*\left(\mathbf{u}^1\wedge\cdots\wedge\mathbf{u}^N\right)\right] &= \det\left[v^i{}_j\right]\det\left[u^k{}_l\right]\\ &= \det\left[v^i{}_j\right]\det\left[u^k{}_l\right]^T = \det\left(\left[v^i{}_j\right]\left[u^k{}_l\right]^T\right)\\ &= \det\left[\sum_{k=1}^N v^i{}_k u^i{}_k\right] = \det\left[\mathbf{v}^i\cdot\mathbf{u}^j\right]\end{aligned}$$

which is the desired result.

By exactly the same argument, we have also the following theorem.

Theorem 40.2. If $\mathbf{U}$ is a unit density as before, then

$$\mathbf{U}(\mathbf{v}_1,...,\mathbf{v}_N)\mathbf{U}(\mathbf{u}_1,...,\mathbf{u}_N) = \det\left[\mathbf{v}_i \cdot \mathbf{u}_j\right] \tag{40.35}$$

for all $\mathbf{v}_1,...,\mathbf{v}_N$ and $\mathbf{u}_1,...,\mathbf{u}_N$ in $\mathscr{V}$.

Exercises

40.1 If $\mathbf{A}$ is an endomorphism of $\mathscr{V}$ show that the determinant of $\mathbf{A}$ can be characterized by the following basis-free condition:

$$\mathbf{D}(\mathbf{A}\mathbf{v}_1,...,\mathbf{A}\mathbf{v}_N) = (\det \mathbf{A})\mathbf{D}(\mathbf{v}_1,...,\mathbf{v}_N) \tag{40.36}$$

for all densities $\mathbf{D} \in \hat{\mathscr{T}}_N(\mathscr{V})$ and all vectors $\mathbf{v}_1,...,\mathbf{v}_N$ in $\mathscr{V}$.

Note. A complete proof of this result consists of the following two parts:

(i) There exists a unique scalar α, depending on $\mathbf{A}$, such that

$$\mathbf{D}(\mathbf{A}\mathbf{v}_1,...,\mathbf{A}\mathbf{v}_N) = \alpha\mathbf{D}(\mathbf{v}_1,...,\mathbf{v}_N)$$

for all $\mathbf{D} \in \hat{\mathscr{T}}_N(\mathscr{V})$ and all vectors $\mathbf{v}_1,...,\mathbf{v}_N$ in $\mathscr{V}$.

(ii) The scalar α is equal to $\det \mathbf{A}$.

40.2 Use the formula (40.36) and show that the fundamental invariants $\{\mu_1,...,\mu_N\}$ of an endomorphism $\mathbf{A}$ can be characterized by

$$\mu_k\mathbf{D}(\mathbf{v}_1,...,\mathbf{v}_N) = \sum_{i_1<\cdots<i_K} \mathbf{D}(\mathbf{v}_1,...,\mathbf{A}\mathbf{v}_{i_1},...,\mathbf{A}\mathbf{v}_{i_K},...,\mathbf{v}_N) \tag{40.37}$$

for all $\mathbf{v}_1,...,\mathbf{v}_N \in \mathscr{V}$ and all $\mathbf{D} \in \hat{\mathscr{T}}_N(\mathscr{V})$. Here the summation on the right-hand side of (40.37) is similar to that of (40.23), i.e., there are $\binom{N}{K}$ terms in the summation, each containing the value of $\mathbf{D}$ at the argument with K vectors $\mathbf{v}_{i_1},...,\mathbf{v}_{i_K}$, acted on by $\mathbf{A}$. In particular, taking $K = N$, we recover the formula (40.36), and taking $K = 1$, we obtain

$$(\operatorname{tr}\mathbf{A})\mathbf{D}(\mathbf{v}_1,...,\mathbf{v}_N) = \sum_{i=1}^{N}\mathbf{D}(\mathbf{v}_1,...,\mathbf{A}\mathbf{v}_i,...,\mathbf{v}_N) \tag{40.38}$$

40.3 The *Gramian* is a function (not multilinear) G of K vectors defined by

$$G(\mathbf{v}_1,...,\mathbf{v}_N) = \det\left[\mathbf{v}_i \cdot \mathbf{v}_j\right] \tag{40.39}$$

where the matrix $\left[\mathbf{v}_i \cdot \mathbf{v}_j\right]$ is $K \times K$, of course, Use the result of Theorem 40.2 and show that

$$G(\mathbf{v}_1, ..., \mathbf{v}_N) \geq 0 \tag{40.40}$$

and that the equality holds if and only if $\{\mathbf{v}_1, ..., \mathbf{v}_K\}$ is a linear dependent set. *Note*. When $K = 2$, the result (40.40) reduces to the Schwarz inequality.

40.4 Use the results of this section and prove that

$$\det \mathbf{A} = \det \mathbf{A}^T$$

for an endomorphism $\mathbf{A} \in \mathscr{L}(\mathscr{V}; \mathscr{V})$.

40.5 Show that $\operatorname{adj} \mathbf{A}$, which is used in (26.15), can be defined by

$$(\operatorname{adj} \mathbf{A})\mathbf{v}^1 \wedge \mathbf{v}^2 \wedge \cdots \wedge \mathbf{v}^N = \mathbf{v}^1 \wedge \mathbf{A}\mathbf{v}^2 \wedge \cdots \wedge \mathbf{A}\mathbf{v}^N$$

40.6 Use (40.23) and the result in Exercise 40.5 and show that

$$\mu_{N-1} = \operatorname{tr} \operatorname{adj} \mathbf{A}$$

40.7 Show that

$$\operatorname{adj} \mathbf{AB} = \operatorname{adj} \mathbf{B} \operatorname{adj} \mathbf{A}$$
$$\det \operatorname{adj} \mathbf{A} = (\det \mathbf{A})^{N-1}$$
$$\operatorname{adj}(\operatorname{adj} \mathbf{A}) = (\det \mathbf{A})^{N-2} \mathbf{A}$$

and

$$\det \operatorname{adj}(\operatorname{adj} \mathbf{A}) = (\det \mathbf{A})^{(N-1)^2}$$

for $\mathbf{A}, \mathbf{B} \in \mathscr{L}(\mathscr{V}; \mathscr{V})$ and $N = \dim \mathscr{V}$.

Section 41. Duality

As a consequence of (39.9), the dimension of the space $\hat{\mathscr{T}}_r(\mathscr{V})$ is equal to that of $\hat{\mathscr{T}}_{N-r}(\mathscr{V})$. Hence the spaces $\hat{\mathscr{T}}_r(\mathscr{V})$ and $\hat{\mathscr{T}}_{N-r}(\mathscr{V})$ are isomorphic. The purpose of this section is to establish a particular isomorphism,

$$\mathbf{D}_r : \hat{\mathscr{T}}_r(\mathscr{V}) \to \hat{\mathscr{T}}_{N-r}(\mathscr{V}) \tag{41.1}$$

called the *duality operator*, for an oriented space $\mathscr{V}$. As we shall see, this duality operator gives rise to a definition to the operation of *cross product* or *vector product* when $N = 3$ and $r = 2$.

We recall first that $\hat{\mathscr{T}}_r(\mathscr{V})$ is equipped with the inner product $*$ given by (39.29). Let $\mathbf{E}$ be the distinguished positive unit density in $\hat{\mathscr{T}}_N(\mathscr{V})$ as introduced in the preceding section. Then for any $\mathbf{A} \in \hat{\mathscr{T}}_r(\mathscr{V})$ we define its *dual* $\mathbf{D}_r\mathbf{A}$ in $\hat{\mathscr{T}}_{N-r}(\mathscr{V})$ by the condition

$$\mathbf{E} * (\mathbf{A} \wedge \mathbf{Z}) = (\mathbf{D}_r\mathbf{A}) * \mathbf{Z} \tag{41.2}$$

for all $\mathbf{Z} \in \hat{\mathscr{T}}_{N-r}(\mathscr{V})$. Since the left-hand side of (41.2) is linear in $\mathbf{Z}$, and since $*$ is an inner product, $\mathbf{D}_r\mathbf{A}$ is uniquely determined by (41.2). Further, from (41.2), $\mathbf{D}_r\mathbf{A}$ depends linearly on $\mathbf{A}$, so $\mathbf{D}_r$, is a linear transformation from $\hat{\mathscr{T}}_r(\mathscr{V})$ to $\hat{\mathscr{T}}_{N-r}(\mathscr{V})$.

Theorem 41.1. The duality operator $\mathbf{D}_r$ defined by the condition (41.2) is an isomorphism.

Proof. Since $\hat{\mathscr{T}}_r(\mathscr{V})$ and $\hat{\mathscr{T}}_{N-r}(\mathscr{V})$ are isomorphic, it suffices to show that $\mathbf{D}_r$ is an isometry, i.e.,

$$(\mathbf{D}_r\mathbf{A}) * (\mathbf{D}_r\mathbf{A}) = \mathbf{A} * \mathbf{A} \tag{41.3}$$

for all $\mathbf{A} \in \hat{\mathscr{T}}_r(\mathscr{V})$. The polar identity

$$\mathbf{A} * \mathbf{B} = \frac{1}{2}\{\mathbf{A} * \mathbf{A} + \mathbf{B} * \mathbf{B} - (\mathbf{A} - \mathbf{B}) * (\mathbf{A} - \mathbf{B})\} \tag{41.4}$$

then implies that $\mathbf{D}_r$ preserves the inner product also; i.e.,

$$(\mathbf{D}_r\mathbf{A}) * (\mathbf{D}_r\mathbf{B}) = \mathbf{A} * \mathbf{B} \tag{41.5}$$

for all $\mathbf{A}$ and $\mathbf{B}$ in $\hat{\mathscr{T}}_r(\mathscr{V})$. To prove (41.3), we choose an arbitrary right-handed orthogonal basis $\{\mathbf{i}_j\}$ of $\mathscr{V}$. Then $\mathbf{E}$ is simply the product basis of $\{\mathbf{i}_j\}$

$$\mathbf{E} = \mathbf{i}_1 \wedge \cdots \wedge \mathbf{i}_N \tag{41.6}$$

Of course, we represent $\mathbf{A}, \mathbf{Z}$, and $\mathbf{D}_r\mathbf{A}$ in strict component form relative to $\{\mathbf{i}_j\}$ also; then (41.2) can be written as

$$\sum_{\substack{i_1<\cdots<i_r \\ j_1<\cdots<j_{N-r}}} \varepsilon_{i_1\ldots i_r j_1\ldots j_{N-r}} A_{i_1\ldots i_r} Z_{j_1\ldots j_{N-r}} = \sum_{j_1<\cdots<j_{N-r}} (\mathbf{D}_r\mathbf{A})_{j_1\ldots j_{N-r}} Z_{j_1\ldots j_{N-r}} \tag{41.7}$$

Since $\mathbf{Z}$ is arbitrary, (41.7) implies

$$(\mathbf{D}_r\mathbf{A})_{j_1\ldots j_{N-r}} = \sum_{i_1<\cdots<i_r} \varepsilon_{i_1\ldots i_r j_1\ldots j_{N-r}} A_{i_1\ldots i_r} \tag{41.8}$$

This formula gives the strict components of $\mathbf{D}_r\mathbf{A}$ in terms of those of $\mathbf{A}$ relative to a right-handed orthonormal basis.

From (41.8) we can compute the value of left-hand side of (41.3) by

$$\begin{aligned}(\mathbf{D}_r\mathbf{A}) * (\mathbf{D}_r\mathbf{A}) &= \sum_{\substack{i_1<\cdots<i_r \\ k_1<\cdots<k_r \\ j_1<\cdots<j_{N-r}}} \varepsilon_{i_1\ldots i_r j_1\ldots j_{N-r}} \varepsilon_{k_1\ldots k_r j_1\ldots j_{N-r}} A_{i_1\ldots i_r} A_{k_1\ldots k_r} \\ &= \sum_{i_1<\cdots<i_r} A_{i_1\ldots i_r} A_{i_1\ldots i_r} = \mathbf{A} * \mathbf{A}\end{aligned}$$

which is the desired result.

Having proved that $\mathbf{D}_r$ is an isomorphism from $\hat{\mathscr{T}}_r(\mathscr{V})$ to $\hat{\mathscr{T}}_{N-r}(\mathscr{V})$ relative to the inner product $*$, we can now use the condition (41.2) and (41.5) to compute the inverse of $\mathbf{D}_r$. Indeed, from the skew symmetry of the wedge product [cf. (38.25)] we have

$$\mathbf{A} \wedge \mathbf{Z} = (-1)^{r(N-r)} \mathbf{Z} \wedge \mathbf{A} \tag{41.9}$$

Hence (41.2) yields

$$(\mathbf{D}_r\mathbf{A}) * \mathbf{Z} = (-1)^{r(N-r)} \mathbf{A} * (\mathbf{D}_{N-r}\mathbf{Z}) \tag{41.10}$$

for all $\mathbf{A} \in \hat{\mathscr{T}}_r(\mathscr{V})$ and $\mathbf{Z} \in \hat{\mathscr{T}}_{N-r}(\mathscr{V})$. On the other hand. (41.5) yields

$$(\mathbf{D}_r\mathbf{A})*(\mathbf{Z}) = \mathbf{A}*(\mathbf{D}_r^{-1}\mathbf{Z}) \tag{41.11}$$

when we chose $\mathbf{D}_r\mathbf{B} = \mathbf{Z}$. Comparing (41.10) with (41.11), we obtain

$$\mathbf{D}_r^{-1} = (-1)^{r(N-r)}\mathbf{D}_{N-r} \tag{41.12}$$

which is the desired result.

We notice that the equations (41.8) and (39.21) are very similar to each other, the only difference being that (39.21) applies to all bases while (41.8) is restricted to right-handed orthonormal bases only. Since the quantities given by (39.21) transform according to the rule of an axial tensor density, while the dual $\mathbf{D}_r\mathbf{A}$ is a tensor in $\hat{\mathscr{T}}_{N-r}(\mathscr{V})$, the formula (41.8) is no longer valid if the basis is not right-handed and orthogonal, If $\{\mathbf{e}^i\}$ is an arbitrary basis, then $\mathbf{E}$ is given by (40.31). In this case if we represent $\mathbf{A}, \mathbf{Z}$, and $\mathbf{D}_r\mathbf{A}$ again by their strict components relative to $\{\mathbf{e}^i\}$, then from (39.21) and (40.28) the condition (41.2) can be written as

$$\sum_{\substack{i_1<\cdots<i_r \\ j_1<\cdots<j_{N-r}}} A_{i_1\ldots i_r} Z_{j_1\ldots j_{N-r}} \varepsilon^{i_1\ldots i_r j_1\ldots j_{N-r}} e^* = \sum_{\substack{k_1<<k_{N-r} \\ j_1<\cdots<j_{N-r}}} \det\begin{bmatrix} e^{k_1 j_1} & \cdot & \cdot & \cdot & e^{k_1 j_{N-r}} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ e^{k_{N-r} j_1} & \cdot & \cdot & \cdot & e^{k_{N-r} j_{N-r}} \end{bmatrix} (\mathbf{D}_r\mathbf{A})_{k_1\ldots k_{N-r}} Z_{j_1\ldots j_{N-r}} \tag{41.13}$$

Since $\mathbf{Z}$ is arbitrary, (41.13) implies

$$\sum_{i_1<\cdots<i_r} A_{i_1\ldots i_r} \varepsilon^{i_1\ldots i_r j_1\ldots j_{N-r}} e^* = \sum_{k_1<<k_{N-r}} \det\begin{bmatrix} e^{k_1 j_1} & \cdot & \cdot & \cdot & e^{k_1 j_{N-r}} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ e^{k_{N-r} j_1} & \cdot & \cdot & \cdot & e^{k_{N-r} j_{N-r}} \end{bmatrix} (\mathbf{D}_r\mathbf{A})_{k_1\ldots k_{N-r}} \tag{41.14}$$

We can solve (41.14) for the components of $\mathbf{D}_r\mathbf{A}$ in the following way. We multiply (41.14) by

$$\det\begin{bmatrix} e_{j_1 l_1} & \cdot & \cdot & \cdot & e_{j_1 l_{N-r}} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ e_{j_{N-r} l_1} & \cdot & \cdot & \cdot & e_{j_{N-r} l_{N-r}} \end{bmatrix}$$

and sum on $j_1,...,j_{N-r}$ in increasing order, then we obtain

$$\left(\mathbf{D}_r\mathbf{A}\right)_{l_1\ldots l_{N-r}} = \sum_{\substack{i_1<\cdots<i_r \\ j_1<\cdots<j_{N-r}}} A_{i_1\ldots i_r} e^* \varepsilon^{i_1\ldots i_r j_1\ldots j_{N-r}} \det\begin{bmatrix} e_{j_1 l_1} & \cdot & \cdot & \cdot & e_{j_1 l_{N-r}} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ e_{j_{N-r} l_1} & \cdot & \cdot & \cdot & e_{j_{N-r} l_{N-r}} \end{bmatrix} \tag{41.15}$$

which is the desired result. In deriving (41.15) we have used the identity (21.26) for the matrix $\left[e^{ij}\right]$ and its inverse $\left[e_{ij}\right]$ in order to obtain the formula

$$\sum_{j_1<\cdots<j_{N-r}} \det\begin{bmatrix} e^{k_1 j_1} & \cdot & \cdot & \cdot & e^{k_1 j_{N-r}} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ e^{k_{N-r} j_1} & \cdot & \cdot & \cdot & e^{k_{N-r} j_{N-r}} \end{bmatrix} \det\begin{bmatrix} e_{j_1 l_1} & \cdot & \cdot & \cdot & e_{j_1 l_{N-r}} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ e_{j_{N-r} l_1} & \cdot & \cdot & \cdot & e_{j_{N-r} l_{N-r}} \end{bmatrix} = \delta^{k_1\ldots k_{N-r}}_{l_1\ldots l_{N-r}} \tag{41.16}$$

Equation (41.15) follows, since we have

$$\sum_{k_1<\cdots<k_{N-r}} \delta^{k_1\ldots k_{N-r}}_{l_1\ldots l_{N-r}} \left(\mathbf{D}_r\mathbf{A}\right)_{k_1\ldots k_{N-r}} = \left(\mathbf{D}_r\mathbf{A}\right)_{l_1\ldots l_{N-r}} \tag{41.17}$$

Notice that (41.8) is a special case of (41.15) when the basis is right-handed and orthonormal, since in this case $e^* = 1, e_{ij} = \delta_{ij}$, and, from (21.5),

$$\det\begin{bmatrix} \delta_{j_1 l_1} & \cdot & \cdot & \cdot & \delta_{j_1 l_{N-r}} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \delta_{j_{N-r} l_1} & \cdot & \cdot & \cdot & \delta_{j_{N-r} l_{N-r}} \end{bmatrix} = \delta^{j_1\ldots j_r}_{l_1\ldots l_r}$$

Then as in (41.17) we have

$$\sum_{j_1<\cdots<j_{N-r}} \varepsilon^{i_1..i_r j_1..j_{N-r}} \delta^{j_1\ldots j_r}_{l_1\ldots l_r} = \varepsilon^{i_1..i_r l_1..l_{N-r}}$$

And thus (41.15) reduces to (41.8).

The duality operator can be used to define the *cross product* or the *vector product* for an oriented three-dimensional space. If we take $N = 3$, then $\mathbf{D}_2$ is an isomorphism from $\hat{\mathscr{T}}_2(\mathscr{V})$ to $\hat{\mathscr{T}}_1(\mathscr{V})$

$$\mathbf{D}_2 : \hat{\mathscr{T}}_2(\mathscr{V}) \to \hat{\mathscr{T}}_1(\mathscr{V}) = \mathscr{V} \tag{41.18}$$

so for any $\mathbf{u}$ and $\mathbf{v} \in \mathscr{V}$, $\mathbf{D}_2(\mathbf{u} \wedge \mathbf{v})$ is a vector in $\mathscr{V}$. We put

$$\mathbf{u} \times \mathbf{v} \equiv \mathbf{D}_2(\mathbf{u} \wedge \mathbf{v}) \tag{41.19}$$

called the *cross product* of $\mathbf{u}$ with $\mathbf{v}$.

We now prove that this definition is consistent with the classical definition of a cross product. This fact is more or less obvious. From (41.2), we have

$$\mathbf{E} * (\mathbf{u} \wedge \mathbf{v} \wedge \mathbf{w}) = \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) \qquad \text{for all } \mathbf{w} \in \mathscr{V} \tag{41.20}$$

where we have replaced the $*$ inner product on the right-hand side by the $\cdot$ inner product since $\mathbf{u} \times \mathbf{v}$ and $\mathbf{w}$ belong to $\mathscr{V}$. Equation (41.20) shows that

$$(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{u} = (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{v} = 0 \tag{41.21}_1$$

so that $\mathbf{u} \times \mathbf{v}$ is orthogonal to $\mathbf{v}$. That equation shows also that $\mathbf{u} \times \mathbf{v} = \mathbf{0}$ if and only if $\mathbf{u}, \mathbf{v}$ are linearly dependent. Further, if $\mathbf{u}, \mathbf{v}$ are linearly independent, and if $\mathbf{n}$ is the unit normal of $\mathbf{u}$ and $\mathbf{v}$ such that $\{\mathbf{u}, \mathbf{v}, \mathbf{n}\}$ from a right-handed basis, then

$$(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{n} > 0 \tag{41,21}_2$$

which means that $\mathbf{u} \times \mathbf{v}$ is pointing in the same direction as $\mathbf{n}$. Finally, from (40.33) we obtain

$$\begin{aligned} \left[\mathbf{E} * (\mathbf{u} \wedge \mathbf{v} \wedge \mathbf{w})\right]\left[\mathbf{E} * (\mathbf{u} \wedge \mathbf{v} \wedge \mathbf{w})\right] &= (\mathbf{u} \times \mathbf{v}) \cdot (\mathbf{u} \times \mathbf{v}) \\ &= \det \begin{bmatrix} \mathbf{u} \cdot \mathbf{u} & \mathbf{u} \cdot \mathbf{v} & 0 \\ \mathbf{v} \cdot \mathbf{u} & \mathbf{v} \cdot \mathbf{v} & 0 \\ 0 & 0 & 1 \end{bmatrix} = \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 - (\mathbf{u} \cdot \mathbf{v})^2 \\ &= \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 (1 - \cos^2 \theta) = \|\mathbf{u}\|^2 \|\mathbf{v}\|^2 \sin^2 \theta \end{aligned} \tag{41.22}$$

where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$. Hence we have shown that $\mathbf{u} \times \mathbf{v}$ is in the direction of $\mathbf{n}$ and has the magnitude $\|\mathbf{u}\|\|\mathbf{v}\| \sin\theta$, so that the definition of $\mathbf{u} \times \mathbf{v}$, based on (41.20), is consistent with the classical definition of the cross product.

From (41.19), the usual properties of the cross product are obvious; e.g., $\mathbf{u}\times\mathbf{v}$ is bilinear and skew-symmetric in $\mathbf{u}$ and $\mathbf{v}$. From (41.8), relative to a right-handed orthogonal basis $\{\mathbf{i}_1,\mathbf{i}_2,\mathbf{i}_3\}$ the components of $\mathbf{u}\times\mathbf{v}$ are given by

$$(\mathbf{u}\times\mathbf{v})_k = \sum_{i<j}\varepsilon_{ijk}\left(u_i v_j - u_j v_i\right) = \varepsilon_{ijk}u_i v_j \tag{41.23}$$

where the summations on i, j in the last term are unrestricted. Thus we have

$$\begin{aligned}(\mathbf{u}\times\mathbf{v})_1 &= u_2v_3 - u_3v_2\\ (\mathbf{u}\times\mathbf{v})_2 &= u_3v_1 - u_1v_3\\ (\mathbf{u}\times\mathbf{v})_3 &= u_1v_2 - u_2v_1\end{aligned} \tag{41.24}$$

On the other hand, if we use an arbitrary basis $\{\mathbf{e}^i\}$, then from (41.15) we have

$$(\mathbf{u}\times\mathbf{v})_l = \sum_{i<j}\left(u_i v_j - u_j v_i\right)e^*\varepsilon^{ijk}e_{kj} = e^* e_{kl}\varepsilon^{ijk}u_i v_j \tag{41.25}$$

where e^* is given by (40.31), namely

$$e^* = \varepsilon\left(\det\left[e^{ij}\right]\right)^{1/2}$$

Exercises

41.1 If $\mathbf{D}\in\hat{\mathscr{T}}_N(\mathscr{V})$, what is the value of $\mathbf{D}_N\mathbf{D}$?

41.2 If $\{\mathbf{e}^i\}$ is a basis of $\mathscr{V}$, determine the strict components of the dual $\mathbf{D}_r\left(\mathbf{e}^{i_1}\wedge\cdots\wedge\mathbf{e}^{i_r}\right)$.

Hint. The strict components of $\mathbf{e}^{i_1}\wedge\cdots\wedge\mathbf{e}^{i_r}$ are $\left\{\delta^{i_1\ldots i_r}_{j_1\ldots j_r}, j_1<\cdots<j_r\right\}$ since as in (41.17) we have

$$\sum_{j_1<\cdots<j_r}\delta^{i_1\ldots i_r}_{j_1\ldots j_r}\mathbf{e}^{j_1}\wedge\cdots\wedge\mathbf{e}^{j_r} = \mathbf{e}^{i_1}\wedge\cdots\wedge\mathbf{e}^{i_r} \tag{41.26}$$

41.3 If $\mathbf{A}$ is an endomorphism of a three-dimensional oriented inner product space $\mathscr{V}$, show that

$$\mathbf{Au}\cdot(\mathbf{Av}\times\mathbf{Aw}) = (\det\mathbf{A})\,\mathbf{u}\cdot(\mathbf{v}\times\mathbf{w}) \tag{41.27}$$

and if $\mathbf{A}$ is invertible, show that

$$\mathbf{Av} \times \mathbf{Aw} = \left(\det \mathbf{A}\right)\left(\mathbf{A}^{-1}\right)^{T} \mathbf{v} \times \mathbf{w} \tag{41.28}$$

Section 42. Transformation to the Contravariant Representation

So far we have used the covariant representations of skew-symmetric tensors only. We could, of course, develop the results of exterior algebra using the contravariant representations or even the mixcd representations of skew-symmetric tensors, since we have a fixed rule of transformation among the various representations based on the inner product, as explained in Section 35. In this section, we shall demonstrate the transformation from the covariant representation to the contravariant representation.

Recall that in general if $\mathbf{A}$ is a tensor of order r, then relative to any reciprocal bases $\{\mathbf{e}_i\}$ and $\{\mathbf{e}^i\}$ the contravariant components $A^{i_1\cdots i_r}$ of $\mathbf{A}$ are related to the covariant components $A_{i_1\cdots i_r}$ of $\mathbf{A}$ by [cf. (35.21)]

$$A^{i_1\cdots i_r} = e^{i_1 j_2}\cdots e^{i_r j_r} A_{j_1\cdots j_r} \tag{42.1}$$

This transforn1ation rule is valid for all *r*th order tensors, including skew-symmetric ones. However, as we have explained in Section 39, for skew-symmetric tensors it is convenient to use the strict components. Their transformation rule no longer has the simple form (42.1), since the summations on the repeated indices $j_1\cdots j_r$ on the right-hand side of (42.1) are unrestricted. We shall now derive the transformation rule between the contravariant and the covariant strict components of a skew-symmetric tensor.

Recall that the strict components of a skew-symmetric tensor are simply the ordinary components restricted to an increasing set of indices, as shown in (39.10). In order to obtain an equivalent form of (42.1) using the strict components of $\mathbf{A}$ only, we must replace the right-hand side of that equation by a restricted summation. For this purpose we use the identity [cf. (37.26)]

$$A_{j_1\cdots j_r} = A_{[j_1\cdots j_r]} = \frac{1}{r!}\delta^{k_1\cdots k_r}_{j_1\cdots j_r} A_{k_1\cdots k_r} \tag{42.2}$$

where, by assumption, $\mathbf{A}$ is skew-symmetric. Substituting (42.2) into (42.1), we obtain

$$A^{i_1\cdots i_r} = \frac{1}{r!}\delta^{k_1\cdots k_r}_{j_1\cdots j_r} e^{i_1 j_1}\cdots e^{i_r j_r} A_{k_1\cdots k_r} \tag{42.3}$$

Now, since the coefficient of $A_{k_1 \ldots k_r}$ is skew-symmetric, we can restrict the summations on $k_1 \cdots k_r$ to the increasing order by removing the factor $1/r!$, that is

$$A^{i_1 .. i_r} = \sum_{k_1 < \cdots < k_r} \delta^{k_1 \ldots k_r}_{j_1 \ldots j_r} e^{i_1 j_1} \cdots e^{i_r j_r} A_{k_1 \ldots k_r} \tag{42.4}$$

This is the desired transformation rule from the covariant strict components to the contravariant ones. Clearly, the inverse of (42.4) is

$$A_{i_1 .. i_r} = \sum_{k_1 < \cdots < k_r} \delta^{j_1 \ldots j_r}_{k_1 \ldots k_r} e_{i_1 j_1} \cdots e_{i_r j_r} A^{k_1 \ldots k_r} \tag{42.5}$$

which is the transformation rule from the contravariant strict components to the covariant ones.

From (21.21) we see that (42.4) is equivalent to

$$\begin{aligned} A^{i_1 .. i_r} &= \sum_{k_1 < \cdots < k_r} e^{i_1 .. i_r, k_1 \ldots k_r} A_{k_1 \ldots k_r} \\ &= \sum_{k_1 < \cdots < k_r} \det \begin{bmatrix} e^{i_1 k_1} & \cdot & \cdot & \cdot & e^{i_1 k_r} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ e^{i_r k_1} & \cdot & \cdot & \cdot & e^{i_r k_r} \end{bmatrix} A_{k_1 \ldots k_r} \end{aligned} \tag{42.6}$$

while (42.5) is equivalent to

$$\begin{aligned} A_{i_1 .. i_r} &= \sum_{k_1 < \cdots < k_r} e_{i_1 .. i_r, k_1 \ldots k_r} A^{k_1 \ldots k_r} \\ &= \sum_{k_1 < \cdots < k_r} \det \begin{bmatrix} e_{i_1 k_1} & \cdot & \cdot & \cdot & e_{i_1 k_r} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ e_{i_r k_1} & \cdot & \cdot & \cdot & e_{i_r k_r} \end{bmatrix} A^{k_1 \ldots k_r} \end{aligned} \tag{42.7}$$

In particular, when $r = N$, we have

$$A^{1\cdots N} = \det\left[e^{ij}\right] A_{1\cdots N}, \qquad A_{1\cdots N} = \det\left[e_{ij}\right] A^{1\cdots N} \tag{42.8}$$

From the transformation rules (42.6) and (42.7), or directly from the skew-symmetry of the wedge product, we see that the product basis of any reciprocal bases $\{\mathbf{e}_i\}$ and $\{\mathbf{e}^i\}$ obeys the following transformation rules:

$$\mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r} = \sum_{k_1 < \cdots < k_r} \det \begin{bmatrix} e^{i_1 k_1} & \cdot & \cdot & \cdot & e^{i_1 k_r} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ e^{i_r k_1} & \cdot & \cdot & \cdot & e^{i_r k_r} \end{bmatrix} \mathbf{e}_{k_1} \wedge \cdots \wedge \mathbf{e}_{k_r} \tag{42.9}$$

and

$$\mathbf{e}_{i_1} \wedge \cdots \wedge \mathbf{e}_{i_r} = \sum_{k_1 < \cdots < k_r} \det \begin{bmatrix} e_{i_1 k_1} & \cdot & \cdot & \cdot & e_{i_1 k_r} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ e_{i_r k_1} & \cdot & \cdot & \cdot & e_{i_r k_r} \end{bmatrix} \mathbf{e}^{k_1} \wedge \cdots \wedge \mathbf{e}^{k_r} \tag{42.10}$$

In deriving (42.9) and (42.10) we have used the fact that the strict covariant components of $\mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r}$ are

$$\left\{\delta^{i_1 \cdots i_r}_{j_1 \cdots j_r},\, j_1 < \cdots < j_r\right\}$$

and, likewise, the strict contravariant components of $\mathbf{e}_{i_1} \wedge \cdots \wedge \mathbf{e}_{i_r}$ are

$$\left\{\delta^{j_1 \cdots j_r}_{i_1 \cdots i_r},\, j_1 < \cdots < j_r\right\}$$

as shown by (41.26). If we apply (42.9) and (42.10) to the product bases $\mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N$ and $\mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_N$, we get

$$\begin{aligned}\mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N &= \det\left[e^{ij}\right]\mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_N \\ \mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_N &= \det\left[e_{ij}\right]\mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N\end{aligned} \tag{42.11}$$

The product bases

$$\left\{\mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r}, i_1 < \cdots < i_r\right\} \qquad \text{and} \qquad \left\{\mathbf{e}_{i_1} \wedge \cdots \wedge \mathbf{e}_{i_r}, i_1 < \cdots < i_r\right\}$$

are reciprocal bases with respect to the inner product $*$, since from (39.30) we have

$$\left(\mathbf{e}^{i_1} \wedge \cdots \wedge \mathbf{e}^{i_r}\right) * \left(\mathbf{e}_{j_1} \wedge \cdots \wedge \mathbf{e}_{j_r}\right) = \det\begin{bmatrix} \delta^{i_1}_{j_1} & \cdot & \cdot & \cdot & \delta^{i_1}_{j_r} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \delta^{i_r}_{j_1} & \cdot & \cdot & \cdot & \delta^{i_r}_{j_r} \end{bmatrix} = \delta^{i_1 \cdots i_r}_{j_1 \cdots j_r} \tag{42.12}$$

In particular, when $r = N$ we have

$$\left(\mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N\right) * \left(\mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_N\right) = 1 \tag{42.13}$$

From (42.12), we can compute the $*$ inner product of any two *r*th order skew-symmetric tensors

$$\mathbf{A} * \mathbf{B} = \sum_{k_1 < \cdots < k_r} A^{k_1 \ldots k_r} B_{k_1 \ldots k_r} = \sum_{k_1 < \cdots < k_r} A_{k_1 \ldots k_r} B^{k_1 \ldots k_r} \tag{42.14}$$

These formulas are equivalent to the formula (39.31), which is based on the covariant strict components of $\mathbf{A}$ and $\mathbf{B}$.

For an oriented space we have defined the density e^* of a basis $\left\{\mathbf{e}^i\right\}$ to be the components of $\mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N$ relative to the positive unit density $\mathbf{E}$, namely

$$\mathbf{e}^1 \wedge \cdots \wedge \mathbf{e}^N = e^* \mathbf{E} \tag{42.15}$$

as shown by (40.28). Clearly we can define a similar component for the basis $\left\{\mathbf{e}_i\right\}$, namely

$$\mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_N = e\mathbf{E} \tag{42.16}$$

Further, from (42.13) the components e and e^* are related by

$$ee^* = 1 \tag{42.17}$$

In view of this relation, we call e the *volume* of $\{\mathbf{e}_i\}$. Then $\{\mathbf{e}_i\}$ is *positively oriented* or *right-handed* if its volume e is positive; otherwise, $\{\mathbf{e}_i\}$ is *negatively oriented* or *left-handed.* As before, a unimodular basis $\{\mathbf{e}_i\}$ is defined by the condition that the *absolute volume* $|e|$ is equal to unity.

We can compute the absolute volume by

$$e^2 = \det\left[e_{ij}\right] \tag{42.18}$$

or equivalently

$$|e| = \left(\det\left[e_{ij}\right]\right)^{1/2} \tag{42.19}$$

The proof is the same as that of (40.29) and (40.30). Substituting (42.19) into (42.17), we have also

$$\mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_N = \varepsilon\left(\det\left[e_{ij}\right]\right)^{1/2}\mathbf{E} \tag{42.20}$$

where ε is $+$ if $\{\mathbf{e}_i\}$ is positively oriented and it is $-$ if $\{\mathbf{e}_i\}$ is negatively oriented.

Using the contravariant components, we can simplify the formulas of the preceding section somewhat; e.g., from (42.6) we can rewrite (41.15) as

$$\left(\mathbf{D}_r\mathbf{A}\right)^{j_1\cdots j_{N-r}} = \sum_{i_1<\cdots<i_r} A_{i_1\cdots i_r} e^* \varepsilon^{i_1\cdots i_r j_1\cdots j_{N-r}} \tag{42.21}$$

which is equivalent to

$$\left(\mathbf{D}_r\mathbf{A}\right)_{j_1\cdots j_{N-r}} = \sum_{i_1<\cdots<i_r} A^{i_1\cdots i_r} e\varepsilon_{i_1\cdots i_r j_1\cdots j_{N-r}} \tag{42.22}$$

Similarly, (41.25) can be rewritten as

$$\left(\mathbf{u}\times\mathbf{v}\right)^k = e^*\varepsilon^{ijk}u_i v_j \tag{42.23}$$

which is equivalent to

$$\left(\mathbf{u}\times\mathbf{v}\right)_k = e\varepsilon_{ijk}u^i v^j \tag{42.24}$$

Exercises

42.1 Prove the formula (42.22).

42.2 Use the result of Exercise 41.2 or the transformation rules (42.21)and (42.22) and show that

$$\mathbf{D}_r\left(\mathbf{e}^{i_1}\wedge\cdots\wedge\mathbf{e}^{i_r}\right) = \sum_{j_1<\cdots<j_r} e^*\varepsilon^{i_1\cdots i_r j_1\cdots j_{N-r}}\mathbf{e}_{j_1}\wedge\cdots\wedge\mathbf{e}_{j_{N-r}} \tag{42.25}$$

which is equivalent to

$$\mathbf{D}_r\left(\mathbf{e}_{i_1}\wedge\cdots\wedge\mathbf{e}_{i_r}\right) = \sum_{j_1<\cdots<j_r} e^*\varepsilon_{i_1\cdots i_r j_1\cdots j_{N-r}}\mathbf{e}^{j_1}\wedge\cdots\wedge\mathbf{e}^{j_{N-r}} \tag{42.26}$$

42.3 Show that has the representations

$$\mathbf{E} = \frac{1}{e}\varepsilon^{i_1\cdots i_N}\mathbf{e}_{i_1}\otimes\cdots\otimes\mathbf{e}_{i_N} = e\varepsilon_{j_1\cdots j_N}\mathbf{e}^{j_1}\otimes\cdots\otimes\mathbf{e}^{j_N} \tag{42.27}$$

where

$$\frac{1}{e}\varepsilon^{i_1\cdots i_N} = e\varepsilon_{j_1\cdots j_N} e^{i_1 j_1}\cdots e^{i_N j_N} \tag{42.28}$$

and

$$e = \varepsilon\left(\det\left[e_{ij}\right]\right)^{1/2} \tag{42.29}$$

INDEX

The page numbers in this index refer to the versions of Volumes I and II of the online versions. In addition to the information below, the search routine in Adobe Acrobat will be useful to the reader. Pages 1-294 will be found in the online editions of Volume 1, pages 295-520 in Volume 2