HILDEBRAND

Methods of Applied Mathematics

SECOND EDITION

FRANCIS B. HILDEBRAND

Associate Professor of Mathematics
Massachusetts Institute of Technology

Methods of Applied Mathematics

SECOND EDITION

PRENTICE-HALL, INC.

ENGLEWOOD CLIFFS, NEW JERSEY

 Library of Congress Catalog Card Number: 65–14941. Printed in the United States of America. C-57920.

PRENTICE-HALL INTERNATIONAL, INC., *London*
PRENTICE-HALL OF AUSTRALIA, PTY., LTD., *Sydney*
PRENTICE-HALL OF CANADA, LTD., *Toronto*
PRENTICE-HALL OF INDIA (PRIVATE) LTD., *New Delhi*
PRENTICE-HALL OF JAPAN, INC., *Tokyo*

Preface

The principal aim of this volume is to place at the disposal of the engineer or physicist the basis of an intelligent working knowledge of a number of facts and techniques relevant to some fields of mathematics which often are not treated in courses of the "Advanced Calculus" type, but which are useful in varied fields of application.

Many students in the fields of application have neither the time nor the inclination for the study of detailed treatments of each of these topics from the classical point of view. However, efficient use of facts or techniques depends strongly upon a substantial understanding of the basic underlying principles. For this reason, care has been taken throughout the text either to provide a rigorous proof, when the proof is believed to contribute to an understanding of the desired result, or to state the result as precisely as possible and indicate why it might have been *formally* anticipated.

In each chapter, the treatment consists of showing how typical problems may arise, of establishing those parts of the relevant theory which are of principal practical significance, and of developing techniques for analytical and numerical analysis and problem solving.

Whereas experience gained from a course on the Advanced Calculus level is presumed, the treatments are largely self-contained, so that the nature of this preliminary course is not of great importance.

In order to increase the usefulness of the volume as a basic or supplementary text, and as a reference volume, an attempt has been made to organize the material so that there is little essential interdependence among the chapters, and considerable flexibility exists with regard to the omission of topics within chapters. In addition, a substantial amount of supplementary material is included in annotated problems which complement numerous exercises, of varying difficulty, arranged in correspondence with successive

sections of the text at the end of the chapters. Answers to all problems are either incorporated into their statement or listed at the end of the book.

The first chapter deals principally with *linear algebraic equations, quadratic and Hermitian forms*, and operations with *vectors and matrices*, with special emphasis on the concept of characteristic values. A brief summary of corresponding results in *function space* is included for comparison, and for convenient reference. Whereas a considerable amount of material is presented, particular care was taken here to arrange the demonstrations in such a way that maximum flexibility in selection of topics is present.

The first portion of the second chapter introduces the variational notation and derives the Euler equations relevant to a large class of problems in the *calculus of variations*. More than usual emphasis is placed on the significance of natural boundary conditions. Generalized coordinates, Hamilton's principle, and Lagrange's equations are treated and illustrated within the framework of this theory. The chapter concludes with a discussion of the formulation of minimal principles of more general type, and with the application of direct and semidirect methods of the calculus of variations to the exact and approximate solution of practical problems.

The concluding chapter deals with the formulation and theory of linear *integral equations*, and with exact and approximate methods for obtaining their solutions, particular emphasis being placed on the several equivalent interpretations of the relevant Green's function. Considerable supplementary material is provided here in annotated problems.

The present text is a revision of corresponding chapters of the first edition, published in 1952. It incorporates a number of changes in method of presentation and in notation, as well as some new material and additional problems and exercises. A revised and expanded version of the earlier material on difference equations and on finite difference methods is to appear separately.

Many compromises between mathematical elegance and practical significance were found to be necessary. However, it is hoped that the text will serve to ease the way of the engineer or physicist into the more advanced areas of applicable mathematics, for which his need continues to increase, without obscuring from him the existence of certain *difficulties*, sometimes implied by the phrase "It can be shown," and without failing to warn him of certain *dangers* involved in formal application of techniques beyond the limits inside which their validity has been well established.

The author is indebted to colleagues and students in various fields for help in selecting and revising the content and presentation, and particularly to Professor Albert A. Bennett for many valuable criticisms and suggestions.

FRANCIS B. HILDEBRAND

Contents

CHAPTER TWO

Calculus of Variations and Applications 119

CHAPTER THREE

Integral Equations 222

Methods of Applied Mathematics

CHAPTER ONE

Matrices and Linear Equations

1.1. Introduction. In many fields of analysis we find it necessary to deal with an *ordered set* of elements, which may be numbers or functions. In particular, we may deal with an ordinary *sequence* of the form

$$a_1, a_2, \cdots, a_n$$

or with a two-dimensional *array* such as the rectangular arrangement

$$\begin{array}{llll} a_{11}, & a_{12}, & \cdots, & a_{1n} \\ a_{21}, & a_{22}, & \cdots, & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1}, & a_{m2}, & \cdots, & a_{mn}, \end{array}$$

consisting of m rows and n columns.

When suitable laws of equality, addition and subtraction, and multiplication are associated with sets of such rectangular arrays, the arrays are called *matrices*, and are then designated by a special symbolism. The laws of combination are specified in such a way that the matrices so defined are of frequent usefulness in both practical and theoretical considerations.

Since matrices are perhaps most intimately associated with sets of *linear algebraic equations*, it is desirable to investigate the general nature of the solutions of such sets of equations by elementary methods, and hence to provide a basis for certain definitions and investigations which follow.

1.2. Linear equations. The Gauss-Jordan reduction. We deal first with the problem of attempting to obtain solutions of a set of m linear equations

in n unknown variables $x_1, x_2, \cdots, x_n$, of the form

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= c_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= c_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= c_m \end{aligned}\right\}, \tag{1}$$

by direct calculation.

Under the assumption that (1) does indeed possess a solution, the *Gauss-Jordan reduction* proceeds as follows:

First Step. Suppose that $a_{11} \neq 0$. (Otherwise, renumber the equations or variables so that this is so.) Divide both sides of the first equation by a_{11}, so that the resultant equivalent equation is of the form

$$x_1 + a'_{12}x_2 + \cdots + a'_{1n}x_n = c'_1. \tag{2}$$

Multiply both sides of (2) successively by $a_{21}, a_{31}, \cdots, a_{m1}$, and subtract the respective resultant equations from the second, third, $\cdots$, mth equations of (1), to reduce (1) to the form

$$\left.\begin{aligned} x_1 + a'_{12}x_2 + \cdots + a'_{1n}x_n &= c'_1, \\ a'_{22}x_2 + \cdots + a'_{2n}x_n &= c'_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots & \\ a'_{m2}x_2 + \cdots + a'_{mn}x_n &= c'_m \end{aligned}\right\}. \tag{3}$$

Second Step. Suppose that $a'_{22} \neq 0$. (Otherwise, renumber the equations or variables so that this is so.) Divide both sides of the second equation in (3) by a'_{22}, so that this equation takes the form

$$x_2 + a''_{23}x_3 + \cdots + a''_{2n}x_n = c''_2, \tag{4}$$

and use this equation, as in the first step, to eliminate the coefficient of x_2 in *all other equations* in (3), so that the set of equations becomes

$$\left.\begin{aligned} x_1 + a''_{13}x_3 + \cdots + a''_{1n}x_n &= c''_1, \\ x_2 + a''_{23}x_3 + \cdots + a''_{2n}x_n &= c''_2, \\ a''_{33}x_3 + \cdots + a''_{3n}x_n &= c''_3, \\ \cdots\cdots\cdots\cdots\cdots\cdots & \\ a''_{m3}x_3 + \cdots + a''_{mn}x_n &= c''_m \end{aligned}\right\}. \tag{5}$$

Remaining Steps. Continue the above process r times until it terminates, that is, until $r = m$ *or* until the coefficients of all x's are *zero* in all equations following the rth equation. We shall speak of these $m - r$ equations as the *residual equations*, when $m > r$.

There then exist two alternatives. *First*, it may happen that, with $m > r$, one or more of the residual equations has a nonzero right-hand member, and hence is of the form $0 = c_k^{(r)}$ (where in fact $c_k^{(r)} \neq 0$). In this case, the assumption that a solution of (1) exists leads to a *contradiction*, and hence *no solution exists*. The set (1) is then said to be *inconsistent* or *incompatible*.

Otherwise, no contradiction exists, and the set (1) of m equations is reduced to a set of r equations which, after a transposition, can be written in the form

$$\left.\begin{aligned} x_1 &= \gamma_1 + \alpha_{11}x_{r+1} + \cdots + \alpha_{1,n-r}x_n, \\ x_2 &= \gamma_2 + \alpha_{21}x_{r+1} + \cdots + \alpha_{2,n-r}x_n, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_r &= \gamma_r + \alpha_{r1}x_{r+1} + \cdots + \alpha_{r,n-r}x_n \end{aligned}\right\}, \tag{6}$$

where the γ's and α's are specific constants related to the coefficients in (1). Since each of the steps in the reduction of the set (1) to the set (6) is *reversible*, it follows that the two sets are *equivalent*, in the sense that each set implies the other. Hence, in this case the most general solution of (1) expresses each of the r variables $x_1, x_2, \cdots, x_r$ as a specific constant plus a specific linear combination of the remaining $n - r$ variables, each of which can be assigned *arbitrarily*.

If $r = n$, a *unique* solution is obtained. Otherwise, we say that an *$(n - r)$-parameter family* of solutions exists. The number $n - r = d$ may be called the *defect* of the system (1). We notice that if the system (1) is consistent and r is less than m, then $m - r$ of the equations (namely, those which correspond to the residual equations) are actually ignorable, and hence must be implied by the remaining r equations.

The reduction may be illustrated by considering the four simultaneous equations

$$\left.\begin{aligned} x_1 + 2x_2 - x_3 - 2x_4 &= -1, \\ 2x_1 + x_2 + x_3 - x_4 &= 4, \\ x_1 - x_2 + 2x_3 + x_4 &= 5, \\ x_1 + 3x_2 - 2x_3 - 3x_4 &= -3 \end{aligned}\right\}. \tag{7}$$

It is easily verified that after two steps in the reduction one obtains the equivalent set

$$\left.\begin{aligned} x_1 + x_3 &= 3, \\ x_2 - x_3 - x_4 &= -2, \\ 0 &= 0, \\ 0 &= 0 \end{aligned}\right\}.$$

Hence the system is of defect *two*. If we write $x_3 = C_1$ and $x_4 = C_2$, it follows that the general solution can be expressed in the form

$$x_1 = 3 - C_1, \quad x_2 = -2 + C_1 + C_2, \quad x_3 = C_1, \quad x_4 = C_2, \tag{8a}$$

where C_1 and C_2 are arbitrary constants. This two-parameter family of solutions can also be written in the symbolic form

$$\{x_1, x_2, x_3, x_4\} = \{3, -2, 0, 0\} + C_1\{-1, 1, 1, 0\} + C_2\{0, 1, 0, 1\}. \tag{8b}$$

It follows also that the third and fourth equations of (7) must be consequences of the first two equations. Indeed, the third equation is obtained by subtracting the first from the second, and the fourth by subtracting one-third of the second from five-thirds of the first.

The Gauss-Jordan reduction is useful in actually obtaining numerical solutions of sets of linear equations,* and it has been presented here also for the purpose of motivating certain definitions and terminologies which follow.

1.3. Matrices. The set of equations (1) can be visualized as representing a *linear transformation* in which the set of n numbers $\{x_1, x_2, \cdots, x_n\}$ is transformed into the set of m numbers $\{c_1, c_2, \cdots, c_m\}$.

The rectangular array of the coefficients a_{ij} specifies the transformation. Such an array is often enclosed in square brackets and denoted by a single boldface capital letter,

$$\mathbf{A} \equiv [a_{ij}] \equiv \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \tag{9}$$

and is called an $m \times n$ *matrix* when certain laws of combination, yet to be specified, are laid down. In the symbol a_{ij}, representing a typical element, the *first* subscript (here i) denotes the row and the *second* subscript (here j) the column occupied by the element.

* In place of eliminating x_k from *all* equations except the kth, in the kth step, one may eliminate x_k only in those equations *following* the kth equation. When the process terminates, after r steps, the rth unknown is given explicitly by the rth equation. The $(r-1)$th unknown is then determined by substitution in the $(r-1)$th equation, and the solution is completed by working back in this way to the first equation. The method just outlined is associated with the name of *Gauss*. In order that the "round-off" errors be as small as possible, it is usually desirable that the sequence of eliminations be ordered such that the coefficient of x_k in the equation used to eliminate x_k is as large as possible in absolute value, relative to the remaining coefficients in that equation.

A modification of this method, due to Crout (Reference 3), which is particularly well adapted to the use of desk computing machines, is described in an appendix.

The sets of quantities $x_i\,(i = 1, 2. \cdots, n)$ and $c_i\,(i = 1, 2, \cdots, m)$ are conventionally represented as matrices of *one column* each. In order to emphasize the fact that a matrix consists of only one column, it is sometimes convenient to denote it by a lower-case boldface letter and to enclose it in braces, rather than brackets, and so to write

$$\mathbf{x} \equiv \{x_i\} \equiv \begin{Bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{Bmatrix}, \qquad \mathbf{c} \equiv \{c_i\} \equiv \begin{Bmatrix} c_1 \\ c_2 \\ \cdot \\ \cdot \\ \cdot \\ c_m \end{Bmatrix}. \tag{10a,b}$$

For convenience in writing, the elements of a one-column matrix are frequently arranged horizontally,

$$\mathbf{x} = \{x_1, x_2, \cdots, x_n\},$$

the use of braces then being *necessary* to indicate the transposition.

Other symbols, such as parentheses or double vertical lines, are also used to enclose matrix arrays.

If we interpret (1) as stating that the matrix $\mathbf{A}$ transforms the one-column matrix $\mathbf{x}$ into the one-column matrix $\mathbf{c}$, it is natural to write the transformation in the form

$$\mathbf{A}\,\mathbf{x} = \mathbf{c}, \tag{11}$$

where $\mathbf{A} = [a_{ij}]$, $\mathbf{x} = \{x_i\}$, and $\mathbf{c} = \{c_i\}$.

On the other hand, the set of equations (1) can be written in the form

$$\sum_{k=1}^{n} a_{ik}x_k = c_i \qquad (i = 1, 2, \cdots, m), \tag{12a}$$

which leads to the matrix equation

$$\left\{\sum_{k=1}^{n} a_{ik}x_k\right\} = \{c_i\}. \tag{12b}$$

Hence, if (11) and (12b) are to be equivalent, we are led to the *definition*

$$\mathbf{A}\,\mathbf{x} = [a_{ik}]\{x_k\} \equiv \left\{\sum_{k=1}^{n} a_{ik}x_k\right\}. \tag{13}$$

Formally, we merely replace the *column* subscript in the general term of the *first* factor by a *dummy index k*, replace the *row* subscript in the general

term of the *second* factor by the same dummy index, and sum over that index.*

The definition clearly is applicable only when the number of *columns* in the *first* factor is equal to the number of *rows* (elements) in the *second* factor. Unless this condition is satisfied, the product is undefined.

We notice that a_{ik} is the element in the ith row and kth column of **A**, and that x_k is the kth element in the one-column matrix **x**. Since i ranges from 1 to m in a_{ij}, the definition (13) states that the product of an $m \times n$ matrix into an $n \times 1$ matrix is an $m \times 1$ matrix (m elements in one column). The ith element in the product is obtained from the ith row of the first factor and the single column of the second factor, by multiplying together the first elements, second elements, and so forth, and adding these products together algebraically.

Thus, for example, the definition leads to the result

$$\begin{bmatrix} 1 & 0 \\ 2 & 1 \\ -1 & 2 \end{bmatrix} \begin{Bmatrix} 1 \\ 2 \end{Bmatrix} = \begin{Bmatrix} 1 \cdot 1 + 0 \cdot 2 \\ 2 \cdot 1 + 1 \cdot 2 \\ -1 \cdot 1 + 2 \cdot 2 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 4 \\ 3 \end{Bmatrix}.$$

Now suppose that the n variables $x_1, \cdots, x_n$ are expressed as linear combinations of s new variables $y_1, \cdots, y_s$, that is, that a set of relations holds of the form

$$x_i = \sum_{k=1}^{s} b_{ik} y_k \qquad (i = 1, 2, \cdots, n). \tag{14}$$

If the original variables satisfy (12a), the equations satisfied by the new variables are obtained by introducing (14) into (12a). In addition to replacing i by k in (14), for this introduction, we must replace k in (14) by a *new*

* Very frequently, in the literature, use is made of the so-called *summation convention*, in which the sigma symbol is omitted in a sum such as

$$\sum_{k=1}^{n} a_{ik} x_k$$

with the understanding that the notation $a_{i\kappa} x_k$ then is to indicate the result of *summing* the product with respect to the *repeated* index, over the range of that index. Similarly, with this convention one would write $a_{ik} b_{kl} c_{lj}$ when summations with respect to both k and l are intended. An explicit statement then must be made when the *element* a_{kk} is to be distinguished from the *sum*

$$\sum_{k=1}^{n} a_{kk}$$

or in other cases when the summation convention temporarily is to be abandoned. The summation convention will not be used in this text.

dummy index, say l, to avoid ambiguity of notation. The result of the substitution then takes the form

$$\sum_{k=1}^{n} a_{ik}\left(\sum_{l=1}^{s} b_{kl}y_l\right) = c_i \qquad (i = 1, 2, \cdots, m), \tag{15a}$$

or, since the order in which the finite sums are formed is immaterial,

$$\sum_{l=1}^{s}\left(\sum_{k=1}^{n} a_{ik}b_{kl}\right)y_l = c_i \qquad (i = 1, 2, \cdots, m). \tag{15b}$$

In matrix notation, the transformation (14) takes the form

$$\mathbf{x} = \mathbf{B}\,\mathbf{y} \tag{16}$$

and, corresponding to (15a), the introduction of (16) into (11) gives

$$\mathbf{A}(\mathbf{B}\,\mathbf{y}) = \mathbf{c}. \tag{17}$$

But if we write

$$p_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj} \qquad \begin{pmatrix} i = 1, 2, \cdots, m \\ j = 1, 2, \cdots, s \end{pmatrix} \tag{18}$$

equation (15b) takes the form

$$\sum_{l=1}^{s} p_{il}y_l = c_i \qquad (i = 1, 2, \cdots, m),$$

and hence, in accordance with (12a) and (13), the matrix form of the transformation (15b) is

$$\mathbf{P}\,\mathbf{y} = \mathbf{c}, \tag{19}$$

where $\mathbf{P} = [p_{ij}]$.

Thus it follows that the result of operating on $\mathbf{y}$ by $\mathbf{B}$, and on the product by $\mathbf{A}$ [given by the left-hand member of (17)], is the same as the result of operating on $\mathbf{y}$ directly by the matrix $\mathbf{P}$. We accordingly *define* this matrix to be the product $\mathbf{A}\,\mathbf{B}$,

$$\mathbf{A}\,\mathbf{B} = [a_{ik}][b_{kj}] = \left[\sum_{k=1}^{n} a_{ik}b_{kj}\right]. \tag{20}$$

The desirable relation

$$\mathbf{A}(\mathbf{B}\,\mathbf{y}) = (\mathbf{A}\,\mathbf{B})\mathbf{y}$$

then is a consequence of this definition.

Recalling that the first subscript in each case is the row index and the second the column index, we see that if the first factor of (20) has m rows and n columns, and the second n rows and s columns, the index i in the right-hand member may vary from 1 to m while the index j in that member may vary from 1 to s. Hence, the *product of an $m \times n$ matrix into an $n \times s$ matrix is an $m \times s$ matrix.* The element p_{ij} in the ith row and jth column of the product is formed by multiplying together corresponding elements of

the ith *row* of the *first* factor and the jth *column* of the *second* factor, and adding the results algebraically. In particular, the definition (20) properly reduces to (13) when $s = 1$.

Thus, for example, we have

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

$$= \begin{bmatrix} (1 \cdot 1 + 0 \cdot 1 + 1 \cdot 2) & (1 \cdot 2 + 0 \cdot 0 + 1 \cdot 1) & (1 \cdot 1 + 0 \cdot 1 + 1 \cdot 0) \\ (1 \cdot 1 - 2 \cdot 1 + 1 \cdot 2) & (1 \cdot 2 - 2 \cdot 0 + 1 \cdot 1) & (1 \cdot 1 - 2 \cdot 1 + 1 \cdot 0) \end{bmatrix}$$

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

We notice that **A B** is defined only if the number of *columns* in **A** is equal to the number of *rows* in **B**. In this case, the two matrices are said to be *conformable* in the order stated.

If **A** is an $m \times n$ matrix and **B** an $n \times m$ matrix, then **A** and **B** are conformable in either order, the product **A B** then being a *square* matrix of order m and the product **B A** a square matrix of order n. Even in the case when **A** and **B** are square matrices of the same order the products **A B** and **B A** are not generally equal. For example, in the case of two square matrices of order two we have

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

and also

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

Thus, in multiplying **B** by **A** in such cases, we must carefully distinguish *pre*multiplication (**A B**) from *post*multiplication (**B A**).

Two $m \times n$ matrices are said to be *equal* if and only if corresponding elements in the two matrices are equal.

The *sum* of two $m \times n$ matrices $[a_{ij}]$ and $[b_{ij}]$ is defined to be the matrix $[a_{ij} + b_{ij}]$. Further, the product of a number k and a matrix $[a_{ij}]$ is defined to be the matrix $[ka_{ij}]$, in which *each* element of the original matrix is multiplied by k.

From the preceding definitions, it is easily shown that, if **A**, **B**, and **C** are each $m \times n$ matrices, *addition* is *commutative* and *associative*:

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}, \qquad \mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}. \tag{21}$$

Also, if the relevant products are defined, *multiplication* of matrices is *associative*,

$$\mathbf{A}(\mathbf{B}\,\mathbf{C}) = (\mathbf{A}\,\mathbf{B})\mathbf{C}, \tag{22}$$

and *distributive*,

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{A}\,\mathbf{B} + \mathbf{A}\,\mathbf{C}, \qquad (\mathbf{B} + \mathbf{C})\mathbf{A} = \mathbf{B}\,\mathbf{A} + \mathbf{C}\,\mathbf{A}, \tag{23}$$

but, in general, *not commutative.**

It is consistent with these definitions to divide a given matrix into smaller submatrices, the process being known as the *partitioning* of a matrix. Thus, for example, we may partition a square matrix **A** of order three *symmetrically* as follows:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \equiv \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}$$

where the elements of the partitioned form are the *matrices*

$$\mathbf{B}_{11} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \qquad \mathbf{B}_{12} = \begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix},$$

$$\mathbf{B}_{21} = [a_{31} \quad a_{32}], \qquad \mathbf{B}_{22} = [a_{33}].$$

If, for example, a second square matrix of order three is similarly partitioned, the submatrices can be treated as single elements and the usual laws of matrix multiplication and addition can be applied to the two matrices so partitioned, as is easily verified.

More generally, if two conformable matrices in a product are partitioned, necessary and sufficient conditions that this statement apply are that to each vertical partition line separating *columns* r and $r + 1$ in the *first* factor there correspond a horizontal partition line separating *rows* r and $r + 1$ in the *second* factor, and that no additional *horizontal* partition lines be present in the *second* factor.

In particular, if we think of the matrices **B** and **C** in the relation $\mathbf{A}\,\mathbf{B} = \mathbf{C}$ as being partitioned into columns, we rediscover the fact that each column of **C** can be obtained by premultiplying the corresponding column of **B** by the matrix **A**. Similarly, by thinking of **A** and **C** as being partitioned into *rows*, we see that each row of **C** is the result of postmultiplying the corresponding row of **A** by the matrix **B**.

* Except when otherwise noted, it will be supposed in this text that the *scalars* which comprise the *elements* of matrices are real or complex numbers. However, many of the results to be established are also valid (when suitably interpreted) for many other sets of permissible elements.

1.4. Determinants. Cramer's rule. In this section we review certain properties of *determinants*. Associated with any *square* matrix $[a_{ij}]$ of order n we define the *determinant* $|\mathbf{A}| = |a_{ij}|$,

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix},$$

as a *number* obtained as the sum of all possible products in each of which there appears one and only one element from each row and each column, each such product being prefixed by a plus or minus sign according to the following rule: *Let the elements involved in a given product be joined in pairs by line segments. If the total number of such segments sloping upward to the right is even, prefix a plus sign to the product. Otherwise, prefix a negative sign.**

From this definition, the following properties of determinants, which greatly simplify their actual evaluation, are easily established:

1. If all elements of any row or column of a square matrix are zeros, its determinant is zero.

2. The value of the determinant is unchanged if the rows and columns of the matrix are interchanged.

3. If two rows (or two columns) of a square matrix are interchanged, the sign of its determinant is changed.

4. If all elements of one row (or one column) of a square matrix are multiplied by a number k, the determinant is multiplied by k.

5. If corresponding elements of two rows (or two columns) are equal or in a constant ratio, then the determinant is zero.

6. If each element in one row (or one column) is expressed as the sum of two terms, then the determinant is equal to the sum of two determinants, in each of which one of the two terms is deleted in each element of that row (or column).

7. If to the elements of any row (column) are added k times the corresponding elements of any other row (column), the determinant is unchanged.†

* This statement of the rule of signs is equivalent to the statement which involves *inversions* of subscripts. It possesses the advantage of being readily applicable in actual cases when the elements are numbers (or functions) and are not provided with explicit subscripts. Also, with this statement of the rule, the proofs of the properties which follow are in general simplified.

† It can be shown that if we impose the condition that Properties 4 and 7 hold, and in addition impose the requirement that the determinant be unity when the diagonal elements are unity and all other elements are zero, then these conditions imply all other properties of determinants, and may serve as the definition of a determinant.

From these properties others may be deduced. For example, from Property 7 it follows immediately that to any row (column) of a square matrix may be added any *linear combination* of the *other* rows (columns) without changing the determinant of the matrix. By combining this result with Property 1, we deduce that *if any row (column) of a square matrix is a linear combination of the other rows (columns), then the determinant of that matrix is zero.*

If the row and column containing an element a_{ij} in a square matrix $\mathbf{A}$ are deleted, the determinant of the remaining square array is called the *minor* of a_{ij}, and is denoted here by M_{ij}. The *cofactor* of a_{ij}, denoted here by A_{ij}, is then defined by the relation

$$A_{ij} = (-1)^{i+j} M_{ij}. \tag{24}$$

Thus if the sum of the row and column indices of an element is *even*, the cofactor and the minor of that element are identical; otherwise they differ in sign.

It is a consequence of the definition of a determinant that *the cofactor of a_{ij} is the coefficient of a_{ij} in the expansion of* $|\mathbf{A}|$. This fact leads to the important *Laplace expansion formulas*:

$$|\mathbf{A}| = \sum_{k=1}^{n} a_{ik} A_{ik} \qquad \text{and} \qquad |\mathbf{A}| = \sum_{k=1}^{n} a_{kj} A_{kj}, \tag{25a,b}$$

for any relevant value of i or j. These formulas state that *the determinant of a square matrix is equal to the sum of the products of the elements of any single row or column of that matrix by their cofactors.*

If a_{ik} is replaced by a_{rk} in (25a), the result $\Sigma\, a_{rk} A_{ik}$ must accordingly be the determinant of a new matrix in which the elements of the ith row are replaced by the corresponding elements of the rth row, and hence must vanish if $r \neq i$ by virtue of Property 5. An analogous result follows if a_{kj} is replaced by a_{ks} in (25b), when $s \neq j$. Thus, in addition to (25), we have the relations

$$\sum_{k=1}^{n} a_{rk} A_{ik} = 0 \quad (r \neq i), \qquad \sum_{k=1}^{n} a_{ks} A_{kj} = 0 \quad (s \neq j). \tag{26a,b}$$

These results lead directly to *Cramer's rule* for solving a set of n linear equations in n unknown quantities, of the form

$$\sum_{k=1}^{n} a_{ik} x_k = c_i \qquad (i = 1, 2, \cdots, n), \tag{27}$$

in the case when the determinant of the matrix of coefficients is not zero,

$$|a_{ij}| \neq 0. \tag{28}$$

For if we *assume* the existence of a solution and multiply both sides of

(27) by A_{ir}, where r is any integer between 1 and n, and then sum the results with respect to i, there follows (after an interchange of order of summation)

$$\sum_{k=1}^{n}\left(\sum_{i=1}^{n} a_{ik}A_{ir}\right)x_k = \sum_{i=1}^{n} c_i A_{ir} \qquad (r = 1, 2, \cdots, n). \tag{29}$$

By virtue of (25b) and (26b), the inner sum on the left in (29) vanishes unless $k = r$ and is equal to $|\mathbf{A}|$ in that case. Hence (29) takes the form

$$|\mathbf{A}|\, x_r = \sum_{i=1}^{n} A_{ir}c_i \qquad (r = 1, 2, \cdots, n). \tag{30}$$

Thus, if the system (27) possesses a solution, then that solution also must satisfy the equation set (30). When $|\mathbf{A}| \neq 0$, the only possible solution of (27) accordingly is given by

$$x_r = \frac{1}{|\mathbf{A}|}\sum_{i=1}^{n} A_{ir}c_i \qquad (r = 1, 2, \cdots, n). \tag{31}$$

To verify that (31) does in fact satisfy (27), we introduce (31) into (27), first changing the dummy index i in (31) to another symbol, say j, to avoid ambiguity, after which the left-hand side of the ith equation of (27) becomes

$$\frac{1}{|\mathbf{A}|}\sum_{k=1}^{n} a_{ik}\sum_{j=1}^{n} A_{jk}c_j = \frac{1}{|\mathbf{A}|}\sum_{j=1}^{n}\left(\sum_{k=1}^{n} a_{ik}A_{jk}\right)c_j$$

after an interchange of order of summation. The use of (26a) and (25a) properly reduces this expression to

$$\frac{1}{|\mathbf{A}|}\left(\sum_{k=1}^{n} a_{ik}A_{ik}\right)c_i = c_i,$$

and hence establishes the validity of (31) when $|\mathbf{A}| \neq 0$.

The expansion on the right in (30) differs from the right-hand member of the expansion

$$|\mathbf{A}| = \sum_{i=1}^{n} A_{ir}a_{ir}$$

only in the fact that the column $\{c_i\}$ replaces the column $\{a_{ir}\}$ of the coefficients of x_r in $\mathbf{A}$. Thus we deduce *Cramer's rule*, which can be stated as follows:

When the determinant $|\mathbf{A}|$ of the matrix of coefficients in a set of n linear algebraic equations in n unknowns $x_1, \cdots, x_n$ is not zero, that set of equations has a unique solution. The value of any x_r can be expressed as the ratio of two determinants, the denominator being the determinant of the matrix of coefficients, and the numerator being the determinant of the matrix obtained by replacing the column of the coefficients of x_r in the coefficient matrix by the column of the right-hand members.

In the case when all right-hand members c_i are zero, the equations are said to be *homogeneous.* In this case, one solution is clearly the *trivial* one $x_1 = x_2 = \cdots = x_n = 0$. The preceding result then states that this is the only possible solution if $|\mathbf{A}| \neq 0$, so that *a set of n linear homogeneous equations in n unknowns cannot possess a nontrivial solution unless the determinant of the coefficient matrix vanishes.*

We postpone the treatment of the case when $|\mathbf{A}| = 0$, as well as the case when the number of equations differs from the number of unknowns, to Sections 1.9 and 1.11.

It can be shown that the *determinant of the product* of two square matrices of the same order is equal to the *product of the determinants:**

$$|\mathbf{A}\,\mathbf{B}| = |\mathbf{A}|\,|\mathbf{B}|. \tag{32}$$

A square matrix whose determinant vanishes is called a *singular* matrix; a *nonsingular* matrix is a square matrix whose determinant is *not* zero. From (32) it then follows that *the product of two nonsingular matrices is also nonsingular.*

It is true also that if a square matrix $\mathbf{M}$ is of one of the special forms

$$\mathbf{M} = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{0} \\ \hline \mathbf{C} & \mathbf{B} \end{array}\right] \quad \text{or} \quad \mathbf{M} = \left[\begin{array}{c|c} \mathbf{A} & \mathbf{C} \\ \hline \mathbf{0} & \mathbf{B} \end{array}\right], \tag{33a}$$

where $\mathbf{A}$ and $\mathbf{B}$ are square submatrices and where $\mathbf{0}$ is a submatrix whose elements are all *zeros,* then

$$|\mathbf{M}| = |\mathbf{A}|\,|\mathbf{B}|. \tag{33b}$$

Here the dimensions of $\mathbf{A}$ and $\mathbf{B}$ need not be equal and the zero submatrix $\mathbf{0}$ of $\mathbf{M}$ need not be square.

It follows from the definitions that the determinant of the negative of a square matrix is *not* necessarily the negative of the determinant, but that one has the relationship

$$|-\mathbf{A}| = (-1)^n\,|\mathbf{A}|,$$

where n is the order of the matrix $\mathbf{A}$.

1.5. Special matrices. In this section we define certain special matrices which are of importance, and investigate some of their properties.

That matrix which is obtained from $\mathbf{A} = [a_{ij}]$ by interchanging rows and columns is called the *transpose* of $\mathbf{A}$, and is here indicated by $\mathbf{A}^T$:

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

* While this result is in no sense profound, the existent proofs of (32) are either lengthy or indirect.

Thus the transpose of an $m \times n$ matrix is an $n \times m$ matrix. If the element in row r and column s of $\mathbf{A}$ is a_{rs}, where r may vary from 1 to m and s from 1 to n, then the element a'_{rs} in row r and column s of $\mathbf{A}^T$ is given by $a'_{rs} = a_{sr}$, where now r may vary from 1 to n and s from 1 to m.

If $\mathbf{A}$ is an $m \times l$ matrix and $\mathbf{B}$ is an $l \times n$ matrix, then both the products $\mathbf{A}\,\mathbf{B}$ and $\mathbf{B}^T\mathbf{A}^T$ exist, the former being an $m \times n$ matrix, and the latter an $n \times m$ matrix. We show next that the latter matrix is the transpose of the former. Since the element in row r and column s of the product $\mathbf{A}\,\mathbf{B} \equiv \mathbf{C}$ is given by

$$\sum_{k=1}^{l} a_{rk} b_{ks} \equiv c_{rs},$$

where r may vary from 1 to m and s from 1 to n, whereas the element c'_{rs} in row r and column s of the product $\mathbf{B}^T\,\mathbf{A}^T$ is given by

$$c'_{rs} \equiv \sum_{k=1}^{l} b'_{rk} a'_{ks} = \sum_{k=1}^{l} b_{kr} a_{sk} = c_{sr},$$

where now r may vary from 1 to n and s from 1 to m, it follows that $\mathbf{B}^T\,\mathbf{A}^T$ is indeed the transpose of $\mathbf{A}\,\mathbf{B}$.

Thus, we have shown that *the transpose of the product* $\mathbf{A}\,\mathbf{B}$ *is the product of the transposes in reverse order:*

$$(\mathbf{A}\,\mathbf{B})^T = \mathbf{B}^T\,\mathbf{A}^T. \tag{34}$$

This result will be of frequent usefulness.

When $\mathbf{A}$ is a *square* matrix, the matrix obtained from $\mathbf{A}$ by replacing each element by its cofactor and then interchanging rows and columns is called the *adjoint* of $\mathbf{A}$:

$$\operatorname{Adj} \mathbf{A} = \begin{bmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{bmatrix} = [A_{ji}].$$

The adjoint of a product is found to be equal to the product of the adjoints *in the reverse order.*

When the elements of a matrix $\mathbf{A}$ are complex, we denote by $\bar{\mathbf{A}}$ the matrix obtained by replacing each element of $\mathbf{A}$ by its complex conjugate, and call the matrix $\bar{\mathbf{A}}$ the *conjugate* of $\mathbf{A}$.

The *unit matrix* $\mathbf{I}$ of order n is the *square* $n \times n$ matrix having ones in its *principal diagonal* and zeros elsewhere,

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{bmatrix},$$

while a *zero matrix* **0** has zeros for *all* its elements. When it is necessary to indicate the dimensions of such matrices, the notation $\mathbf{I}_n$ may be used to denote the $n \times n$ unit matrix and the notation $\mathbf{0}_{m,n}$ to denote the $m \times n$ zero matrix. It is readily verified that for any matrix **A** there follow

$$\mathbf{A}\,\mathbf{I} = \mathbf{A}, \qquad \mathbf{I}\,\mathbf{A} = \mathbf{A} \tag{35a,b}$$

and

$$\mathbf{A}\,\mathbf{0} = \mathbf{0}, \qquad \mathbf{0}\,\mathbf{A} = \mathbf{0}. \tag{36a,b}$$

Here, if **A** is an $m \times n$ matrix, the symbol **I** necessarily stands for $\mathbf{I}_n$ in (35a) and for $\mathbf{I}_m$ in (35b), while **0** stands for $\mathbf{0}_{n,s}$ on the left and $\mathbf{0}_{m,s}$ on the right in (36a), and for $\mathbf{0}_{r,m}$ on the left and $\mathbf{0}_{r,n}$ on the right in (36b), where r and s are any positive integers.

The notation of the so-called *Kronecker delta*,

$$\delta_{pq} = \begin{cases} 0 & \textit{when} \quad p \neq q, \\ 1 & \textit{when} \quad p = q, \end{cases} \tag{37}$$

is frequently useful. With this notation, the general term of the unit matrix is merely δ_{ij}; that is, we can write

$$\mathbf{I} = [\delta_{ij}].$$

More generally, if all elements of a square matrix except those in the principal diagonal are zeros, the matrix is said to be a *diagonal matrix*. A diagonal matrix can thus be written in the form

$$\mathbf{D} = [d_i\delta_{ij}] = [d_j\delta_{ij}]$$

where the diagonal elements, for which $i = j$, are $d_1, d_2, \cdots, d_n$. The notation

$$\mathbf{D} = \operatorname{diag}(d_1, d_2, \cdots, d_n) \tag{38}$$

is also useful.

*Pre*multiplication of a matrix **A** by **D** multiplies the ith *row* of **A** by d_i; *post*multiplication multiplies the jth *column* by d_j. These results follow from the calculations

$$\mathbf{D}\,\mathbf{A} = [d_i\delta_{ik}][a_{kj}] = \left[\sum_{k=1}^{n} d_i\delta_{ik}a_{kj}\right] = [d_i a_{ij}] \tag{39}$$

and

$$\mathbf{A}\,\mathbf{D} = [a_{ik}][d_k\delta_{kj}] = \left[\sum_{k=1}^{n} a_{ik}d_k\delta_{kj}\right] = [a_{ij}d_j] = [d_j a_{ij}]. \tag{40}$$

A diagonal matrix whose diagonal elements are all *equal* is called a *scalar matrix*. Thus, a scalar matrix must be of the form

$$\mathbf{S} = k\mathbf{I} = [k\delta_{ij}].$$

1.6. The inverse matrix. With the notation of (37), the two equations (25a) and (26a) can be combined in the form

$$\sum_{k=1}^{n} a_{ik}A_{jk} = |\mathbf{A}|\,\delta_{ij}, \tag{41a}$$

while (25b) and (26b) lead to the relation

$$\sum_{k=1}^{n} a_{kj}A_{ki} = |\mathbf{A}|\,\delta_{ij}. \tag{41b}$$

If we write temporarily

$$m_{ij} = \frac{A_{ji}}{|\mathbf{A}|}, \tag{42}$$

under the assumption that $|\mathbf{A}| \neq 0$, these equations become

$$\sum_{k=1}^{n} a_{ik}m_{kj} = \delta_{ij}, \qquad \sum_{k=1}^{n} m_{ik}a_{kj} = \delta_{ij}. \tag{43a,b}$$

Hence, reviewing the definition (20) of the matrix product, we see that these equations imply the matrix equations

$$[a_{ik}][m_{kj}] = \mathbf{I}, \qquad [m_{ik}][a_{kj}] = \mathbf{I}. \tag{44}$$

That is, the matrix $\mathbf{M} = [m_{ij}]$ has the property that

$$\mathbf{A}\,\mathbf{M} = \mathbf{M}\,\mathbf{A} = \mathbf{I}, \tag{45}$$

where $\mathbf{I}$ is the unit matrix. It is natural to define a matrix satisfying (45) to be an *inverse* or *reciprocal* of $\mathbf{A}$, and to write $\mathbf{M} = \mathbf{A}^{-1}$.

If $\mathbf{A}$ is not *square*, there is no matrix $\mathbf{M}$ for which both $\mathbf{A}\,\mathbf{M}$ and $\mathbf{M}\,\mathbf{A}$ are unit matrices (see Problem 31), and hence $\mathbf{A}$ then cannot possess an inverse. Further, *a square matrix can have only one inverse*. To prove this statement, we suppose that $\mathbf{M}'$ is any matrix such that

$$\mathbf{A}\,\mathbf{M}' = \mathbf{I}, \tag{46}$$

where $\mathbf{A}$ is square. Then since accordingly $|\mathbf{A}|\,|\mathbf{M}'| = 1$, it follows that $\mathbf{A}$ must be *nonsingular*, and hence the associated matrix $\mathbf{M}$ exists. If we premultiply both sides of (46) by $\mathbf{M}$ and use (45) and (35), there follows

$$(\mathbf{M}\,\mathbf{A})\mathbf{M}' = \mathbf{M}\,\mathbf{I} \qquad or \qquad \mathbf{M}' = \mathbf{M},$$

as was to be shown.

We conclude that *if and only if the matrix* $\mathbf{A} = [a_{ij}]$ *is nonsingular*, it possesses an inverse $\mathbf{A}^{-1}$ such that

$$\mathbf{A}^{-1}\,\mathbf{A} = \mathbf{A}\,\mathbf{A}^{-1} = \mathbf{I}, \tag{47}$$

and that inverse is of the form

$$\mathbf{A}^{-1} = [m_{ij}] \quad where \quad m_{ij} = \frac{A_{ji}}{|\mathbf{A}|}. \tag{48}$$

Thus, to obtain the inverse of a nonsingular square matrix $[a_{ij}]$ we may first replace a_{ij} by its cofactor $A_{ij} = (-1)^{i+j}M_{ij}$, then interchange rows and columns and divide each element by the determinant $|a_{ij}|$. In the terminology of Section 1.5, the inverse of $\mathbf{A}$ is the adjoint of $\mathbf{A}$ divided by the determinant of $\mathbf{A}$:

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \operatorname{Adj} \mathbf{A}. \tag{49}$$

This equation also implies the useful relations

$$\mathbf{A}(\operatorname{Adj} \mathbf{A}) = |\mathbf{A}|\,\mathbf{I}, \qquad (\operatorname{Adj} \mathbf{A})\mathbf{A} = |\mathbf{A}|\,\mathbf{I}. \tag{50a,b}$$

It may be noticed that equations (50a,b) also follow directly from (41a,b) and hence are valid even when $|\mathbf{A}| = 0$.

To determine the inverse of a *product* of nonsingular square matrices, we write

$$\mathbf{A}\,\mathbf{B} = \mathbf{C}.$$

If we premultiply both sides of this equation successively by $\mathbf{A}^{-1}$ and $\mathbf{B}^{-1}$, there follows

$$\mathbf{I} = \mathbf{B}^{-1}\,\mathbf{A}^{-1}\,\mathbf{C}$$

and hence, by postmultiplying both sides of this equation by $\mathbf{C}^{-1}$ and replacing $\mathbf{C}$ by $\mathbf{A}\,\mathbf{B}$, we obtain the rule

$$(\mathbf{A}\,\mathbf{B})^{-1} = \mathbf{B}^{-1}\,\mathbf{A}^{-1}. \tag{51}$$

To illustrate the use of the inverse matrix, we again consider the problem of solving the set of linear equations (27) under the assumption (28). In matrix notation we have

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

and hence, after premultiplying both sides by $\mathbf{A}^{-1}$, there follows

$$\mathbf{x} = \mathbf{A}^{-1}\,\mathbf{c} \tag{52a}$$

or

$$\begin{Bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{Bmatrix} = \frac{1}{|\mathbf{A}|} \begin{bmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{bmatrix} \begin{Bmatrix} c_1 \\ c_2 \\ \cdot \\ \cdot \\ \cdot \\ c_n \end{Bmatrix} \tag{52b}$$

or

$$x_i = \frac{1}{|\mathbf{A}|}(A_{1i}c_1 + A_{2i}c_2 + \cdots + A_{ni}c_n) \qquad (i = 1, 2, \cdots, n), \tag{52c}$$

in accordance with the expanded form of Cramer's rule.

1.7. Rank of a matrix. Before proceeding to the analytical treatment of *general* sets of linear equations, to which Cramer's rule may not apply, it is desirable to introduce an additional definition and to establish certain preliminary basic results.

We define the *rank* of any matrix **A** as *the order of the largest square submatrix of* **A** (formed by deleting certain rows and/or columns of **A**) *whose determinant does not vanish.*

Suppose now that a certain matrix **A** is of rank r. We next show that if a set of r rows of **A** containing a nonsingular $r \times r$ submatrix **R** is selected, then any *other* row of **A** is a linear combination of those r rows.

To simplify the notation, we suppose that the square array **R** of order r in the upper left corner of the matrix **A** has a nonvanishing determinant, and consider the following submatrix of **A**:

$$\mathbf{M} = \begin{bmatrix} a_{11} & \cdots & a_{1r} & a_{1s} \\ \cdots & \cdots & \cdots & \cdots \\ a_{r1} & \cdots & a_{rr} & a_{rs} \\ a_{q1} & \cdots & a_{qr} & a_{qs} \end{bmatrix} = \left[\begin{array}{ccc:c} & & & a_{1s} \\ & \mathbf{R} & & \vdots \\ & & & a_{rs} \\ \hdashline a_{q1} & \cdots & a_{qr} & a_{qs} \end{array}\right] \tag{53}$$

where $s > r$ and $q > r$. Then, since the matrix **A** is of rank r, the determinant of this square submatrix must vanish for all such s and q.

Now it is possible to determine constants $\lambda_1, \lambda_2, \cdots, \lambda_r$ such that the equations

$$\left.\begin{aligned} \lambda_1 a_{11} + \lambda_2 a_{21} + \cdots + \lambda_r a_{r1} &= a_{q1}, \\ \lambda_1 a_{12} + \lambda_2 a_{22} + \cdots + \lambda_r a_{r2} &= a_{q2}, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ \lambda_1 a_{1r} + \lambda_2 a_{2r} + \cdots + \lambda_r a_{rr} &= a_{qr} \end{aligned}\right\} \tag{54}$$

are satisfied, since the coefficient determinant $|\mathbf{R}^T| = |\mathbf{R}|$ assuredly does not vanish and Cramer's rule applies. Hence, with these constants of combination, we can determine a row of elements which is a linear combination of the first r rows of **M**, and which will have its first r elements identical with the first r elements of the last row. Let the last element of that combination be denoted by a'_{qs}. In evaluating the *determinant* of **M**, we may subtract this linear combination of the first r rows from the last row without changing the value of the determinant, to obtain the result

$$|\mathbf{M}| = \begin{vmatrix} a_{11} & \cdots & a_{1r} & a_{1s} \\ \cdots & \cdots & \cdots & \cdots \\ a_{r1} & \cdots & a_{rr} & a_{rs} \\ 0 & \cdots & 0 & a_{qs} - a'_{qs} \end{vmatrix}. \tag{55}$$

But since $|\mathbf{M}|$ is equal, by the Laplace expansion, to the product of $a_{qs} - a'_{qs}$ and the determinant $|\mathbf{R}|$ which does *not* vanish, by hypothesis, and since $|\mathbf{M}| = 0$, it follows that $a'_{qs} = a_{qs}$. Hence we see that *the last row of* $\mathbf{M}$ *is a linear combination of the first* r *rows.* Since this is true for *any* q and s greater than r, the result can be stated as follows:

If a matrix is of rank r, *and a set of* r *rows containing a nonsingular submatrix of order* r *is selected, then any other row in the matrix is a linear combination of these* r *rows.*

The same statement is easily seen to be true, by a similar argument, if the word "row" is replaced by "column" throughout.

As a special case, we deduce that if a square matrix $\mathbf{A}$ is singular, then at least one of the rows of $\mathbf{A}$ must be a linear combination of the others, and the same statement applies to the columns. Since the converse has already been established in Section 1.7, it follows that *a square matrix is singular if and only if one of its rows is a linear combination of the others. The same statement applies to the columns.*

It is obvious that the process of interchanging rows and columns does not affect the rank of a matrix, so that *the two matrices* $\mathbf{A}$ *and* $\mathbf{A}^T$ *have the same rank.* The following section treats certain other operations on matrices which also leave the rank invariant.

1.8. Elementary operations. Associated with a set of m linear equations in n unknowns,

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= c_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= c_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= c_m \end{aligned}\right\}, \tag{56}$$

we consider two matrices: the $m \times n$ matrix of coefficients $[a_{ij}]$, and the $m \times (n + 1)$ matrix formed by joining to the columns of $[a_{ij}]$ the column of constants $\{c_i\}$. We refer to the former matrix as the *coefficient matrix* and to the second as the *augmented matrix.*

As was shown in Section 1.2, we may use the Gauss-Jordan reduction (renumbering certain equations or variables, if necessary) to replace (56) by an equivalent set of equations of the form

$$\left.\begin{aligned} x_1 \quad - \alpha_{11}x_{r+1} - \cdots - \alpha_{1,n-r}x_n &= \gamma_1, \\ x_2 \quad - \alpha_{21}x_{r+1} - \cdots - \alpha_{2,n-r}x_n &= \gamma_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ x_r - \alpha_{r1}x_{r+1} - \cdots - \alpha_{r,n-r}x_n &= \gamma_r, \\ 0 &= \gamma_{r+1}, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ 0 &= \gamma_m \end{aligned}\right\}, \tag{57}$$

by a process which involves, in addition to possible renumbering, only the multiplication of equal quantities by equal nonzero quantities and the addition of equals to equals. Accordingly, the augmented matrix of (56) is transformed to the augmented matrix of (57), which is of the form

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & -\alpha_{11} & -\alpha_{12} & \cdots & -\alpha_{1,n-r} & \gamma_1 \\ 0 & 1 & \cdots & 0 & -\alpha_{21} & -\alpha_{22} & \cdots & -\alpha_{2,n-r} & \gamma_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & -\alpha_{r1} & -\alpha_{r2} & \cdots & -\alpha_{r,n-r} & \gamma_r \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \gamma_{r+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \gamma_m \end{bmatrix}, \tag{58}$$

and, at the same time, the coefficient matrix of (56) transforms into the result of deleting the last column of the matrix (58).

The steps in the reduction involve only the following so-called *elementary row and column operations:*

1. The interchange of two rows (or of two columns).
2. The multiplication of the elements of a row (or column) by a number other than zero.
3. The addition, to the elements of a row (column), of k times the corresponding elements of another row (column).

It is clear that the transformed *coefficient* matrix in the above case is of rank r, whereas if one or more of the numbers $\gamma_{r+1}, \gamma_{r+2}, \cdots, \gamma_m$ is not zero, the rank of the transformed *augmented* matrix is $r + 1$. If $\gamma_{r+1} = \gamma_{r+2} = \cdots = \gamma_m = 0$, *both* transformed matrices are of rank r.

It is next shown that *the ranks of the two matrices associated with* (56) *are the same as the ranks of the corresponding transformed matrices,* that is, that *the rank of a matrix is not changed by the elementary operations.*

We need show only that a nonvanishing determinant of *largest* order r is not reduced to zero, and that no nonvanishing determinant of higher order is *introduced,* by any such operation.

Operation 1 is equivalent to renumbering rows or columns, and obviously cannot affect over-all vanishing or nonvanishing of determinants.

Similarly, *operation 2* can only multiply certain determinants by a nonzero constant.

According to Property 7 of determinants (page 10), *operation 3* does not change the value of any determinant which involves either *both* or *neither* of the two rows (or columns) concerned. To simplify the notation in the remaining case, we again suppose that one nonsingular $r \times r$ submatrix **R**

is in the upper left corner of **A**, and that k times the qth row of **A** is to be added to the ith row, with $i \leqq r < q$. If we then consider the effect of this row operation on the submatrix

$$\mathbf{S} = \begin{bmatrix} a_{11} & \cdots & a_{1r} \\ \cdots & \cdots & \cdots \\ a_{i1} & \cdots & a_{ir} \\ \cdots & \cdots & \cdots \\ a_{r1} & \cdots & a_{rr} \\ \cdots & \cdots & \cdots \\ a_{q1} & \cdots & a_{qr} \end{bmatrix} \equiv \begin{bmatrix} & \mathbf{R} & \\ & & \\ & & \\ \cdots & \cdots & \cdots \\ a_{q1} & \cdots & a_{qr} \end{bmatrix}$$

we may use the last result of Section 1.7 to deduce that the rank of **A** is not *reduced* by *row* operation 3.

For either the last row of **S** is a linear combination of rows of **R** *excluding* the ith, in which case $|\mathbf{R}|$ is unchanged, or the $r \times r$ subarray of **S** which is obtained by deleting the ith row (and which is *unaffected* by the operation) is nonsingular. Thus there is at least *one* nonsingular $r \times r$ submatrix in **S** after the operation is effected.

Conversely, this operation cannot *increase* the rank of a matrix, since the reversed operation (which would be of the same type) then would *reduce* the rank of the *new* matrix.

The same argument applies to operation 3 effected on *columns*. Hence we conclude that *the elementary operations, applied to rows or to columns, do not change the rank of a matrix.*

1.9. Solvability of sets of linear equations. If we notice that no *column* operations are involved in the Gauss-Jordan reduction, except perhaps a renumbering of certain columns of the *coefficient* matrix, we conclude both that the augmented matrices of (56) and (57) are of equal rank, and also that the same is true of the coefficient matrices.

If and only if one or more of the numbers $\gamma_{r+1}, \gamma_{r+2}, \cdots, \gamma_m$ in (57) is not zero, the given set of equations possesses no solution. But if and only if this is so, the rank of the augmented matrix is greater than the rank of the coefficient matrix. Thus we deduce the following basic result:

A set of linear equations possesses a solution if and only if the rank of the augmented matrix is equal to the rank of the coefficient matrix.

If the two ranks *are* both equal to r, and if we select a set of r equations whose coefficient matrix contains a nonsingular submatrix of order r, then we may disregard all other equations, since they are implied by the r basic equations (that is, their coefficients are linear combinations of the coefficients in the r basic equations). The $n - r$ unknowns whose coefficients are *not* involved in the nonsingular $r \times r$ submatrix can be assigned arbitrary

values, after which the remaining r unknowns can be determined in terms of them (by Cramer's rule or otherwise).*

In particular, *if $r = n$ the unknowns are determined uniquely.* Otherwise, if $n - r = d > 0$, the most general solution involves d independent arbitrary parameters.

In the *homogeneous* case, when the right-hand members of (56) are all zeros, the coefficient matrix and the augmented matrix are automatically of equal rank, and a solution *always* exists. But this fact is obvious, since such a system is always satisfied by the *trivial solution* $x_1 = x_2 = \cdots = x_n = 0$. If the rank r of the coefficient matrix is equal to the number n of unknowns, then this is the *only* solution, in accordance with the special results of Section 1.4. However, if $r < n$ (in particular, if the number of equations is less than the number of unknowns) infinitely many solutions exist, the number of independent arbitrary parameters involved being given by the difference $n - r$.

We notice that, in consequence of the *linearity* of the relevant equations, the general solution of a *nonhomogeneous* set of equations is the sum of any one *particular* solution of that set and the most general solution of the associated homogeneous set.

A case of particular interest is that of a set of n homogeneous equations in n unknowns, in which the coefficient matrix is of rank $n - 1$; that is, a set of the form

$$\sum_{k=1}^{n} a_{ik}x_k = 0 \qquad (i = 1, 2, \cdots, n) \tag{59}$$

where

$$|a_{ij}| = 0 \tag{60}$$

but where the determinant of at least one square submatrix of $\mathbf{A}$ of order $n - 1$ does not vanish. In consequence of equations (41a) and (60), these equations are satisfied by the expressions

$$x_i = CA_{si} \qquad (i = 1, 2, \cdots, n) \tag{61}$$

where C is an arbitrary constant and s may take on any value from 1 to n. Since here $d = n - r = 1$, and since (61) contains one arbitrary parameter, (61) must represent the most general solution of (59) unless, for the particular value of s chosen, all cofactors A_{si} happen to vanish. (This exception cannot exist for *all* values of s if the rank of $\mathbf{A}$ is $n - 1$.) With this reservation, the result obtained is equivalent to the statement that, in the case under consideration, *the unknowns are proportional to the cofactors of their coefficients in any row of the matrix* $[a_{ij}]$.

* In actual numerical cases, a procedure such as that of the Gauss or Gauss-Jordan reduction avoids the necessity of perhaps evaluating a large number of determinants. However, the results obtained here are of great importance in more general considerations.

1.10. Linear vector space. The preceding results have interesting and instructive interpretations in terms of so-called "vector space," which is briefly discussed in this section.

It is conventional to speak of a *one-column* matrix as a *column vector* or, more briefly, as a *vector*. (Such an array is often called a *numerical* vector, to distinguish it from a *geometrical* or *physical* vector quantity—such as a force or an acceleration—which it may *represent*.) In accordance with the notation introduced in Section 1.3, a lower-case boldface letter is used in this text to specifically denote a vector, so that we write

$$\mathbf{x} = \{x_1, x_2, \cdots, x_n\}.$$

A *one-row* matrix $[x_1, x_2, \cdots, x_n]$, which is the *transpose* of the vector $\mathbf{x}$, will be termed a *row vector* and will be denoted by $\mathbf{x}^T$ when a special symbolism is desirable.

In *two-dimensional* space, the *elements* of the vector $\{x_1, x_2\}$ can be interpreted as the *components* of $\mathbf{x}$ in the directions of rectangular coordinate (x_1 and x_2) axes. The square of the *length* of this vector is then given by $l^2 = x_1^2 + x_2^2 = \mathbf{x}^T\mathbf{x}$.* Also, if $\mathbf{u}$ and $\mathbf{v}$ are two vectors in two-dimensional space, the *scalar product* of $\mathbf{u}$ and $\mathbf{v}$ is defined to be $u_1v_1 + u_2v_2 = \mathbf{u}^T\mathbf{v} = \mathbf{v}^T\mathbf{u}$. It is seen that the scalar product $\mathbf{u}^T\mathbf{v}$ here is the equivalent, in matrix notation, of the "dot product" $\mathbf{u} \cdot \mathbf{v}$ in vector analysis. We recall that the vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal (perpendicular) if and only if this scalar product vanishes. The vectors $\mathbf{i}_1 = \{1, 0\}$ and $\mathbf{i}_2 = \{0, 1\}$ are the orthogonal *unit vectors* ordinarily denoted by $\mathbf{i}$ and $\mathbf{j}$, respectively, in vector analysis.

The above terminology is extended by analogy to the general case of n *dimensions*. When $n > 3$, it is impossible to visualize the vectors geometrically. However, we use the language associated with space of two or three dimensions, and say that an n-dimensional rectangular coordinate system comprises n mutually orthogonal axes, that a point has n corresponding coordinates, and that a vector has n components along these axes.

The *scalar product* of two vectors $\mathbf{u}$ and $\mathbf{v}$ is defined to be

$$\mathbf{u}^T\mathbf{v} = \mathbf{v}^T\mathbf{u} = u_1v_1 + u_2v_2 + \cdots + u_nv_n \tag{62}$$

and the *square of the length* of a vector $\mathbf{u}$ is defined to be

$$l^2(\mathbf{u}) = \mathbf{u}^T\mathbf{u} = u_1^2 + u_2^2 + \cdots + u_n^2. \tag{63}$$

It is sometimes convenient to denote the scalar product by the abbreviation $(\mathbf{u}, \mathbf{v})$, so that

$$(\mathbf{u}, \mathbf{v}) = \mathbf{u}^T\mathbf{v} = \mathbf{v}^T\mathbf{u} = (\mathbf{v}, \mathbf{u}). \tag{64}$$

* A product of the form $\mathbf{x}^T\mathbf{x}$, which is truly a one-element matrix, is conventionally treated as a scalar.

The abbreviation

$$\mathbf{u}^2 \equiv (\mathbf{u}, \mathbf{u})$$

is frequently used in the special case when $\mathbf{v} = \mathbf{u}$.

Two vectors $\mathbf{u}$ and $\mathbf{v}$ are said to be *orthogonal* if their scalar product vanishes, $(\mathbf{u}, \mathbf{v}) = 0$. A *zero vector* is thus orthogonal to *all* vectors of the same dimension. A vector is said to be a *unit vector* if its length is unity, so that $(\mathbf{u}, \mathbf{u}) = 1$.

When the components of $\mathbf{u}$ and $\mathbf{v}$ are *real*, it can be shown that

$$|(\mathbf{u}, \mathbf{v})| \leqq (\mathbf{u}, \mathbf{u})^{1/2}(\mathbf{v}, \mathbf{v})^{1/2} \tag{65}$$

(see Problem 43), that is, that the magnitude of the *scalar product* of $\mathbf{u}$ and $\mathbf{v}$ is not larger than the product of the *lengths* of $\mathbf{u}$ and $\mathbf{v}$. This important result is known as the *Schwarz inequality*.

Hence the natural definition

$$\cos\theta = \frac{(\mathbf{u}, \mathbf{v})}{(\mathbf{u}, \mathbf{u})^{1/2}(\mathbf{v}, \mathbf{v})^{1/2}} \qquad (-\pi < \theta \leqq \pi) \tag{66}$$

makes $|\cos\theta| \leqq 1$ and so yields a real interpretation for the "*angle* θ between the real vectors $\mathbf{u}$ and $\mathbf{v}$" in the general n-dimensional case.

When the components of the vectors are *complex* numbers, the preceding properties of length, angle, and orthogonality usually are of limited significance and it is desirable to define more appropriate properties, which reduce to the preceding ones in the real case.

For this purpose, the *Hermitian scalar product* of a vector $\mathbf{u}$ into a vector $\mathbf{v}$ may be defined as

$$(\bar{\mathbf{u}}, \mathbf{v}) = \bar{\mathbf{u}}^T\mathbf{v} = \bar{u}_1v_1 + \bar{u}_2v_2 + \cdots + \bar{u}_nv_n = (\mathbf{v}, \bar{\mathbf{u}}), \tag{67}$$

and is complex and *not*, in general, equal to its conjugate $(\mathbf{u}, \bar{\mathbf{v}})$. The square of the *Hermitian length* (or *absolute length*) of a vector $\mathbf{u}$ with complex components is then defined to be the nonnegative *real* quantity

$$l_H^2(\mathbf{u}) = (\bar{\mathbf{u}}, \mathbf{u}) = \bar{\mathbf{u}}^T\mathbf{u} = \bar{u}_1u_1 + \bar{u}_2u_2 + \cdots + \bar{u}_nu_n, \tag{68}$$

and the *Hermitian angle* θ_H between $\mathbf{u}$ and $\mathbf{v}$ is a real quantity defined by the relation

$$\cos\theta_H = \frac{(\bar{\mathbf{u}}, \mathbf{v}) + (\bar{\mathbf{v}}, \mathbf{u})}{2(\bar{\mathbf{u}}, \mathbf{u})^{1/2}(\bar{\mathbf{v}}, \mathbf{v})^{1/2}} \qquad (-\pi < \theta_H \leqq \pi) \tag{69}$$

(see Problem 44). Two complex vectors $\mathbf{u}$ and $\mathbf{v}$ are then said to be *orthogonal in the Hermitian sense* when $(\bar{\mathbf{u}}, \mathbf{v}) = 0$, and hence also $(\bar{\mathbf{v}}, \mathbf{u}) = 0$.

When the elements involved are real, they are equal to their complex conjugates, and it is seen that (67), (68), and (69) reduce to (62), (63), and

(66). However, it should be noted, for example, that if $\mathbf{v} = \{1, i\}$, where $i^2 = -1$, then there follows $l(\mathbf{v}) = 0$ but $l_H(\mathbf{v}) = \sqrt{2}$.

A set of m vectors $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m$ is said to be *linearly independent* if *no* set of constants $c_1, c_2, \ldots, c_m$, at least one of which is not zero, exists such that

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_m\mathbf{u}_m = \mathbf{0}. \tag{70}$$

In two-dimensional space, the existence of c_1 and c_2 (not both zero) such that

$$c_1\mathbf{u}_1 + c_2\mathbf{u}_2 = \mathbf{0}$$

would imply that one of the two-dimensional vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ is a scalar multiple of the other. Hence any two vectors which are not multiples of the same vector (parallel to a line) are linearly independent in two-dimensional space. Further, geometrical considerations indicate that any *three* vectors which are not *parallel to a plane* are linearly independent in three-dimensional space.

To obtain an analytical criterion for linear dependence of a set of m vectors with *real* components, we suppose that c's *do* exist, at least one of which is not zero, such that (70) is satisfied. Then, by successively forming the scalar products of $\mathbf{u}_1, \cdots, \mathbf{u}_m$ into both sides of (70), we find that the constants c_i must also satisfy the equations

$$\begin{aligned}
c_1\mathbf{u}_1^2 + c_2(\mathbf{u}_1, \mathbf{u}_2) + \cdots + c_m(\mathbf{u}_1, \mathbf{u}_m) &= 0,\\
c_1(\mathbf{u}_2, \mathbf{u}_1) + c_2\mathbf{u}_2^2 + \cdots + c_m(\mathbf{u}_2, \mathbf{u}_m) &= 0,\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots &\\
c_1(\mathbf{u}_m, \mathbf{u}_1) + c_2(\mathbf{u}_m, \mathbf{u}_2) + \cdots + c_m\mathbf{u}_m^2 &= 0.
\end{aligned}$$

These conditions clearly require merely that the left-hand member of (70) be simultaneously orthogonal to $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m$. But, according to Cramer's rule, this set of m equations in the m constants c_i cannot possess a nontrivial solution unless the determinant of the matrix of coefficients vanishes:

$$G \equiv \begin{vmatrix} \mathbf{u}_1^2 & (\mathbf{u}_1, \mathbf{u}_2) & \cdots & (\mathbf{u}_1, \mathbf{u}_m) \\ (\mathbf{u}_2, \mathbf{u}_1) & \mathbf{u}_2^2 & \cdots & (\mathbf{u}_2, \mathbf{u}_m) \\ \cdots & \cdots & \cdots & \cdots \\ (\mathbf{u}_m, \mathbf{u}_1) & (\mathbf{u}_m, \mathbf{u}_2) & \cdots & \mathbf{u}_m^2 \end{vmatrix} = 0. \tag{71}$$

This determinant is called the Gram determinant or *Gramian* of $\mathbf{u}_1, \cdots, \mathbf{u}_m$. Thus, if the vectors are linearly dependent the Gramian must vanish. The converse can also be shown to be true (see Problem 42). Hence it follows that *a set of real vectors is linearly dependent if and only if its Gramian vanishes.*

For vectors with *complex* components, this theorem is still true if all scalar products in the definition of the Gramian are replaced by *Hermitian* scalar products.

If $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m$ are n-dimensional vectors, then the set of all vectors $\mathbf{v}$ which can be expressed in the form

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_m\mathbf{u}_m \tag{72}$$

is called the *vector space* generated or *spanned* by the vector set $\mathbf{u}_1, \cdots, \mathbf{u}_m$. If r and only r of the $\mathbf{u}$'s are linearly independent, the space so generated is said to be of *dimension* r. Thus the *dimension* of the generated *space* is also the *rank* of a *matrix* which has the elements of the successive generating vectors as its successive rows or columns.

When $m > r$, we see that $m - r$ of the $\mathbf{u}$'s can be expressed as linear combinations of the r *independent* $\mathbf{u}$'s, so that any vector $\mathbf{v}$ in the vector space generated by the set of m vectors also can be generated by a subset of r independent ones. Such a set of r *linearly independent* vectors is said to form a *basis* for the r-dimensional vector space which it spans.

The space of *all* n-dimensional vectors, that is, the set of all vectors having n elements, will be of principal importance in this work and will be referred to, for brevity, as *n-space*. It is clear that n-space is spanned by *any* set of n linearly independent n-dimensional vectors $\mathbf{u}_1, \cdots, \mathbf{u}_n$. For, if $\mathbf{v}$ is any n-dimensional vector, then in the n equations which equate the n components of the two members of the relation

$$\mathbf{v} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_n\mathbf{u}_n, \tag{73}$$

the matrix of coefficients of the c's has the property that no column is a linear combination of the others, and hence the determinant of the coefficient matrix cannot vanish. Thus c's can be determined so that (73) is true.

To determine the constants in (73) by an alternative method, we may form the scalar product of each $\mathbf{u}$ into the equal members of (73). The resultant set of n scalar equations always can be solved for the c's, when the $\mathbf{u}$'s are linearly independent, since the determinant of the relevant coefficient matrix is the Gramian of the $\mathbf{u}$'s, and hence does not vanish. In particular, it follows that if the n-dimensional vector $\mathbf{v}$ is *orthogonal* to each of the n linearly independent $\mathbf{u}$'s, then $\mathbf{v}$ must be the *zero* vector.*

A set of r linearly independent n-dimensional vectors is said to span an r-dimensional *subspace* of the space of *all* n-dimensional vectors, when $r < n$, and is said to be of *defect* (or *nullity*) $d = n - r$ in n-space.

Clearly, any set of n nonzero *mutually orthogonal* n-dimensional vectors is a basis in n-space, since its Gramian is the determinant of a diagonal

* In the *complex* case, the scalar products and the orthogonality are here to be defined in the *Hermitian* sense.

matrix with nonzero diagonal elements, and hence cannot vanish. An especially convenient basis comprises the particular orthogonal *unit* vectors

$$\mathbf{i}_1 = \{1, 0, 0, \cdots, 0\}, \qquad \mathbf{i}_2 = \{0, 1, 0, \cdots, 0\}, \qquad \cdots,$$
$$\mathbf{i}_n = \{0, 0, 0, \cdots, 1\}, \tag{74}$$

and is sometimes called the *standard basis* of n-space.

1.11. Linear equations and vector space. We now indicate briefly the interpretation of the basic results of Section 1.9, relevant to a set of m linear algebraic equations in n unknowns, with reference to the terminology of Section 1.10.

The set of equations

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= c_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= c_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= c_m \end{aligned}\right\}, \tag{75}$$

which corresponds to the vector equation

$$\mathbf{A}\,\mathbf{x} = \mathbf{c}, \tag{76}$$

can also be written in the form

$$(\boldsymbol{\alpha}_i, \mathbf{x}) = c_i \qquad (i = 1, 2, \cdots, m). \tag{77}$$

Here $\mathbf{x}$ is the n-dimensional vector $\{x_1, \cdots, x_n\}$ and

$$\boldsymbol{\alpha}_i = \{a_{i1}, a_{i2}, \cdots, a_{in}\} \qquad (i = 1, 2, \cdots, m) \tag{78}$$

is the n-dimensional vector whose elements comprise the ith *row* of the coefficient matrix $\mathbf{A}$ of (75) so that, schematically,

$$\mathbf{A} = [a_{ij}] = \begin{bmatrix} -\boldsymbol{\alpha}_1{}^T \rightarrow \\ \cdots\cdots\cdots \\ -\boldsymbol{\alpha}_m{}^T \rightarrow \end{bmatrix}.$$

Thus the equations in (75) can be interpreted as prescribing the scalar product of the unknown vector $\mathbf{x}$ and the transpose of each of the m *row vectors* of $\mathbf{A}$.

In particular, when the c's are all zeros the equations (75) become *homogeneous* and require that $\mathbf{x}$ be *orthogonal* to each of the m vectors $\boldsymbol{\alpha}_i$. When the rank of $\mathbf{A}$ is equal to n, the results of Section 1.9 state that $\mathbf{x} = \mathbf{0}$ is the only solution of the associated matrix equation

$$\mathbf{A}\,\mathbf{x} = \mathbf{0}. \tag{79}$$

This result is in accordance with the fact that in this case the m vectors $\boldsymbol{\alpha}_i$ *span* all of n-space. However, when the rank of $\mathbf{A}$ is r, where $r < n$, Section 1.9 states that there is an $(n - r)$-fold infinity of vectors, each of which satisfies (79) and hence is orthogonal to each of the m vectors $\boldsymbol{\alpha}_i$. This means that when the vectors $\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_m$ span only a *subspace* of dimension $r < n$, it is possible to find a set of $d = n - r$ linearly independent nonzero vectors, say $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_d$, each of which is orthogonal to all the $\boldsymbol{\alpha}$'s. Any vector $\mathbf{x}$ which is a linear combination of these vectors,

$$\mathbf{x} = C_1\mathbf{u}_1 + C_2\mathbf{u}_2 + \cdots + C_d\mathbf{u}_d, \tag{80}$$

will satisfy the equation (79), and its components will satisfy (75) with $c_i = 0$.

Here the *solution space* (80), which is generated by the basis $\mathbf{u}_1, \cdots, \mathbf{u}_{n-r}$, and the complementary vector space, known as the *row space* of $\mathbf{A}$, which is generated by any r linearly independent vectors in the set $\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_m$, together comprise all of n-space. For since the Gramian of the combination of the *two* bases just considered is the product of the separate nonvanishing Gramians [see equation (33)] it follows that their combination comprises an n-member basis for the space of *all* n-dimensional vectors. Further, it is easily seen that no vector other than the zero vector can be in *both* subspaces (see Problem 48). Hence it may be deduced that *a nonzero vector* $\mathbf{x}$ *satisfies* (79) *if and only if it is not in the row space of* $\mathbf{A}$.

In order to display the general solution of (75) or, equivalently, (57) in the form (80) when the right-hand members of (75) vanish, we may write $x_{r+1} = C_1, x_{r+2} = C_2, \cdots, x_n = C_{n-r}$, where the C's are arbitrary. The solution can then be written in the vector form

$$\begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_r \\ x_{r+1} \\ x_{r+2} \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} = C_1 \begin{pmatrix} \alpha_{11} \\ \alpha_{21} \\ \cdot \\ \cdot \\ \alpha_{r1} \\ 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix} + C_2 \begin{pmatrix} \alpha_{12} \\ \alpha_{22} \\ \cdot \\ \cdot \\ \alpha_{r2} \\ 0 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix} + \cdots + C_{n-r} \begin{pmatrix} \alpha_{1,n-r} \\ \alpha_{2,n-r} \\ \cdot \\ \cdot \\ \alpha_{r,n-r} \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix}. \tag{80'}$$

It is clear from the form of the $n - r$ solution vectors that these vectors are indeed linearly independent.

In the more general case when equations (75) are *nonhomogeneous,* so that the scalar products of $\mathbf{x}$ and the vectors $\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_m$ are each to take on prescribed values, the most general vector $\mathbf{x}$ having this property is expressible as the sum of any *particular* vector having this property (if such exist) and an arbitrary linear combination of all vectors which are *orthogonal* to all the $\boldsymbol{\alpha}$'s.

It is useful to notice also that the m equations (75) can be combined into the single vector equation

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \cdot \\ \cdot \\ \cdot \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \cdot \\ \cdot \\ \cdot \\ a_{m2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \cdot \\ \cdot \\ \cdot \\ a_{mn} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \cdot \\ \cdot \\ \cdot \\ c_m \end{pmatrix} \tag{81}$$

or

$$x_1\boldsymbol{\beta}_1 + x_2\boldsymbol{\beta}_2 + \cdots + x_n\boldsymbol{\beta}_n = \mathbf{c}, \tag{82}$$

where $\mathbf{c}$ is the m-dimensional vector $\{c_1, \cdots, c_m\}$ and

$$\boldsymbol{\beta}_j = \{a_{1j}, a_{2j}, \cdots, a_{mj}\} \qquad (j = 1, 2, \cdots, n) \tag{83}$$

is the m-dimensional vector whose elements comprise the jth *column* of the matrix $\mathbf{A}$,

$$\mathbf{A} = [a_{ij}] = \begin{bmatrix} | & & | \\ \boldsymbol{\beta}_1 & \cdots & \boldsymbol{\beta}_n \\ \downarrow & & \downarrow \end{bmatrix}.$$

Thus, with this interpretation, we see that the equations (75) are compatible if and only if $\mathbf{c}$ is representable as a linear combination of the *column vectors* comprising $\mathbf{A}$, and hence if and only if $\mathbf{c}$ is in the *column space* of $\mathbf{A}$. The components of $\mathbf{x}$, when they exist, then are to be the constants of combination in such a representation.

In some considerations, it is desirable to associate with the relation $\mathbf{A}\,\mathbf{x} = \mathbf{c}$, corresponding to the m equations (75) in n unknowns $x_1, \cdots, x_n$, the *transposed homogeneous set*

$$\left.\begin{aligned} a_{11}x_1' + a_{21}x_2' + \cdots + a_{m1}x_m' &= 0, \\ a_{12}x_1' + a_{22}x_2' + \cdots + a_{m2}x_m' &= 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{1n}x_1' + a_{2n}x_2' + \cdots + a_{mn}x_m' &= 0 \end{aligned}\right\} \tag{84}$$

of n *homogeneous* equations in m *new* unknowns $x_1', \cdots, x_m'$, corresponding to the relation $\mathbf{A}^T\,\mathbf{x}' = \mathbf{0}$. These equations also can be written in the forms

$$(\boldsymbol{\beta}_j, \mathbf{x}') = 0 \qquad (j = 1, 2, \cdots, n) \tag{85}$$

and

$$\sum_{i=1}^{m} x_i' \boldsymbol{\alpha}_i = \mathbf{0}, \tag{86}$$

with the notation of (78) and (83), in accordance with the fact that the row and column vectors of $\mathbf{A}$ are the column and row vectors, respectively, of $\mathbf{A}^T$.

From the relations (82) and (85), we may deduce the following useful result:

The nonhomogeneous equation $\mathbf{A}\,\mathbf{x} = \mathbf{c}$ *possesses a vector solution* $\mathbf{x}$ *if and only if the vector* $\mathbf{c}$ *is orthogonal to all vector solutions of the associated homogeneous equation* $\mathbf{A}^T\,\mathbf{x}' = \mathbf{0}$.

In order to establish this result, we notice first that if $\mathbf{A}\,\mathbf{x} = \mathbf{c}$ has a solution, then $\mathbf{c}$ is in the *column* space of $\mathbf{A}$ and hence in the *row* space of $\mathbf{A}^T$. Hence $\mathbf{c}$ then is a linear combination of the vectors $\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_n$, each of which is orthogonal to every $\mathbf{x}'$ which satisfies $\mathbf{A}^T\,\mathbf{x}' = \mathbf{0}$, according to (85). Thus $\mathbf{c}$ also is orthogonal to each $\mathbf{x}'$. Conversely, if $\mathbf{c}$ is orthogonal to each solution of $\mathbf{A}^T\,\mathbf{x}' = \mathbf{0}$, then $\mathbf{c}$ is in the complement in n-space of the associated solution space. Thus $\mathbf{c}$ is in the *row* space of $\mathbf{A}^T$ and hence in the *column* space of $\mathbf{A}$, so that $x_1, \cdots, x_n$ exist such that (82) is satisfied and accordingly $\mathbf{A}\,\mathbf{x} = \mathbf{c}$ has a solution.

1.12. Characteristic-value problems. A frequently encountered problem is that of determining those values of a constant λ for which *nontrivial* solutions exist to a homogeneous set of equations of the form

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= \lambda x_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= \lambda x_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= \lambda x_n \end{aligned}\right\}. \tag{87}$$

Such a problem is known as a characteristic-value problem; values of λ for which nontrivial solutions exist are called *characteristic values* (also *eigenvalues* or *latent roots*) of the problem or *of the matrix* $\mathbf{A}$, and corresponding vector solutions are known as *characteristic vectors* (also *eigenvectors*) of the problem or *of the matrix* $\mathbf{A}$. A column made up of the elements of a characteristic vector is often called a *modal column*.

In many practical considerations in which such problems arise, the matrix $\mathbf{A}$ is *real and symmetric*, so that two elements which are symmetrically placed with respect to the principal diagonal are equal:

$$a_{ji} = a_{ij}. \tag{88}$$

More generally, when the coefficients are *complex* the most important cases are those in which symmetrically situated elements are *complex conjugates:*

$$a_{ji} = \bar{a}_{ij}. \tag{89}$$

Matrices having the symmetry property (89) are known as *Hermitian matrices*, and are considered in Section 1.17.

The discussion of the present section is to be restricted to *real symmetric matrices*, for which the symmetry property (88) applies. Whereas certain of the results to be obtained hold also for symmetric matrices with imaginary elements, such matrices are of limited importance in applications.

In matrix notation, equation (87) takes the form

$$\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{x} \qquad or \qquad (\mathbf{A} - \lambda\,\mathbf{I})\mathbf{x} = \mathbf{0}, \tag{90}$$

where $\mathbf{I}$ is the unit matrix of order n. This homogeneous problem possesses nontrivial solutions if and only if the determinant of the coefficient matrix $\mathbf{A} - \lambda\,\mathbf{I}$ vanishes:

$$|\mathbf{A} - \lambda\,\mathbf{I}| \equiv \begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix} = 0. \tag{91}$$

This condition requires that λ be a root of an algebraic equation of degree n, known as the *characteristic* (or *secular*) *equation*. The n solutions $\lambda_1, \lambda_2, \cdots, \lambda_n$, which need not all be distinct, are the characteristic numbers or latent roots of the matrix $\mathbf{A}$.

Corresponding to each such value λ_k, there exists at least one vector solution (modal column) of (87) or (90), which is determined within an arbitrary multiplicative constant.* Now let λ_1 and λ_2 be two *distinct* characteristic numbers and denote corresponding characteristic vectors by $\mathbf{u}_1$ and $\mathbf{u}_2$, respectively, so that the equations

$$\mathbf{A}\,\mathbf{u}_1 = \lambda_1\mathbf{u}_1, \qquad \mathbf{A}\,\mathbf{u}_2 = \lambda_2\mathbf{u}_2 \qquad (\lambda_1 \neq \lambda_2) \tag{92a,b}$$

are satisfied. If we postmultiply the transpose of (92a) by $\mathbf{u}_2$ there follows

$$(\mathbf{A}\,\mathbf{u}_1)^T\,\mathbf{u}_2 = \lambda_1\mathbf{u}_1{}^T\,\mathbf{u}_2$$

or, using (34),

$$\mathbf{u}_1{}^T\,\mathbf{A}^T\,\mathbf{u}_2 = \lambda_1\mathbf{u}_1{}^T\,\mathbf{u}_2. \tag{93a}$$

Also, by premultiplying (92b) by $\mathbf{u}_1{}^T$, we obtain

$$\mathbf{u}_1{}^T\,\mathbf{A}\,\mathbf{u}_2 = \lambda_2\mathbf{u}_1{}^T\,\mathbf{u}_2. \tag{93b}$$

The result of subtracting (93a) from (93b), and noticing that for a *symmetric* matrix $\mathbf{A}^T = \mathbf{A}$, is then the relation

$$(\lambda_2 - \lambda_1)(\mathbf{u}_1, \mathbf{u}_2) = 0, \tag{93c}$$

* As was shown in Section 1.9, the components of this solution vector can be expressed as arbitrary multiples of the cofactors of the elements in a row of the matrix $\mathbf{A} - \lambda_k\mathbf{I}$ unless all those cofactors vanish.

and, since we have assumed that $\lambda_1 \neq \lambda_2$, we thus have the following important result:

Two characteristic vectors of a real symmetric matrix, corresponding to different characteristic numbers, are orthogonal,

$$(\mathbf{u}_1, \mathbf{u}_2) = 0. \tag{94}$$

A second basic result is that the characteristic numbers of such a matrix are always *real.* To establish this fact, we suppose that $\lambda_1 = \alpha + i\beta$ is a root of (91), where α and β are real.

Then if $\mathbf{u}_1$ is a corresponding characteristic vector, so that

$$\mathbf{A}\,\mathbf{u}_1 = \lambda_1 \mathbf{u}_1,$$

there follows also

$$\mathbf{A}\,\bar{\mathbf{u}}_1 = \bar{\lambda}_1 \bar{\mathbf{u}}_1,$$

by virtue of the fact that the conjugate of a product is the product of the conjugates, and of the fact that here $\mathbf{A}$ is real. Thus $\lambda_2 = \alpha - i\beta = \bar{\lambda}_1$ must also be a characteristic number with $\mathbf{u}_2 = \bar{\mathbf{u}}_1$ as an associated characteristic vector. By premultiplying the first of the relations

$$\mathbf{A}\,\mathbf{u}_1 = \lambda_1 \mathbf{u}_1, \qquad \mathbf{A}\,\bar{\mathbf{u}}_1 = \bar{\lambda}_1\,\bar{\mathbf{u}}_1 \tag{95a,b}$$

by $\bar{\mathbf{u}}_1{}^T$ and postmultiplying the transpose of the second by $\mathbf{u}_1$, we obtain the relation

$$(\lambda_1 - \bar{\lambda}_1)(\bar{\mathbf{u}}_1, \mathbf{u}_1) = \bar{\mathbf{u}}_1{}^T\,\mathbf{A}\,\mathbf{u}_1 - (\mathbf{A}\,\bar{\mathbf{u}}_1)^T\,\mathbf{u}_1 = 0. \tag{96}$$

But now, since the product $(\bar{\mathbf{u}}_1, \mathbf{u}_1)$ is a *positive* quantity, it follows that $\lambda_1 - \bar{\lambda}_1 = 2i\beta$ must vanish, so that λ_1 must be real. Thus we conclude that *all characteristic numbers of a real symmetric matrix are real.*

Accordingly, all characteristic *vectors* also can be taken to be real, by rejecting permissible imaginary multiplicative factors.

If a characteristic number, say λ_1, of a symmetric matrix is a multiple root of multiplicity s, that is, if the left-hand member of (91) possesses the factor $(\lambda - \lambda_1)^s$, *then to λ_1 there correspond s linearly independent characteristic vectors,* any nontrivial linear combinations of which accordingly also have the same property. Proof of this important fact is postponed to Section 1.18.

The preceding statement is *not* necessarily true for *nonsymmetric* matrices, as can be seen by considering the equations

$$\left.\begin{aligned} x_1 + x_2 &= \lambda x_1, \\ -x_1 - x_2 &= \lambda x_2 \end{aligned}\right\},$$

for which

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}.$$

Here the characteristic equation is readily found to be merely $\lambda^2 = 0$, so that $\lambda = 0$ is a characteristic number of multiplicity *two*. However, when $\lambda = 0$, the only possible solution is given by $x_1 = C_1$, $x_2 = -C_1$ and hence $\mathbf{x} = C_1\{1, -1\}$. Thus, here the double root $\lambda = 0$ corresponds to only *one* characteristic vector

As is shown in the following section, it is always possible to choose s linearly independent vectors corresponding to a characteristic number of multiplicity s in such a way that they are orthogonal *to each other*, in addition to being (automatically) orthogonal to all other characteristic vectors. Thus, if multiple roots of (91) are counted separately, we obtain always exactly n characteristic numbers, and we can determine a corresponding set of n mutually orthogonal characteristic vectors. By virtue of the results of Section 1.10, this set of vectors comprises a *basis* in n-space; that is, *any vector in n-space can be expressed as some linear combination of these n vectors.*

In particular, if each of the n orthogonal characteristic vectors has been divided by its length, and so is a *unit* vector, we say that the resultant set of vectors is an *orthonormal* set. If these vectors are denoted by $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$, so that there follows also

$$(\mathbf{e}_i, \mathbf{e}_k) = \delta_{ik}, \tag{97}$$

then the ith coefficient in the representation

$$\mathbf{v} = \sum_{k=1}^{n} \alpha_k \mathbf{e}_k \tag{98}$$

of an arbitrary n-dimensional vector $\mathbf{v}$ is easily obtained by forming the scalar product of $\mathbf{e}_i$ with the equal members of (98), in the form

$$\alpha_i = (\mathbf{e}_i, \mathbf{v}). \tag{99}$$

Consider now the *nonhomogeneous* equation

$$\mathbf{A}\,\mathbf{x} - \lambda\,\mathbf{x} = \mathbf{c}, \tag{100}$$

where $\mathbf{A}$ is a real symmetric matrix. This equation reduces to (87) or (90) when $\mathbf{c} = \mathbf{0}$. If (100) has a solution, then that solution can be expressed as a linear combination of the characteristic vectors of $\mathbf{A}$. We suppose that n orthogonal unit characteristic vectors $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$ are known, and notice that they satisfy the respective equations

$$\mathbf{A}\,\mathbf{e}_1 = \lambda_1 \mathbf{e}_1, \qquad \cdots, \qquad \mathbf{A}\,\mathbf{e}_n = \lambda_n \mathbf{e}_n. \tag{101}$$

The solution of (100) then can be assumed in the form

$$\mathbf{x} = \sum_{k=1}^{n} \alpha_k \mathbf{e}_k, \tag{102}$$

where the constants α_k are to be determined. The introduction of (102) into (100), and the use of (101), then leads to the requirement

$$\sum_{k=1}^{n}(\lambda_k - \lambda)\alpha_k \mathbf{e}_k = \mathbf{c}. \tag{103}$$

By forming the scalar product of any $\mathbf{e}_i$ with both sides of (103) we may deduce that (102) satisfies (100) if and only if the ith coefficient α_i satisfies the equation

$$(\lambda_i - \lambda)\alpha_i = (\mathbf{e}_i, \mathbf{c}) \qquad (i = 1, 2, \cdots, n). \tag{104}$$

Hence, *if λ is not a characteristic number*, the solution (102) is obtained in the form

$$\mathbf{x} = \sum_{k=1}^{n} \frac{(\mathbf{e}_k, \mathbf{c})}{\lambda_k - \lambda}\, \mathbf{e}_k. \tag{105}$$

Thus a unique solution of the nonhomogeneous problem is obtained when λ is not a characteristic number. *If $\lambda = \lambda_p$, no solution exists unless the vector* $\mathbf{c}$ *is orthogonal to the characteristic vector* (*or vectors*) *corresponding to* λ_p. In case this condition is satisfied, equation (104) shows that the corresponding coefficient (or coefficients) α_p may be chosen *arbitrarily*, so that *infinitely many* solutions then exist.

In particular, if $\lambda = 0$, equation (100) reduces to the equation

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

which was studied previously. This equation thus has a unique solution unless $\lambda = 0$ is a characteristic number of $\mathbf{A}$, that is, unless the equation $\mathbf{A}\,\mathbf{x} = \mathbf{0}$ has nontrivial solutions. In this exceptional situation no solution exists unless $\mathbf{c}$ is orthogonal to the vectors which satisfy $\mathbf{A}\,\mathbf{x} = \mathbf{0}$, in which case infinitely many solutions exist. This result is in accordance with the results of the preceding section, where it was shown that the requirement for the existence of a solution in the exceptional case is that $\mathbf{c}$ be orthogonal to the vectors which satisfy the equation $\mathbf{A}^T\mathbf{x} = \mathbf{0}$, since in the present case we have considered only a symmetric matrix, for which $\mathbf{A}^T = \mathbf{A}$.

The existence criterion obtained here, in the *more general* case when $\lambda = \lambda_p$, is also obtainable from the last result of the preceding section, by replacing $\mathbf{A}$ by $\mathbf{A} - \lambda_p\mathbf{I}$ in that result, and noticing that the latter matrix is symmetric when $\mathbf{A}$ is symmetric.

1.13. Orthogonalization of vector sets. It is often desirable, as in the preceding section, to form from a set of s linearly independent vectors $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_s$ an *orthogonal* set of s linear combinations of the original vectors. It is also convenient to "normalize" the vectors in such a way that each is a *unit vector*. The following procedure is a simple one, and it can be extended by analogy to other similar problems.

We first select *any one* of the original vectors, say $\mathbf{v}_1 = \mathbf{u}_1$, and divide it by its length. This is the first member of the desired set:

$$\mathbf{e}_1 = \frac{\mathbf{u}_1}{l(\mathbf{u}_1)}. \tag{106}$$

We next choose a second vector, say $\mathbf{u}_2$, from the original set and write $\mathbf{v}_2 = \mathbf{u}_2 - c\,\mathbf{e}_1$. The requirement that $\mathbf{v}_2$ be orthogonal to $\mathbf{e}_1$ leads to the determination

$$(\mathbf{e}_1, \mathbf{v}_2) = (\mathbf{e}_1, \mathbf{u}_2) - c(\mathbf{e}_1, \mathbf{e}_1) = 0$$

or

$$c = (\mathbf{e}_1, \mathbf{u}_2),$$

so that

$$\mathbf{v}_2 = \mathbf{u}_2 - (\mathbf{e}_1, \mathbf{u}_2)\mathbf{e}_1. \tag{107a}$$

Since $\mathbf{e}_1$ is a unit vector, the familiar geometrical interpretation of the scalar product in two or three dimensions leads us to say that $(\mathbf{e}_1, \mathbf{u}_2)$ is "the *scalar component* of $\mathbf{u}_2$ in the direction of $\mathbf{e}_1$," and hence that in (107a) we have "subtracted off the $\mathbf{e}_1$ component of $\mathbf{u}_2$."

The second member, $\mathbf{e}_2$, of the desired set of orthogonal unit vectors is obtained by dividing $\mathbf{v}_2$ by its length:

$$\mathbf{e}_2 = \frac{\mathbf{v}_2}{l(\mathbf{v}_2)}. \tag{107b}$$

In the third step we write $\mathbf{v}_3 = \mathbf{u}_3 - c_1\mathbf{e}_1 - c_2\mathbf{e}_2$. The requirement that $\mathbf{v}_3$ be simultaneously orthogonal to $\mathbf{e}_1$ and $\mathbf{e}_2$ then determines values of c_1 and c_2 which are in accordance with the geometrical interpretation described above, and there follows

$$\mathbf{v}_3 = \mathbf{u}_3 - (\mathbf{e}_1, \mathbf{u}_3)\mathbf{e}_1 - (\mathbf{e}_2, \mathbf{u}_3)\mathbf{e}_2, \tag{108a}$$

so that the "$\mathbf{e}_1$ and $\mathbf{e}_2$ components" of $\mathbf{u}_3$ are subtracted off. The third required vector $\mathbf{e}_3$ is then given by

$$\mathbf{e}_3 = \frac{\mathbf{v}_3}{l(\mathbf{v}_3)}. \tag{108b}$$

A continuation of this process finally determines the sth member of the required set in the form

$$\mathbf{e}_s = \frac{\mathbf{v}_s}{l(\mathbf{v}_s)} \quad \textit{where} \quad \mathbf{v}_s = \mathbf{u}_s - \sum_{k=1}^{s-1} (\mathbf{e}_k, \mathbf{u}_s)\mathbf{e}_k. \tag{109}$$

This method, which is often called the *Gram-Schmidt orthogonalization procedure*, would fail if and only if at some stage $\mathbf{v}_r = \mathbf{0}$. But this would mean that $\mathbf{u}_r$ is a linear combination of $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_{r-1}$, and hence also a linear combination of $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_{r-1}$, in contradiction with the hypothesis that the set of $\mathbf{u}$'s is linearly independent.

It is seen that this procedure permits the determination of an *orthonormal basis* (that is, a basis comprising *mutually orthogonal unit vectors*) for a vector space when *any* set of spanning vectors is known.

When the vectors $\mathbf{u}_1, \cdots, \mathbf{u}_s$ are *complex*, the same procedure clearly applies if *Hermitian* products and lengths are used throughout.

1.14. Quadratic forms. A homogeneous expression of second degree, of the form

$$A \equiv a_{11}x_1^2 + a_{22}x_2^2 + \cdots + a_{nn}x_n^2 + 2a_{12}x_1x_2 + 2a_{13}x_1x_3 + \cdots + 2a_{n-1,n}x_{n-1}x_n, \quad (110)$$

is called a *quadratic form* in $x_1, x_2, \cdots, x_n$. It will be supposed here that the elements a_{ij} and the variables x_i are *real*. In two-dimensional space the equation $A =$ constant represents a second-degree curve (conic) with center at the origin, while in three-dimensional space the equation $A =$ constant represents a central *quadric surface* with center at the origin. Many problems associated with such forms are intimately related to problems associated with sets of linear equations.

We may notice first that if we write

$$y_i = \frac{1}{2}\frac{\partial A}{\partial x_i} \qquad (i = 1, 2, \cdots, n),$$

we obtain the equations

$$\left.\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= y_1, \\ a_{12}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= y_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{1n}x_1 + a_{2n}x_2 + \cdots + a_{nn}x_n &= y_n \end{aligned}\right\}. \quad (111)$$

This set of equations can be written in the form

$$\mathbf{A}\,\mathbf{x} = \mathbf{y}, \quad (112)$$

where $\mathbf{A} = [a_{ij}]$ is a *symmetric* matrix. That is, the elements satisfy the symmetry condition

$$a_{ji} = a_{ij}. \quad (113)$$

On the other hand, it is easily seen that (110) is equivalent to the relation $A \equiv (\mathbf{x}, \mathbf{y})$; that is, (110) can be written in the form

$$A \equiv \mathbf{x}^T\,\mathbf{A}\,\mathbf{x}. \quad (114)$$

In many cases it is desirable to express $x_1, x_2, \cdots, x_n$ as linear combinations of new real variables $x_1', x_2', \cdots, x_n'$ in such a way that A is reduced to a linear combination of only *squares* of the new variables, the cross-product terms being eliminated. A form of this type is said to be a *canonical form*.

Let the vector $\mathbf{x}$ be expressed in terms of $\mathbf{x}'$ by the equation

$$\mathbf{x} = \mathbf{Q}\,\mathbf{x}', \tag{115}$$

where $\mathbf{Q}$ is a square matrix of order n. The introduction of (115) into (114) then gives

$$A = (\mathbf{Q}\,\mathbf{x}')^T\,\mathbf{A}\,\mathbf{Q}\,\mathbf{x}' = \mathbf{x}'^T\,\mathbf{Q}^T\,\mathbf{A}\,\mathbf{Q}\,\mathbf{x}' \tag{116}$$

or

$$A = \mathbf{x}'^T\,\mathbf{A}'\,\mathbf{x}', \tag{117}$$

where the new matrix $\mathbf{A}'$ is defined by the equation

$$\mathbf{A}' = \mathbf{Q}^T\,\mathbf{A}\,\mathbf{Q}. \tag{118}$$

Thus we see that, if A is to involve only squares of the variables x_i', the matrix $\mathbf{Q}$ in (115) must be chosen so that $\mathbf{Q}^T\,\mathbf{A}\,\mathbf{Q}$ is a *diagonal matrix;* that is, so that all elements for which $i \neq j$ vanish.

We show next that if the characteristic numbers and corresponding characteristic vectors of the real symmetric matrix $\mathbf{A}$ are known, a matrix $\mathbf{Q}$ having this property can be very easily constructed. Suppose that the characteristic numbers of $\mathbf{A}$ are $\lambda_1, \lambda_2, \cdots, \lambda_n$, repeated roots of the characteristic equation being numbered separately, and denote the corresponding members of the orthogonalized set of n characteristic unit vectors by $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$. We then have the relations

$$\mathbf{A}\,\mathbf{e}_1 = \lambda_1\mathbf{e}_1, \qquad \cdots, \qquad \mathbf{A}\,\mathbf{e}_n = \lambda_n\mathbf{e}_n. \tag{119}$$

Let a matrix $\mathbf{Q}$ be constructed in such a way that the elements of the unit vectors $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$ are the elements of the successive *columns* of $\mathbf{Q}$:

$$\mathbf{Q} = \begin{bmatrix} e_{11} & e_{21} & \cdots & e_{n1} \\ e_{12} & e_{22} & \cdots & e_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ e_{1n} & e_{2n} & \cdots & e_{nn} \end{bmatrix}. \tag{120}$$

Then, if use is made of (119), it is easily seen that

$$\mathbf{A}\,\mathbf{Q} = \begin{bmatrix} \lambda_1 e_{11} & \lambda_2 e_{21} & \cdots & \lambda_n e_{n1} \\ \lambda_1 e_{12} & \lambda_2 e_{22} & \cdots & \lambda_n e_{n2} \\ \cdots & \cdots & \cdots & \cdots \\ \lambda_1 e_{1n} & \lambda_2 e_{2n} & \cdots & \lambda_n e_{nn} \end{bmatrix} \tag{121a}$$

or

$$\mathbf{A}\,\mathbf{Q} = \mathbf{Q} \cdot \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}. \tag{121b}$$

This relation follows directly from the fact that the product of **A** into the kth column of **Q** is the kth column of the right-hand member of (121a). Since the vectors $\mathbf{e}_1, \cdots, \mathbf{e}_n$ are linearly independent, it follows that $|\mathbf{Q}| \neq 0$. Thus the inverse $\mathbf{Q}^{-1}$ exists, and by premultiplying the equal members of (121b) by $\mathbf{Q}^{-1}$ we obtain the result

$$\mathbf{Q}^{-1}\,\mathbf{A}\,\mathbf{Q} = [\lambda_i \delta_{ij}]. \tag{122}$$

Hence, the matrix **A** is diagonalized by the indicated operations, the diagonal elements being merely the characteristic numbers of **A**.

However, the *desired* diagonalization (118) was to be of the form $\mathbf{Q}^T\,\mathbf{A}\,\mathbf{Q}$. Thus, the matrix **Q** defined by (120) is not acceptable for present purposes unless it can be shown that

$$\mathbf{Q}^T = \mathbf{Q}^{-1} \qquad or \qquad \mathbf{Q}^T\,\mathbf{Q} = \mathbf{I}. \tag{123}$$

But the typical term p_{ij} of the product $\mathbf{Q}^T\,\mathbf{Q}$ is of the form

$$p_{ij} = \sum_{k=1}^{n} e_{ik} e_{jk},$$

where $$\mathbf{e}_i = \{e_{i1}, e_{i2}, \cdots, e_{in}\},$$

and since the **e**'s are *orthogonal* the indicated sum *vanishes* unless $i = j$, in which case the sum is *unity* since the **e**'s are *unit vectors*.

Hence there follows $p_{ij} = \delta_{ij}$; that is, $\mathbf{Q}^T\,\mathbf{Q} = [\delta_{ij}] = \mathbf{I}$, as is required by (123). Further, since $|\mathbf{Q}| = |\mathbf{Q}^T|$, we may obtain from (123) the useful result

$$|\mathbf{Q}|^2 = 1: \qquad |\mathbf{Q}| = \pm 1. \tag{124}$$

It follows that the matrix **Q** defined by (120) does indeed have the property that *the quadratic form*

$$A = \mathbf{x}^T\,\mathbf{A}\,\mathbf{x} \tag{125}$$

is reduced by the change in variables

$$\mathbf{x} = \mathbf{Q}\,\mathbf{x}' \tag{126}$$

to the form

$$A = \mathbf{x}'^T\,\mathbf{A}'\,\mathbf{x}' \qquad where \qquad \mathbf{A}' = [\lambda_i \delta_{ij}],$$

that is, to the form

$$A = \lambda_1 x_1'^2 + \lambda_2 x_2'^2 + \cdots + \lambda_n x_n'^2, \tag{127}$$

where the numbers λ_i are the characteristic numbers of **A**.

A matrix whose columns comprise the elements of n linearly independent characteristic vectors of a given matrix **A**, of order n, is called a *modal matrix* of **A**. In particular, when those n vectors are mutually orthogonal and of unit length, it is convenient to say that the modal matrix is *orthonormal*. Thus the matrix **Q** is an *orthonormal modal matrix* of **A**.

We notice that if $\lambda = 0$ is an s-fold root of the characteristic equation the form (127) has only $n - s$ nonvanishing terms. It is shown in Section 1.18 that this situation arises if and only if the symmetric matrix $\mathbf{A}$ is of *rank* $r = n - s$.

The new variables x_i' are related to the original ones, in accordance with (115) and (123), by the equation

$$\mathbf{x}' = \mathbf{Q}^{-1}\mathbf{x} = \mathbf{Q}^T\mathbf{x}, \tag{128}$$

and hence are of the form

$$\left.\begin{aligned} x_1' &= e_{11}x_1 + e_{12}x_2 + \cdots + e_{1n}x_n, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ x_n' &= e_{n1}x_1 + e_{n2}x_2 + \cdots + e_{nn}x_n \end{aligned}\right\} \tag{129}$$

or

$$x_i' = (\mathbf{e}_i, \mathbf{x}) \qquad (i = 1, 2, \cdots, n). \tag{129'}$$

When all the characteristic numbers of $\mathbf{A}$ are *distinct*, the orthonormal modal matrix $\mathbf{Q}$ is uniquely determined except for the ordering of the columns and the arbitrary algebraic sign associated with each column. However, if a root is of multiplicity s, the corresponding s orthogonalized unit vectors can be chosen in infinitely many ways, as was shown in Section 1.13.

We remark that the modal matrix $\mathbf{Q}$ specified by (120) is not the *only* matrix which can be used in (126) to reduce a quadratic form to a sum of squares. However, it is the only such matrix which possesses the useful property that

$$\mathbf{Q}^T = \mathbf{Q}^{-1}.$$

A matrix having this property is called an *orthogonal matrix*.

It is easily seen that *a square matrix is an orthogonal matrix if and only if its columns comprise the elements of real mutually orthogonal unit vectors.*

1.15. A numerical example. To illustrate the preceding reduction in a specific numerical case, we consider the quadratic form

$$A = 25x_1^2 + 34x_2^2 + 41x_3^2 - 24x_2x_3.$$

The corresponding matrix $\mathbf{A}$ is then of the form

$$\mathbf{A} = \begin{bmatrix} 25 & 0 & 0 \\ 0 & 34 & -12 \\ 0 & -12 & 41 \end{bmatrix}$$

and the equations $\mathbf{A}\mathbf{x} - \lambda\mathbf{x} = \mathbf{0}$ become

$$\begin{aligned} (25 - \lambda)x_1 \qquad\qquad\quad &= 0, \\ (34 - \lambda)x_2 \quad -12x_3 &= 0, \\ -12x_2 + (41 - \lambda)x_3 &= 0. \end{aligned}$$

The characteristic equation $|\mathbf{A} - \lambda \mathbf{I}| = 0$ then takes the form

$$(25 - \lambda)(\lambda^2 - 75\lambda + 1250) = 0,$$

from which the characteristic numbers are determined:

$$\lambda_1 = \lambda_2 = 25, \qquad \lambda_3 = 50.$$

When $\lambda = \lambda_1 = \lambda_2 = 25$, the equations $\mathbf{A}\,\mathbf{x} - \lambda\,\mathbf{x} = \mathbf{0}$ become

$$\begin{aligned} 0 &= 0, \\ 9x_2 - 12x_3 &= 0, \\ -12x_2 + 16x_3 &= 0, \end{aligned}$$

with the general solution $x_1 = C_1$, $x_2 = C_2$, $x_3 = \frac{3}{4}C_2$. In vector form we may write $\mathbf{x} = C_1\mathbf{u}_1 + C_2\mathbf{u}_2$, where $\mathbf{u}_1 = \{1, 0, 0\}$ and $\mathbf{u}_2 = \{0, 1, \frac{3}{4}\}$. Since it happens that $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal, we need only divide them by their lengths $l_1 = 1$ and $l_2 = \frac{5}{4}$ to obtain the two orthogonal unit characteristic vectors

$$\mathbf{e}_1 = \{1, 0, 0\}, \qquad \mathbf{e}_2 = \{0, \tfrac{4}{5}, \tfrac{3}{5}\}.$$

In a similar way, a unit characteristic vector corresponding to $\lambda = \lambda_3 = 50$ is found to be

$$\mathbf{e}_3 = \{0, \tfrac{3}{5}, -\tfrac{4}{5}\}.$$

Hence the orthonormal modal matrix $\mathbf{Q}$ of equation (120) can be taken in the form

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{4}{5} & \frac{3}{5} \\ 0 & \frac{3}{5} & -\frac{4}{5} \end{bmatrix}$$

and the new coordinates defined by (129) are then given by

$$\begin{aligned} x_1' &= x_1, \\ x_2' &= \tfrac{4}{5}x_2 + \tfrac{3}{5}x_3, \\ x_3' &= \tfrac{3}{5}x_2 - \tfrac{4}{5}x_3. \end{aligned}$$

With this choice of the new coordinates, (127) states that the quadratic form under consideration takes the form

$$A \equiv 25x_1'^2 + 25x_2'^2 + 50x_3'^2.$$

In particular, it follows that the quadric surface with the equation

$$25x_1^2 + 34x_2^2 + 41x_3^2 - 24x_2x_3 = 25$$

takes the standard form

$$x_1'^2 + x_2'^2 + 2x_3'^2 = 1$$

with the introduction of the new coordinates. It is shown in Section 1.21 that *the new x_1'-x_n' coordinate system defined by* (126) *is also a rectangular system when* $\mathbf{Q}$ *is an orthogonal matrix* and that *length and angle then are preserved by the transformation.* Hence the quadric surface just considered is an *oblate spheroid* with semiaxes of length 1, 1, $\sqrt{2}/2$.

In the present example it happens that the chosen matrix $\mathbf{Q}$ is *symmetric*, and hence also is such that $\mathbf{Q} = \mathbf{Q}^T = \mathbf{Q}^{-1}$.

It may be noticed that, by the usual method of "completing squares," we may, for example, also reduce the form A as follows:

$$\begin{aligned} A &= 25x_1^2 + 34[x_2^2 - \tfrac{24}{34}x_2x_3 + (\tfrac{12}{34})^2x_3^2] + (41 - \tfrac{144}{34})x_3^2 \\ &= 25x_1^2 + 34(x_2 - \tfrac{6}{17}x_3)^2 + \tfrac{625}{17}x_3^2. \end{aligned}$$

Hence, if we introduce new variables by the relations

$$\begin{aligned} x_1' &= x_1, \\ x_2' &= x_2 - \tfrac{6}{17}x_3, \\ x_3' &= x_3, \end{aligned}$$

we can reduce A to the form

$$A = 25x_1'^2 + 34x_2'^2 + \tfrac{625}{17}x_3'^2.$$

However, here the matrix $\mathbf{Q}$ for which $\mathbf{x} = \mathbf{Q}\,\mathbf{x}'$, and which takes the *triangular* form

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \frac{6}{17} \\ 0 & 0 & 1 \end{bmatrix},$$

is *not* an orthogonal matrix. Consequently, as is shown in Section 1.21, the new $x_1'x_2'x_3'$ coordinate system is *not* a rectangular system in this case; that is, the new coordinate axes are *not mutually perpendicular.* Nevertheless, the matrix $\mathbf{Q}$ does have the property that $\mathbf{Q}^T\mathbf{A}\,\mathbf{Q}$ is a diagonal matrix.

1.16. Equivalent matrices and transformations. Two matrices $\mathbf{A}$ and $\mathbf{B}$ which can be obtained from each other by a finite number of successive applications of the *elementary operations* (Section 1.8) to rows and/or columns are said to be *equivalent* (but not necessarily *equal*) *matrices.*

It can be shown that any such sequence of operations on the *rows* of $\mathbf{A}$ can be effected by *pre*multiplying $\mathbf{A}$ by some nonsingular matrix $\mathbf{P}$, while corresponding operations on *columns* can always be effected by *post*-multiplying $\mathbf{A}$ by a nonsingular matrix $\mathbf{Q}$. This result is a consequence of the easily established fact that an elementary operation on rows (columns) of $\mathbf{A}$ may be accomplished by first performing that operation on the *unit*

matrix **I** of appropriate order, and then premultiplying (postmultiplying) **A** by the resultant matrix (see Problems 32 and 33).

The converse of the preceding statement is also true (see Problem 34); that is, *the matrices* **A** *and* **B** *are equivalent if and only if nonsingular matrices* **P** *and* **Q** *exist such that* $\mathbf{B} = \mathbf{P}\,\mathbf{A}\,\mathbf{Q}$.

Since the elementary operations do not change the rank of a matrix, it follows that *two equivalent matrices have the same rank.*

Transformations of the form $\mathbf{P}\,\mathbf{A}\,\mathbf{Q}$ are classified according to restrictions imposed on **P** and **Q**. Thus if $\mathbf{P} = \mathbf{Q}^T = \mathbf{Q}^{-1}$, as in the reduction of Section 1.14, the transformation is called an *orthogonal* transformation. If only $\mathbf{P} = \mathbf{Q}^T$, as is required by equation (118), the resulting transformation $\mathbf{Q}^T\,\mathbf{A}\,\mathbf{Q}$ is called a *congruence* transformation, whereas a transformation of the form $\mathbf{Q}^{-1}\,\mathbf{A}\,\mathbf{Q}$, for which $\mathbf{P} = \mathbf{Q}^{-1}$, is called a *similarity* transformation. This terminology is motivated by certain geometrical considerations. We notice that *an orthogonal transformation is both a congruence and a similarity transformation.*

Conjunctive and *unitary* transformations, which are of importance in dealing with matrices of *complex* elements, are defined in the following section.

1.17. Hermitian matrices. We now consider a matrix with *complex* elements which satisfy the relation

$$a_{ji} = \bar{a}_{ij}. \tag{130}$$

Such a matrix is hence of the special form

$$\mathbf{H} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \bar{a}_{12} & a_{22} & a_{23} & \cdots & a_{2n} \\ \bar{a}_{13} & \bar{a}_{23} & a_{33} & \cdots & a_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \bar{a}_{1n} & \bar{a}_{2n} & \bar{a}_{3n} & \cdots & a_{nn} \end{bmatrix}, \tag{131}$$

and is known as a *Hermitian* matrix. Thus a Hermitian matrix has the property that two elements situated symmetrically with respect to the principal diagonal are *complex conjugates*. In particular, (130) requires that the elements in the principal diagonal ($i = j$) be *real.*

We see that the *conjugate* of the matrix **H**, obtained by replacing each element by its complex conjugate, and denoted by $\overline{\mathbf{H}}$, is equal to the transpose of **H**:

$$\mathbf{H}^T = \overline{\mathbf{H}}. \tag{132}$$

The product

$$H \equiv \bar{\mathbf{x}}^T\,\mathbf{H}\,\mathbf{x} \tag{133}$$

is known as a *Hermitian form.* In two dimensions, the general Hermitian form is thus given by

$$H \equiv [\bar{x}_1 \quad \bar{x}_2]\begin{bmatrix} a_{11} & a_{12} \\ \bar{a}_{12} & a_{22} \end{bmatrix}\begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix}$$
$$\equiv a_{11}\bar{x}_1 x_1 + (a_{12}\bar{x}_1 x_2 + \bar{a}_{12}\bar{x}_2 x_1) + a_{22}\bar{x}_2 x_2. \tag{134}$$

Although the elements a_{ij} and variables x_i may be complex, *the values assumed by a Hermitian form are always real.* To establish this fact, we recall first that *the conjugate of a product of complex quantities is equal to the product of the conjugates.* Thus, if H were imaginary (nonreal), and given by (133), then its conjugate $\bar{H}$ would be given by

$$\bar{H} = \mathbf{x}^T \overline{\mathbf{H}}\, \bar{\mathbf{x}} = \mathbf{x}^T \mathbf{H}^T \bar{\mathbf{x}} = (\mathbf{H}\,\mathbf{x})^T \bar{\mathbf{x}} = \bar{\mathbf{x}}^T (\mathbf{H}\,\mathbf{x}) = H. \tag{135}$$

But $\bar{H} = H$ only if H is real, as was to be shown.

Also, we can show that the characteristic numbers of a Hermitian matrix are real. For if $\mathbf{u}_1$ is a characteristic vector corresponding to λ_1, we must have

$$\mathbf{H}\,\mathbf{u}_1 = \lambda_1 \mathbf{u}_1, \tag{136}$$

and hence also, after premultiplying both sides by $\bar{\mathbf{u}}_1{}^T$,

$$\bar{\mathbf{u}}_1{}^T \mathbf{H}\,\mathbf{u}_1 = \lambda_1 \bar{\mathbf{u}}_1{}^T \mathbf{u}_1. \tag{137}$$

But since $\bar{\mathbf{u}}_1{}^T \mathbf{H}\,\mathbf{u}_1$ and $\bar{\mathbf{u}}_1{}^T \mathbf{u}_1$ are both real, and $\bar{\mathbf{u}}_1{}^T \mathbf{u}_1 \neq 0$, λ_1 must also be real.

Further, let $\mathbf{u}_2$ be a characteristic vector corresponding to a second characteristic number $\lambda_2 \neq \lambda_1$, so that

$$\mathbf{H}\,\mathbf{u}_2 = \lambda_2 \mathbf{u}_2. \tag{138}$$

If the transposed conjugate of (136) is postmultiplied by $\mathbf{u}_2$, there follows

$$(\overline{\mathbf{H}}\,\bar{\mathbf{u}}_1)^T \mathbf{u}_2 = \lambda_1 \bar{\mathbf{u}}_1{}^T \mathbf{u}_2,$$

while premultiplication of (138) by $\bar{\mathbf{u}}_1{}^T$ leads to the relation

$$\bar{\mathbf{u}}_1{}^T \mathbf{H}\,\mathbf{u}_2 = \lambda_2 \bar{\mathbf{u}}_1{}^T \mathbf{u}_2.$$

By subtracting these equations from each other, and using equations (34) and (132), there follows

$$(\lambda_2 - \lambda_1)\bar{\mathbf{u}}_1{}^T \mathbf{u}_2 = \bar{\mathbf{u}}_1{}^T \mathbf{H}\,\mathbf{u}_2 - (\overline{\mathbf{H}}\,\bar{\mathbf{u}}_1)^T \mathbf{u}_2$$
$$= \bar{\mathbf{u}}_1{}^T \mathbf{H}\,\mathbf{u}_2 - \bar{\mathbf{u}}_1{}^T \overline{\mathbf{H}}^T \mathbf{u}_2$$
$$= 0.$$

Hence we conclude that *two characteristic vectors of a Hermitian matrix, corresponding to different characteristic numbers, are orthogonal in the Hermitian sense:*

$$(\bar{\mathbf{u}}_1, \mathbf{u}_2) \equiv \bar{\mathbf{u}}_1{}^T \mathbf{u}_2 = 0. \tag{139}$$

The vectors $\mathbf{u}_i$ then can be divided by their Hermitian lengths

$$l_H(\mathbf{u}_i) = (\bar{\mathbf{u}}_i, \mathbf{u}_i)^{1/2},$$

to give a set of orthogonal *unit* vectors $\mathbf{e}_i$ corresponding to successive non-repeated roots of the characteristic equation. Corresponding to a root of multiplicity s there exists a set of s linearly independent characteristic vectors (see Section 1.18), which can be orthogonalized and reduced to absolute length unity, by a procedure completely analogous to that given in Section 1.13. Thus we may again obtain a set of n mutually orthogonal unit characteristic vectors $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$, orthogonality and length being defined in the Hermitian sense.

It then follows that any complex n-dimensional vector $\mathbf{v}$ can be expressed in the form

$$\mathbf{v} = \sum_{k=1}^{n} \alpha_k \mathbf{e}_k, \tag{140}$$

where the kth coefficient is given by the formula

$$\alpha_k = (\bar{\mathbf{e}}_k, \mathbf{v}). \tag{141}$$

In order to solve the equation

$$\mathbf{H}\,\mathbf{x} - \lambda\,\mathbf{x} = \mathbf{c}, \tag{142}$$

we may assume the expansion

$$\mathbf{x} = \sum_{k=1}^{n} \alpha_k \mathbf{e}_k, \tag{143}$$

as in the real case, so that (142) takes the form

$$\sum_{k=1}^{n} (\lambda_k - \lambda)\alpha_k \mathbf{e}_k = \mathbf{c}$$

and there follows

$$(\lambda_k - \lambda)\alpha_k = (\bar{\mathbf{e}}_k, \mathbf{c}).$$

Thus, if λ is not a characteristic number, the solution becomes

$$\mathbf{x} = \sum_{k=1}^{n} \frac{(\bar{\mathbf{e}}_k, \mathbf{c})}{\lambda_k - \lambda} \mathbf{e}_k \tag{144}$$

in analogy with (105). If $\lambda = \lambda_p$, no solution exists unless $\mathbf{c}$ is such that $(\bar{\mathbf{e}}_p, \mathbf{c}) = 0$, in which case α_p is arbitrary, and infinitely many solutions exist.

The reduction of a Hermitian form to a sum of the *canonical form*

$$H = \lambda_1 \bar{x}_1' x_1' + \lambda_2 \bar{x}_2' x_2' + \cdots + \lambda_n \bar{x}_n' x_n' \tag{145}$$

may be accomplished by a method analogous to that of Section 1.14. Thus, if we write

$$\mathbf{x} = \mathbf{U}\,\mathbf{x}', \tag{146}$$

the form H of (133) becomes

$$H = (\bar{\mathbf{U}}\,\bar{\mathbf{x}}')^T\,\mathbf{H}\,\mathbf{U}\,\mathbf{x}' = \bar{\mathbf{x}}'^T(\bar{\mathbf{U}}^T\,\mathbf{H}\,\mathbf{U})\mathbf{x}'. \tag{147}$$

This form will be identified with (145) if and only if $\mathbf{U}$ is such that

$$\bar{\mathbf{U}}^T\,\mathbf{H}\,\mathbf{U} = [\lambda_i\delta_{ij}]. \tag{148}$$

As in Section 1.13, a permissible choice of $\mathbf{U}$ consists of an orthonormal modal matrix formed by arranging n orthogonal unit characteristic vectors of $\mathbf{H}$ as its columns, in which case the scalars λ_i are the characteristic numbers of $\mathbf{H}$. For the matrix $\mathbf{U}$ it is found that

$$\bar{\mathbf{U}}^T = \mathbf{U}^{-1} \qquad or \qquad \bar{\mathbf{U}}^T\,\mathbf{U} = \mathbf{I}. \tag{149}$$

A matrix $\mathbf{U}$ having the property (149) is called a *unitary* (or *Hermitian orthogonal*) *matrix*, and the product $\bar{\mathbf{U}}^T\,\mathbf{H}\,\mathbf{U}$ is then called a *unitary transformation* of $\mathbf{H}$. More generally, a transformation of the form $\bar{\mathbf{U}}^T\,\mathbf{H}\,\mathbf{U}$, where $\mathbf{U}$ does not necessarily satisfy (149), is called a *conjunctive transformation* of $\mathbf{H}$.

It is easily seen that *a square matrix is unitary if and only if its columns comprise the elements of mutually orthogonal unit vectors, length and orthogonality being defined in the Hermitian sense.*

1.18. Multiple characteristic numbers of symmetric matrices. We next establish the assertion, made in Section 1.12, that a real symmetric matrix $\mathbf{A}$ with a characteristic number λ_1 of multiplicity s has an s-parameter family of characteristic vectors corresponding to λ_1.

For this purpose, suppose first that $\mathbf{A}$ is a symmetric $n \times n$ matrix such that the characteristic polynomial

$$F(\lambda) = |\mathbf{A} - \lambda\,\mathbf{I}|$$

has $(\lambda - \lambda_1)^2$ as a factor, and let $\mathbf{e}_1$ be *one* unit characteristic vector of $\mathbf{A}$ corresponding to λ_1. Then, if $\mathbf{Q}$ is *any* $n \times n$ *orthogonal* matrix having $\mathbf{e}_1$ as its *first* column vector, there follows

$$\mathbf{Q}^T\,\mathbf{A}\,\mathbf{Q} = \begin{bmatrix} \text{—}\mathbf{e}_1^T\text{→} \\ \cdots\cdots \\ \cdots\cdots \end{bmatrix} \mathbf{A} \begin{bmatrix} | & \vdots & \vdots \\ \mathbf{e}_1 & \vdots & \vdots \\ \downarrow & \vdots & \vdots \end{bmatrix}$$

$$= \begin{bmatrix} \text{—}\mathbf{e}_1^T\text{→} \\ \cdots\cdots \\ \cdots\cdots \end{bmatrix} \begin{bmatrix} | & \vdots & \vdots \\ \lambda_1\mathbf{e}_1 & \vdots & \vdots \\ \downarrow & \vdots & \vdots \end{bmatrix}. \tag{150}$$

Since each vector whose elements comprise a row of $\mathbf{Q}^T$, except the first, is orthogonal to the column $\lambda_1\mathbf{e}_1$, each element of the first column of the

product except the leading element will vanish. Thus the result will be of the form

$$\mathbf{Q}^T \mathbf{A}\, \mathbf{Q} = \begin{bmatrix} \lambda_1 & \alpha_{12} & \cdots & \alpha_{1n} \\ 0 & \alpha_{22} & \cdots & \alpha_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \alpha_{n2} & \cdots & \alpha_{nn} \end{bmatrix} \tag{151}$$

where the α's are certain constants, depending upon the remaining columns of $\mathbf{Q}$. But, since the symmetry of $\mathbf{A}$ implies the symmetry of $\mathbf{Q}^T \mathbf{A}\, \mathbf{Q}$, the elements $\alpha_{12}, \cdots, \alpha_{1n}$ *also* must vanish.

Hence, if $\mathbf{Q}$ is *any* orthogonal matrix having $\mathbf{e}_1$ as its first column vector, the product $\mathbf{Q}^T \mathbf{A}\, \mathbf{Q} = \mathbf{Q}^{-1} \mathbf{A}\, \mathbf{Q}$ is of the form

$$\mathbf{Q}^{-1} \mathbf{A}\, \mathbf{Q} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \alpha_{22} & \cdots & \alpha_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \alpha_{2n} & \cdots & \alpha_{nn} \end{bmatrix} \tag{152}$$

and also there follows

$$\mathbf{Q}^{-1} \mathbf{A}\, \mathbf{Q} - \lambda \mathbf{I} = \begin{bmatrix} \lambda_1 - \lambda & 0 & \cdots & 0 \\ 0 & \alpha_{22} - \lambda & \cdots & \alpha_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \alpha_{2n} & \cdots & \alpha_{nn} - \lambda \end{bmatrix}. \tag{153}$$

Next we notice that, in consequence of the relation

$$\mathbf{Q}^{-1} \mathbf{A}\, \mathbf{Q} - \lambda \mathbf{I} = \mathbf{Q}^{-1}(\mathbf{A} - \lambda \mathbf{I})\mathbf{Q}, \tag{154}$$

there follows also

$$|\mathbf{Q}^{-1} \mathbf{A}\, \mathbf{Q} - \lambda \mathbf{I}| = |\mathbf{A} - \lambda \mathbf{I}| \tag{155}$$

and hence (153) implies the relation

$$\begin{aligned} |\mathbf{A} - \lambda \mathbf{I}| &= \begin{vmatrix} \lambda_1 - \lambda & 0 & \cdots & 0 \\ 0 & \alpha_{22} - \lambda & \cdots & \alpha_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \alpha_{2n} & \cdots & \alpha_{nn} - \lambda \end{vmatrix} \\ &= (\lambda_1 - \lambda) \begin{vmatrix} \alpha_{22} - \lambda & \cdots & \alpha_{2n} \\ \cdots & \cdots & \cdots \\ \alpha_{2n} & \cdots & \alpha_{nn} - \lambda \end{vmatrix}. \end{aligned} \tag{156}$$

But since, by hypothesis, the left-hand member of (156) has $(\lambda_1 - \lambda)^2$ as a factor, it follows that the coefficient of $(\lambda_1 - \lambda)$ in the right-hand member

has $(\lambda_1 - \lambda)$ as a factor, and hence *vanishes when* $\lambda = \lambda_1$. Thus *the matrix* $\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q} - \lambda\mathbf{I}$ *in* (153) *is of rank* $n - 2$ *or less when* $\lambda = \lambda_1$.

Finally, since (154) states that the matrix $\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q} - \lambda_1\mathbf{I}$ is *similar* to the matrix $\mathbf{A} - \lambda_1\mathbf{I}$, we deduce that *the rank of the matrix* $\mathbf{A} - \lambda_1\mathbf{I}$ *is not greater than* $n - 2$ *when* λ_1 *is a characteristic number of* $\mathbf{A}$ *of multiplicity two or more*, that is, that *the matrix equation* $\mathbf{A}\mathbf{x} = \lambda_1\mathbf{x}$ *has at least a two-parameter family of solutions in that case.*

If the multiplicity of λ_1 is *greater* than two, then by taking $\mathbf{Q}$ to be any orthogonal matrix having *two* orthogonal unit vectors $\mathbf{e}_1$ and $\mathbf{e}_2$, both corresponding to λ_1, as its first *two* column vectors, we deduce in a completely analogous way that

$$\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q} - \lambda\mathbf{I} = \begin{bmatrix} \lambda_1 - \lambda & 0 & 0 & \cdots & 0 \\ 0 & \lambda_1 - \lambda & 0 & \cdots & 0 \\ 0 & 0 & \alpha_{33} - \lambda & \cdots & \alpha_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \alpha_{3n} & \cdots & \alpha_{nn} - \lambda \end{bmatrix} \tag{157}$$

and the same argument leads to the conclusion that the matrix $\mathbf{A} - \lambda\mathbf{I}$ is of rank not greater than $n - 3$ when $\lambda = \lambda_1$, so that at least a *three-*parameter family of corresponding characteristic vectors exists when the multiplicity of λ_1 is at least three.

By inductive reasoning, we thus deduce that if λ_1 is a characteristic number of a symmetric matrix $\mathbf{A}$, of multiplicity s, then the rank of the matrix $\mathbf{A} - \lambda\mathbf{I}$ is not greater than $n - s$ when $\lambda = \lambda_1$, so that *at least* s linearly independent characteristic vectors corresponding to λ_1 exist. However, the rank also cannot be *less* than $n - s$, for if this were so, more than s linearly independent characteristic vectors would correspond to λ_1, in which case the total number of linearly independent characteristic vectors corresponding to *all* characteristic numbers would exceed the dimension of n-space.

Thus we deduce the desired result:

If λ_1 *is a characteristic number of multiplicity* s, *of a real symmetric matrix* $\mathbf{A}$ *of order* n, *then the rank of the matrix* $\mathbf{A} - \lambda\mathbf{I}$ *is exactly* $n - s$ *when* $\lambda = \lambda_1$; *that is, there exist exactly* s *linearly independent corresponding characteristic vectors.*

This statement does not apply *in general* to a *nonsymmetric* matrix, as was shown by an example in Section 1.12. However, an argument analogous to that given above shows that *the same statement applies to Hermitian matrices.*

1.19. Definite forms. If the real quadratic form $\mathbf{x}^T\mathbf{A}\mathbf{x}$, associated with a real symmetric matrix $\mathbf{A}$, is *nonnegative* for all real values of the variables x_i, and is *zero* only if *each* of those n variables is zero, then that quadratic

form is said to be *positive definite*. It is then conventional to say also that the *matrix* **A** is positive definite.

Similarly, a *Hermitian* matrix **H** is said to be positive definite if the associated *Hermitian* form $\bar{\mathbf{x}}^T \mathbf{H}\,\mathbf{x}$ is nonnegative for any real or complex vector **x**, and vanishes only when $\mathbf{x} = \mathbf{0}$.

If a real quadratic form $A = \mathbf{x}^T \mathbf{A}\,\mathbf{x}$ is reducible by a transformation of the form $\mathbf{x} = \mathbf{Q}\,\mathbf{x}'$, where **Q** is a nonsingular real matrix, to the sum of *squares* of the n new variables, each with a *positive* coefficient, then it is clear that A is a positive definite form relative to the real variables $x_1', \cdots, x_n'$. But from the relation $\mathbf{x}' = \mathbf{Q}^{-1}\,\mathbf{x}$, which is a consequence of the assumed nonvanishing of $|\mathbf{Q}|$, we see that a real vector **x** then corresponds always to a real vector $\mathbf{x}'$, and that the vectors $\mathbf{x} = \mathbf{0}$ and $\mathbf{x}' = \mathbf{0}$ then correspond uniquely. Hence it follows in this case that A is also positive definite relative to the *original* real variables $x_1, \cdots, x_n$.

Similarly, if a Hermitian form is reducible by a nonsingular complex transformation to the canonical form (145), wherein all coefficients are positive, the form is then nonnegative for any *complex* values of the variables, and is zero if and only if all the n variables vanish.

We notice that if the coefficients of the squares of any of the n variables are zero, then the vanishing of the form does not imply the vanishing of *those* variables, and hence the form is then *not* positive definite relative to the entire set of n variables.

It then follows from the results of preceding sections [see equations (127) and (145)] that *a real quadratic (or Hermitian) form is positive definite if and only if the characteristic numbers of the associated real symmetric (or Hermitian) matrix are all positive.*

A form is often said to be *positive semidefinite* when it takes on only *nonnegative* values for all permissible values of the variables, but *vanishes* for *some* nonzero values of the variables. The preceding argument leads easily to the fact that *a quadratic (or Hermitian) form is positive semidefinite if and only if the associated symmetric (or Hermitian) matrix is singular and possesses no negative characteristic numbers.*

Positive definite forms are of particular importance in applications, and are found to possess certain useful properties. In particular, we show next that if at least one of the *two* real quadratic forms

$$A = \mathbf{x}^T \mathbf{A}\,\mathbf{x}, \qquad B = \mathbf{x}^T \mathbf{B}\,\mathbf{x} \tag{158a,b}$$

is *positive definite*, then it is always possible to reduce the two forms *simultaneously* to linear combinations of only squares of new variables, that is, to canonical forms, by a nonsingular real transformation. For this purpose, suppose that the form B is positive definite. Then, by proceeding exactly as in Section 1.14, we first set

$$\mathbf{x} = \mathbf{Q}\,\mathbf{y}, \tag{159}$$

where $\mathbf{Q}$ is an *orthonormal modal matrix* of $\mathbf{B}$, defined in that section, and so reduce B to the form

$$B = \mu_1 y_1^2 + \mu_2 y_2^2 + \cdots + \mu_n y_n^2, \tag{160}$$

where here μ_i is written for the ith characteristic number of the symmetric matrix $\mathbf{B}$. Since B is positive definite, the μ's are all positive. Hence we may make the real substitution

$$\eta_i = \sqrt{\mu_i}\, y_i \qquad (i = 1, 2, \cdots, n), \tag{161}$$

and thus reduce (160) to the form

$$B = \eta_1^2 + \eta_2^2 + \cdots + \eta_n^2 = \boldsymbol{\eta}^T \boldsymbol{\eta}. \tag{162}$$

At the same time, the substitution (159) reduces A to the form

$$A = (\mathbf{Q}\,\mathbf{y})^T \mathbf{A}\,\mathbf{Q}\,\mathbf{y} = \mathbf{y}^T (\mathbf{Q}^T \mathbf{A}\,\mathbf{Q})\,\mathbf{y} \tag{163}$$

and the subsequent substitution (161) reduces this form to the expression

$$A = \boldsymbol{\eta}^T (\mathbf{Q}'^T \mathbf{A}\,\mathbf{Q}')\boldsymbol{\eta}, \tag{164}$$

where $\mathbf{Q}'$ is a matrix obtained from $\mathbf{Q}$ by dividing each element of the ith column of $\mathbf{Q}$ by $\sqrt{\mu_i}$. Hence, if we write

$$\mathbf{G} = \mathbf{Q}'^T \mathbf{A}\,\mathbf{Q}', \tag{165}$$

equation (164) takes the form

$$A = \boldsymbol{\eta}^T \mathbf{G}\,\boldsymbol{\eta}. \tag{166}$$

Now $\mathbf{G}$ is a symmetric matrix, since

$$\mathbf{G}^T = (\mathbf{Q}'^T \mathbf{A}\,\mathbf{Q}')^T = \mathbf{Q}'^T \mathbf{A}^T \mathbf{Q}' = \mathbf{Q}'^T \mathbf{A}\,\mathbf{Q}' = \mathbf{G}. \tag{167}$$

Hence we may reduce (166) to canonical form by setting

$$\boldsymbol{\eta} = \mathbf{R}\,\boldsymbol{\alpha}, \tag{168}$$

where $\mathbf{R}$ is made up of the characteristic vectors of $\mathbf{G}$ just as $\mathbf{Q}$ is formed from those of $\mathbf{B}$, and (166) is reduced to the form

$$A = \lambda_1 \alpha_1^2 + \lambda_2 \alpha_2^2 + \cdots + \lambda_n \alpha_n^2 \tag{169}$$

where λ_i is the ith characteristic number of the matrix $\mathbf{G}$.

At the same time, the final substitution (168) reduces (162) to

$$B = \boldsymbol{\eta}^T \boldsymbol{\eta} = (\mathbf{R}\,\boldsymbol{\alpha})^T (\mathbf{R}\,\boldsymbol{\alpha}) = \boldsymbol{\alpha}^T \mathbf{R}^T \mathbf{R}\,\boldsymbol{\alpha}. \tag{170}$$

But since the matrix $\mathbf{R}$ is an *orthogonal* matrix, there follows $\mathbf{R}^T \mathbf{R} = \mathbf{I}$, and hence we have the result

$$B = \boldsymbol{\alpha}^T \boldsymbol{\alpha} = \alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2. \tag{171}$$

Thus, finally, with the substitution

$$\mathbf{x} = \mathbf{Q}\,\mathbf{y} = \mathbf{Q}'\,\boldsymbol{\eta} = \mathbf{Q}'\,\mathbf{R}\,\boldsymbol{\alpha}, \tag{172}$$

the two forms (158a,b) are simultaneously reduced to the canonical forms (169) and (171).

If we define the diagonal matrix

$$\mathbf{M} = \begin{bmatrix} m_1 & 0 & \cdots & 0 \\ 0 & m_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & m_n \end{bmatrix}, \qquad m_i = \frac{1}{\sqrt{\mu_i}}, \tag{173}$$

it follows that

$$\mathbf{Q}' = \mathbf{Q}\,\mathbf{M} \tag{174}$$

and (172) becomes

$$\mathbf{x} = \mathbf{Q}\,\mathbf{M}\,\mathbf{R}\,\boldsymbol{\alpha}. \tag{175}$$

Since $\mathbf{Q}$ and $\mathbf{R}$ are orthogonal matrices, with determinants equal to unity in absolute value [see equation (124)], and since clearly $|\mathbf{M}| \neq 0$, it follows that the transformation (172) is indeed nonsingular.

In certain applications to dynamical problems (see Section 2.14) the positive definite form B (*kinetic energy*) involves the time derivative $d\mathbf{x}/dt$ in place of $\mathbf{x}$, whereas the form A (*potential energy*) involves only $\mathbf{x}$ itself. The above reduction is still applicable, however, since $\mathbf{x}$ and $d\mathbf{x}/dt$ are transformed in the same way at each step of the process.

Another method of accomplishing the same reduction, which is usually more conveniently applied in practice, is presented in Section 1.25 [see equations (265)–(267)].

1.20. Discriminants and invariants. It is frequently of importance to determine whether a quadratic or Hermitian form which involves cross-product terms is or is not a *positive definite* form, without reducing it to a canonical form or determining the characteristic numbers of the associated matrix. This problem is to be considered in the present section.

If we write the characteristic equation $|\mathbf{A} - \lambda\,\mathbf{I}| = 0$ of a square matrix $\mathbf{A}$ in the form

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda \end{vmatrix}$$

$$\equiv (-1)^n[\lambda^n - \beta_1\lambda^{n-1} + \beta_2\lambda^{n-2} - \cdots + (-1)^n\beta_n] = 0, \tag{176}$$

and denote the n roots of this equation as $\lambda_1, \lambda_2, \cdots, \lambda_n$, numbering multiple roots separately, it follows that

$$\lambda^n - \beta_1\lambda^{n-1} + \beta_2\lambda^{n-2} - \cdots + (-1)^n\beta_n \equiv (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n). \tag{177}$$

By comparing coefficients of λ in the two sides of (177) it can be shown that

$$\left.\begin{aligned} \beta_1 &= \lambda_1 + \lambda_2 + \cdots + \lambda_n, \\ \beta_2 &= \lambda_1\lambda_2 + \lambda_1\lambda_3 + \cdots + \lambda_{n-1}\lambda_n, \\ \beta_3 &= \lambda_1\lambda_2\lambda_3 + \cdots + \lambda_{n-2}\lambda_{n-1}\lambda_n, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \beta_n &= \lambda_1\lambda_2\lambda_3\cdots\lambda_n \end{aligned}\right\}. \tag{178}$$

Now, for either a *real symmetric* or a *Hermitian* matrix, we have shown that the roots of (176) are all *real*. Hence, by Descartes' rule of signs, we see that in such cases *the roots of the characteristic equation* (176) *are all positive if and only if the quantities* $\beta_1, \beta_2, \cdots, \beta_n$ *are all positive.*

From (176) it follows that β_n is the value of $|\mathbf{A} - \lambda\,\mathbf{I}|$ when $\lambda = 0$; that is, β_n is the value of the determinant of $\mathbf{A}$:

$$\beta_n = |a_{ij}|. \tag{179}$$

Further, it is easily seen that the coefficient of λ^{n-1} in the expansion of the determinant in (176) is merely

$$(-1)^{n+1}(a_{11} + a_{22} + \cdots + a_{nn});$$

that is, β_1 is the sum of the *diagonal elements* of $\mathbf{A}$:

$$\beta_1 = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{k=1}^{n} a_{kk}. \tag{180}$$

This sum is called the *trace* of $\mathbf{A}$.

More generally, it can be shown that β_i *is the sum of all determinants formed from square arrays of order* i *whose principal diagonals lie along the principal diagonal of* $\mathbf{A}$. Such determinants are called the *principal minors* of $\mathbf{A}$.

Thus it follows that *a real quadratic (or Hermitian) form is positive definite if and only if the sums* β_i, *relevant to the associated symmetric (or Hermitian) matrix, are all positive.*

In illustration, the real quadratic form

$$A = a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + 2a_{12}x_1x_2 + 2a_{23}x_2x_3 + 2a_{13}x_1x_3 \tag{181}$$

in three dimensions, which is associated with the real matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}, \tag{182}$$

is positive definite if and only if the three conditions

$$a_{11} + a_{22} + a_{33} > 0, \tag{183a}$$

$$(a_{11}a_{22} - a_{12}^2) + (a_{22}a_{33} - a_{23}^2) + (a_{11}a_{33} - a_{13}^2) > 0, \tag{183b}$$

$$|a_{ij}| > 0, \tag{183c}$$

are satisfied.

It is readily verified by direct expansion that the determinant of the *symmetric* matrix (182) can be written in the form

$$|a_{ij}| = \frac{(a_{11}a_{22} - a_{12}^2)(a_{11}a_{33} - a_{13}^2) - (a_{11}a_{23} - a_{12}a_{13})^2}{a_{11}}, \tag{184}$$

and also in two further equivalent forms obtained by cyclic permutation of the subscripts.

Suppose that we require only that

$$a_{11} > 0, \qquad a_{11}a_{22} - a_{12}^2 > 0, \qquad |a_{ij}| > 0. \tag{185a,b,c}$$

It then follows from (185a,b) that we must have $a_{22} > 0$, and also, by referring to (184), we see that (185a,b,c) imply that $a_{11}a_{33} - a_{13}^2 > 0$. This relation, together with (185a), in turn implies that $a_{33} > 0$. By considering the permutation of the right-hand member of (184) in which $1 \to 2$, $2 \to 3$, $3 \to 1$, we then deduce similarly that (185a,b,c) also imply the inequality $a_{22}a_{33} - a_{23}^2 > 0$. Thus it follows that *the three conditions* (185) *imply the three conditions* (183).

By considering the conditions that (181) still be positive definite when, first, only *one* variable differs from zero and when, second, only *two* variables differ from zero, it is easily seen that *each* diagonal term a_{ii} must be positive and also that *each* principal minor of second order must be positive. Hence these conditions imply and must be implied by either the conditions (183) or the more convenient conditions (185).

More generally, if for *any* real symmetric (or Hermitian) matrix **A** we define the mth *discriminant* Δ_m to be the determinant of the submatrix $\mathbf{D}_m$ obtained by deleting all elements which do not simultaneously lie in the first m rows and columns of **A**, it can be shown that *the real symmetric (or Hermitian) matrix* **A**, *and the corresponding quadratic (or Hermitian) form, is positive definite if and only if each of the n discriminants* Δ_m *is positive.* If and only if this is so, *all* the principal minors of **A** are positive.

To establish the sufficiency of this criterion, we need only prove that, if $\mathbf{D}_m$ is positive definite and $\Delta_{m+1} \equiv |\mathbf{D}_{m+1}|$ is positive, then $\mathbf{D}_{m+1}$ is also positive definite. Suppose, on the contrary, that $\mathbf{D}_{m+1}$ is *not* positive definite. Then, since $|\mathbf{D}_{m+1}|$ is the *product* of the characteristic numbers of $\mathbf{D}_{m+1}$, it follows that an *even* number of these characteristic numbers must be negative. Let γ_1 and γ_2 be two such numbers, and denote by $\mathbf{u}_1$ and $\mathbf{u}_2$ corresponding orthogonal unit characteristic vectors of $\mathbf{D}_{m+1}$, length and orthogonality being defined in the Hermitian sense. If we define the $(m+1)$-dimensional vector

$$\mathbf{x}^* = c_1\mathbf{u}_1 + c_2\mathbf{u}_2,$$

where at least one of the c's does not vanish, and notice that then we have $\mathbf{D}_{m+1}\,\mathbf{u}_1 = \gamma_1\mathbf{u}_1$ and $\mathbf{D}_{m+1}\,\mathbf{u}_2 = \gamma_2\mathbf{u}_2$, there follows easily

$$\bar{\mathbf{x}}^{*T}\,\mathbf{D}_{m+1}\,\mathbf{x}^* = \bar{c}_1 c_1\gamma_1 + \bar{c}_2 c_2\gamma_2 < 0,$$

for any c_1 and c_2. Thus the vector $\mathbf{x}^*$ renders the Hermitian form associated with $\mathbf{D}_{m+1}$ negative. Now let c_1 and c_2 be related in such a way that the component x^*_{m+1} vanishes. If we notice that the Hermitian form $\bar{\mathbf{x}}^T\,\mathbf{D}_{m+1}\,\mathbf{x}$ reduces to the form $\bar{\mathbf{x}}^T\,\mathbf{D}_m\,\mathbf{x}$ when $x_{m+1} = 0$, we conclude that the m-dimensional vector made up of the first m components of the $\mathbf{x}^*$ so determined renders the Hermitian form associated with $\mathbf{D}_m$ negative. Since $\mathbf{D}_m$ is positive definite, this situation is impossible, and the desired contradiction is obtained.

The specialization of the preceding argument to the case of a real symmetric matrix, and its associated real quadratic form, is obtained by deleting the bars indicating complex conjugates. In this case, $\mathbf{u}_1$ and $\mathbf{u}_2$ are real, and the constants c_1 and c_2 are also to be real.

Whereas the requirements that a form or matrix be positive definite thus need not be stated in terms of the *sums* β_i, these sums nevertheless are of considerable importance in themselves. We see from (178) that each β_i is a symmetric function, of degree i, of the characteristic numbers of $\mathbf{A}$. Also, it follows from (176) that *for any two square matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *such that* $|\mathbf{A} - \lambda\mathbf{I}| = |\mathbf{B} - \lambda\mathbf{I}|$ *for all values of* λ, *the* n *quantities* β_i *are the same.*

In order to determine conditions under which this situation exists, let $\mathbf{A}$ and $\mathbf{B}$ be two equivalent matrices. This means that nonsingular matrices $\mathbf{P}$ and $\mathbf{Q}$ exist such that $\mathbf{B} = \mathbf{P}\,\mathbf{A}\,\mathbf{Q}$. Hence we have, for any value of λ,

$$\begin{aligned}\mathbf{B} - \lambda\mathbf{I} &= \mathbf{P}\,\mathbf{A}\,\mathbf{Q} - \lambda\mathbf{I} \\ &= \mathbf{P}(\mathbf{A} - \lambda\,\mathbf{P}^{-1}\,\mathbf{Q}^{-1})\mathbf{Q}\end{aligned}$$

and also

$$|\mathbf{B} - \lambda\mathbf{I}| = |\mathbf{P}|\,|\mathbf{Q}|\,|\mathbf{A} - \lambda\,\mathbf{P}^{-1}\,\mathbf{Q}^{-1}|. \tag{186}$$

Thus, if $\mathbf{P}$ and $\mathbf{Q}$ are such that $\mathbf{P}^{-1}\,\mathbf{Q}^{-1} = \mathbf{I}$ or $\mathbf{P} = \mathbf{Q}^{-1}$, so that

$$\mathbf{B} = \mathbf{Q}^{-1}\,\mathbf{A}\,\mathbf{Q}, \tag{187}$$

there follows $\mathbf{P}\,\mathbf{Q} = \mathbf{I}$, and hence $|\mathbf{P}|\,|\mathbf{Q}| = 1$, and (186) takes the form

$$|\mathbf{B} - \lambda\,\mathbf{I}| = |\mathbf{A} - \lambda\,\mathbf{I}|, \tag{188}$$

for all values of λ.

A transformation of the form (187) has been defined as a *similarity transformation*, and the matrices **A** and **B** are said to be *similar*. Since (188) states that **A** and **B** have the same characteristic equation, it follows that *the quantities β_i are invariant under* (unchanged by) *any similarity transformation.* This result has important consequences in many physical considerations.

Since *orthogonal* and *unitary* transformations are special types of similarity transformations, in which also $\mathbf{Q}^{-1} = \mathbf{Q}^T$ and $\mathbf{Q}^{-1} = \overline{\mathbf{Q}}^T$, respectively, the preceding statement applies to them.

1.21. Coordinate transformations. The elements of an n-dimensional *numerical vector* $\mathbf{x} = \{x_1, x_2, \cdots, x_n\}$ may be interpreted as the *components* of a certain *geometrical vector* $\mathscr{X}$ (such as a *force* or an *acceleration* or a *displacement* from an origin to a point in a geometrical n-space) in the directions of the n basic mutually orthogonal unit vectors

$$\mathbf{i}_1 = \{1, 0, \cdots, 0\}, \qquad \cdots, \qquad \mathbf{i}_n = \{0, 0, \cdots, 1\}$$

which lie along the axes of reference rectangular coordinates $x_1, x_2, \cdots, x_n$ in n-space. This interpretation corresponds to the relationship*

$$\mathbf{x} = \{x_1, x_2, \cdots, x_n\} = \sum_{k=1}^{n} x_k \mathbf{i}_k. \tag{189}$$

Now let a *new* coordinate system in n-space be so chosen that the unit vectors $\mathbf{i}'_1, \mathbf{i}'_2, \cdots, \mathbf{i}'_n$ in the directions of the axes of the new coordinates $x'_1, x'_2, \cdots, x'_n$ are linearly independent and are related to the original unit vectors by the equations

$$\left.\begin{aligned} \mathbf{i}'_1 &= q_{11}\mathbf{i}_1 + q_{21}\mathbf{i}_2 + \cdots + q_{n1}\mathbf{i}_n, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \mathbf{i}'_n &= q_{1n}\mathbf{i}_1 + q_{2n}\mathbf{i}_2 + \cdots + q_{nn}\mathbf{i}_n \end{aligned}\right\} \tag{190}$$

or

$$\mathbf{i}'_k = \sum_{r=1}^{n} q_{rk}\mathbf{i}_r. \tag{191}$$

The geometrical vector $\mathscr{X}$ then can be specified by its components $x'_1, x'_2, \cdots, x'_n$ along the new axes. If we denote the numerical vector comprising this array of components by $\mathbf{x}'$, there follows

$$\mathbf{x}' = x'_1\mathbf{i}'_1 + x'_2\mathbf{i}'_2 + \cdots + x'_n\mathbf{i}'_n = \sum_{k=1}^{n} x'_k\mathbf{i}'_k. \tag{192}$$

* In nongeometric terms, we can say that **x** is the *numerical vector* whose elements are the constants of combination in the representation (189) of an *abstract vector* $\mathscr{X}$ in terms of the *standard basis* $\mathbf{i}_1, \mathbf{i}_2, \cdots, \mathbf{i}_n$.

To determine the new components in terms of the original ones, we transform the representation $\mathbf{x}'$ to the representation $\mathbf{x}$ by introducing (191) into (192):

$$\mathbf{x} = \sum_{k=1}^{n} \sum_{r=1}^{n} x'_k q_{rk} \mathbf{i}_r = \sum_{r=1}^{n} \left(\sum_{k=1}^{n} q_{rk} x'_k \right) \mathbf{i}_r. \tag{193}$$

Then, since the vectors $\mathbf{i}_r$ are mutually orthogonal, their respective coefficients in (189) and (193) must be equal, so that

$$x_r = \sum_{k=1}^{n} q_{rk} x'_k. \tag{194}$$

Thus, if we write $\mathbf{x}' = \{x'_1, x'_2, \cdots, x'_n\}$ for the numerical vector comprising the components of $\mathscr{X}$ in the directions of the new coordinate axes specified by (190), there follows

$$\mathbf{x} = \mathbf{Q}\,\mathbf{x}', \tag{195}$$

where $\mathbf{Q}$ is the *transformation matrix*

$$\mathbf{Q} = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ q_{n1} & q_{n2} & \cdots & q_{nn} \end{bmatrix}, \tag{196}$$

of which the coefficient matrix in (190) is the *transpose*. We notice that each *column* of (196) contains *the components of a new unit vector along the original coordinate axes.*

Here we interpret the matrix $\mathbf{Q}$ of (195) as relating the components of a geometrical vector along the original coordinate axes to the components of the *same* geometrical vector along the new coordinate axes. In other considerations we may suppose that no *change of axes* is involved, and that an equation of the form (195) merely transforms one numerical vector into another one, both vectors then being referred to the same axes. Which interpretation is to be attached to such an equation in practice clearly depends upon the nature of the problem involved.*

In order that the new unit vectors be linearly independent, and hence span n-space, the matrix $\mathbf{Q}$ must be nonsingular. Hence $\mathbf{Q}^{-1}$ then exists,

* In many other applications the two number sets $\{x_1, \cdots, x_n\}$ and $\{x'_1, \cdots, x'_n\}$ are not inherently subject to *any* geometrical interpretations (for example, they could represent the costs of n manufactured items under two different production schedules) and (195) is merely a compact way of stating that the two sets of numbers are related by a certain set of linear algebraic equations. In *abstract* terms, (195) relates the components of an abstract vector $\mathscr{X}$, relative to the *standard* basis $\mathbf{i}_1, \cdots, \mathbf{i}_n$, to the components of $\mathscr{X}$, relative to a *new* basis $\mathbf{i}'_1, \cdots, \mathbf{i}'_n$, when the two bases are related by (191).

and we have also, from (195),

$$\mathbf{x}' = \mathbf{Q}^{-1}\,\mathbf{x}. \tag{197}$$

Suppose now that two numerical vectors $\mathbf{x}$ and $\mathbf{y}$, representing $\mathscr{X}$ and $\mathscr{Y}$, respectively, are related by an equation of the form

$$\mathbf{y} = \mathbf{A}\,\mathbf{x}, \tag{198}$$

when the components refer to the original coordinate frame, and that we require the relationship between the two corresponding numerical vectors $\mathbf{x}'$ and $\mathbf{y}'$ whose components are referred to a new coordinate frame (190). (We may, for example, imagine that $\mathbf{y}$ and $\mathbf{y}'$ represent a *force* and that $\mathbf{x}$ and $\mathbf{x}'$ represent an *acceleration*. In *Newtonian* mechanics, the matrix $\mathbf{A}$ then would be a *scalar* matrix.) By replacing $\mathbf{x}$ by $\mathbf{Q}\,\mathbf{x}'$ and $\mathbf{y}$ by $\mathbf{Q}\,\mathbf{y}'$, we obtain the relation

$$\mathbf{Q}\,\mathbf{y}' = \mathbf{A}\,\mathbf{Q}\,\mathbf{x}'$$

from (198) and hence, after premultiplying both sides by $\mathbf{Q}^{-1}$, we deduce the desired result

$$\mathbf{y}' = (\mathbf{Q}^{-1}\,\mathbf{A}\,\mathbf{Q})\,\mathbf{x}'. \tag{199}$$

Thus we see that the matrix relating $\mathbf{x}'$ and $\mathbf{y}'$ is obtained from that relating $\mathbf{x}$ and $\mathbf{y}$ by a *similarity transformation*. In particular, it follows that the invariance properties discussed at the close of the preceding section apply in the present case. That is, the quantities β_i of that section, pertaining to the matrix $\mathbf{A}$ of (198), are invariant under a nonsingular coordinate transformation. This result is of great importance in many applications.

If the *new* unit vectors are *mutually orthogonal*, we readily obtain from (196) the result

$$\mathbf{Q}^T\,\mathbf{Q} = \mathbf{I} \qquad or \qquad \mathbf{Q}^T = \mathbf{Q}^{-1}, \tag{200}$$

so that $\mathbf{Q}$ then is an *orthogonal* matrix. Thus, *a transformation from one set of orthogonal axes to another is accomplished by an orthogonal transformation.* We may verify that in such a transformation the length of the vector representation in the new system is the same as the length of the representation in the original system (that is, there is no change in *scale*). For if $\mathbf{x} = \mathbf{Q}\,\mathbf{x}'$ there follows

$$l^2 = \mathbf{x}^T\,\mathbf{x} = (\mathbf{Q}\,\mathbf{x}')^T\,\mathbf{Q}\,\mathbf{x}' = \mathbf{x}'^T\,\mathbf{Q}^T\,\mathbf{Q}\,\mathbf{x}' = \mathbf{x}'^T\,\mathbf{x}' = l'^2. \tag{201}$$

Also, the magnitude of the scalar product of the numerical vectors representing two quantities is the same in both systems (that is, the magnitude of an "angle" is also preserved), since

$$(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T\,\mathbf{y} = \mathbf{x}'^T\,\mathbf{Q}^T\,\mathbf{Q}\,\mathbf{y}' = \mathbf{x}'^T\,\mathbf{y}' = (\mathbf{x}', \mathbf{y}'). \tag{202}$$

As these results suggest, it can be shown that any orthogonal transformation in n-space can be interpreted as a combination of *rotations* and *reflections.**

In particular, reference to Section 1.14 shows that, when $\mathbf{A}$ is an $n \times n$ real symmetric matrix, the real quadratic form

$$A = \mathbf{x}^T \mathbf{A}\, \mathbf{x} \tag{203}$$

in the real variables $x_1, \cdots, x_n$ can be reduced to the canonical form

$$A = \lambda_1 x_1'^2 + \cdots + \lambda_n x_n'^2 \tag{204}$$

by a real *coordinate transformation*, comprising only rotations and/or reflections, in which *the axes of the new coordinates* $x_1', \cdots, x_n'$ *are made to coincide with the respective mutually orthogonal characteristic vectors of the matrix* $\mathbf{A}$.

1.22. Functions of symmetric matrices. In this section, we again restrict attention to real *symmetric* matrices, which are of principal interest in applications. We notice first that, as is easily shown, *the sum of two symmetric matrices of the same order is also symmetric*, while *the product of two symmetric matrices of the same order is symmetric if those matrices are commutative.*

Positive integral powers of any square matrix $\mathbf{A}$ are defined by iteration:

$$\mathbf{A}^2 = \mathbf{A}\,\mathbf{A}, \quad \mathbf{A}^3 = \mathbf{A}\,\mathbf{A}^2, \quad \cdots, \quad \mathbf{A}^{n+1} = \mathbf{A}\,\mathbf{A}^n, \quad \cdots. \tag{205}$$

In consequence of this definition, there follows also

$$\mathbf{A}^r\,\mathbf{A}^s = \mathbf{A}^s\,\mathbf{A}^r = \mathbf{A}^{r+s}, \tag{206}$$

when r and s are positive integers. *Negative integral* powers are defined only for *nonsingular* matrices, for which a unique inverse $\mathbf{A}^{-1}$ exists, and are then defined by the relation

$$\mathbf{A}^{-n} = (\mathbf{A}^{-1})^n. \tag{207}$$

If we define also

$$\mathbf{A}^0 = \mathbf{I}, \tag{208}$$

then (206) applies to any nonsingular matrix, for *any* integers r and s. It is clear that *any integral power of a symmetric matrix is also symmetric.*

Polynomial functions of $\mathbf{A}$ are then defined as linear combinations of a finite number of nonnegative integral powers of $\mathbf{A}$. Any polynomial in $\mathbf{A}$ hence can be expressed as a symmetric matrix of the same order as $\mathbf{A}$.

* When two real vectors x and y undergo the *same real transformation*, so that $\mathbf{x} = \mathbf{Q}\,\mathbf{x}'$ and $\mathbf{y} = \mathbf{Q}\,\mathbf{y}'$, the two sets of variables which comprise their components are said to be *cogredient*. When the vectors are transformed separately, so that $\mathbf{x} = \mathbf{P}\,\mathbf{x}'$ and $\mathbf{y} = \mathbf{Q}\,\mathbf{y}'$, in such a way that the condition $(\mathbf{x}, \mathbf{y}) = (\mathbf{x}', \mathbf{y}')$ is satisfied, the two sets of variables are said to be *contragredient*. Equation (202) states that when two vectors undergo the same *orthogonal* transformation their components are *both cogredient and contragredient.*

Suppose now that $\mathbf{A}$ is of order n, and let its characteristic numbers be denoted by $\lambda_1, \lambda_2, \cdots, \lambda_n$ (not necessarily distinct), with corresponding orthogonalized characteristic vectors $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$, so that

$$\mathbf{A}\,\mathbf{u}_i = \lambda_i \mathbf{u}_i \tag{209}$$

for $i = 1, 2, \cdots, n$. If we multiply both sides of (209) by $\mathbf{A}$, and use (209) to simplify the resulting right-hand member, there follows

$$\mathbf{A}^2\,\mathbf{u}_i = \lambda_i \mathbf{A}\,\mathbf{u}_i = \lambda_i^2 \mathbf{u}_i, \tag{210}$$

and, by repeating this process, we deduce from (209) the relation

$$\mathbf{A}^r\,\mathbf{u}_i = \lambda_i^r \mathbf{u}_i \tag{211}$$

for any positive integer r. Similarly, if $\mathbf{A}$ is *nonsingular*, the result of multiplying both sides of (209) by $\mathbf{A}^{-1}$ becomes

$$\mathbf{A}^{-1}\,\mathbf{u}_i = \lambda_i^{-1} \mathbf{u}_i \tag{212}$$

and, by iteration, we find that (211) is then true for any integer r.

Thus we deduce that *if λ_i is a characteristic number of* $\mathbf{A}$, *with a corresponding characteristic vector* $\mathbf{u}_i$, *then λ_i^r is a characteristic number of* $\mathbf{A}^r$, *with the same characteristic vector* $\mathbf{u}_i$. For a *symmetric* matrix of order n, there are exactly n linearly independent characteristic vectors. Hence it follows in this case that $\mathbf{A}^r$ cannot possess *additional* characteristic numbers or linearly independent characteristic vectors.

It should be noticed, however, that $\mathbf{A}^r$ *may* have characteristic vectors which are *not* possessed by $\mathbf{A}$. As a simple example, it is seen that the matrix

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

has the characteristic numbers $\lambda_1 = 1$ and $\lambda_2 = -1$, and corresponding characteristic vectors $\mathbf{u}_1 = \{1, 0\}$ and $\mathbf{u}_2 = \{0, 1\}$. Here the matrix $\mathbf{A}^2 = \mathbf{I}$ has the *double* characteristic number $\lambda_1^2 = \lambda_2^2 = 1$ and it is obvious that *all* nonzero vectors in two-space are characteristic vectors of $\mathbf{A}^2$. Clearly, all such vectors are indeed linear combinations of $\mathbf{u}_1$ and $\mathbf{u}_2$.

Next, consider any polynomial in $\mathbf{A}$, of degree m, of the form

$$P(\mathbf{A}) = \alpha_0 \mathbf{A}^m + \alpha_1 \mathbf{A}^{m-1} + \cdots + \alpha_{m-1}\mathbf{A} + \alpha_m \mathbf{I}. \tag{213}$$

If we consider the product of the matrix $P(\mathbf{A})$ with any characteristic vector of $\mathbf{A}$, and use (211), we obtain the relation

$$P(\mathbf{A})\,\mathbf{u}_i = \alpha_0 \lambda_i^m \mathbf{u}_i + \alpha_1 \lambda_i^{m-1} \mathbf{u}_i + \cdots + \alpha_{m-1}\lambda_i \mathbf{u}_i + \alpha_m \mathbf{u}_i$$

or

$$P(\mathbf{A})\,\mathbf{u}_i = P(\lambda_i)\,\mathbf{u}_i. \tag{214}$$

Hence it follows that the equation

$$[P(\mathbf{A}) - \mu\,\mathbf{I}]\,\mathbf{x} = \mathbf{0} \tag{215}$$

possesses a nontrivial solution when $\mu = P(\lambda_i)$, and that a solution of (215) in this case is an arbitrary multiple of $\mathbf{u}_i$. But, as in the preceding argument, no additional linearly independent solutions of (215) can exist. Thus we find that *if* $\mathbf{A}$ *is symmetric, then all characteristic vectors of* $\mathbf{A}$ *also belong to* $P(\mathbf{A})$, and also, *if the characteristic numbers of* $\mathbf{A}$ *are* $\lambda_1, \cdots, \lambda_n$, *then those of* $P(\mathbf{A})$ *are* $P(\lambda_1), \cdots, P(\lambda_n)$.

Let the λ-polynomial $|\mathbf{A} - \lambda\,\mathbf{I}|$, the vanishing of which determines the characteristic numbers of $\mathbf{A}$, be denoted by $F(\lambda)$:

$$F(\lambda) = |\mathbf{A} - \lambda\,\mathbf{I}| \equiv (-1)^n[\lambda^n - \beta_1\lambda^{n-1} + \cdots + (-1)^n\beta_n]. \tag{216}$$

Then $F(\lambda)$ is a polynomial in λ, of degree n, which vanishes when $\lambda = \lambda_i$,

$$F(\lambda_i) = 0 \qquad (i = 1, 2, \cdots, n). \tag{217}$$

If now we identify the polynomial function P with the function F, equation (214) becomes

$$F(\mathbf{A})\,\mathbf{u}_i = \mathbf{0} \qquad (i = 1, 2, \cdots, n). \tag{218}$$

Thus if we write temporarily $\mathbf{B} = F(\mathbf{A})$, it follows that the equation $\mathbf{B}\,\mathbf{x} = \mathbf{0}$ possesses the n linearly independent solutions $\mathbf{x} = \mathbf{u}_1, \cdots, \mathbf{u}_n$. But since $\mathbf{B}$ is a square matrix of order n, the results of Section 1.9 show that $\mathbf{B}$ must be of rank $n - n = 0$. Hence $\mathbf{B} = F(\mathbf{A})$ must be a *zero matrix*, and it follows that*

$$F(\mathbf{A}) = \mathbf{0}. \tag{219}$$

That is, *if the characteristic equation of a symmetric matrix* $\mathbf{A}$ *is* $F(\lambda) = 0$, *then the matrix* $\mathbf{A}$ *satisfies the equation* $F(\mathbf{A}) = \mathbf{0}$.

This curious and useful result is known as the *Cayley-Hamilton theorem*, and is often stated briefly as follows: "A matrix satisfies its own characteristic equation."

It is important to notice that in deducing (219) from (218) we have made use of the fact that a *symmetric* matrix always has n linearly independent characteristic vectors. Since this statement does not apply to nonsymmetric matrices with repeated characteristic numbers, the preceding proof does not apply in such cases. However, it can be proved by somewhat less direct methods that *the Cayley-Hamilton theorem is true for any square matrix.*

* It should be noticed that $F(\mathbf{M})$ is defined to be the *matrix* obtained by replacing λ by $\mathbf{M}$ in the *polynomial* $F(\lambda) = (-1)^n\lambda^n + \cdots$, with the understanding that λ^0 is to be replaced by $\mathbf{I}$ in the constant term. The interpretation $F(\mathbf{M}) = |\mathbf{A} - \mathbf{M}\,\mathbf{I}| = |\mathbf{A} - \mathbf{M}|$ is *not* intended and obviously would *not* lead to (219).

If $F(\lambda)$ possesses a factor $(\lambda - \lambda_r)^s$, where $s > 1$, so that λ_r is of multiplicity s, the same argument shows that the *symmetric* matrix $\mathbf{A}$ also satisfies the *reduced characteristic equation* $G(\mathbf{A}) = \mathbf{0}$, where $G(\lambda) = F(\lambda)/(\lambda - \lambda_r)^{s-1}$. (The matrix $\mathbf{A}$ considered in Section 1.15 may be used as an illustration.) This statement is *not* necessarily true if $\mathbf{A}$ is nonsymmetric, as may be illustrated by the matrix considered on page 32.

As a verification of the theorem, we notice that, for the matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \tag{220}$$

we have

$$F(\lambda) = \begin{vmatrix} 2 - \lambda & 1 \\ 1 & 2 - \lambda \end{vmatrix} = \lambda^2 - 4\lambda + 3, \tag{221}$$

and the equation $\mathbf{A}^2 - 4\mathbf{A} + 3\mathbf{I} = \mathbf{0}$ becomes

$$\begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix} - \begin{bmatrix} 8 & 4 \\ 4 & 8 \end{bmatrix} + \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

We notice that this theorem permits any positive integral power of a matrix $\mathbf{A}$, and hence *any polynomial* in $\mathbf{A}$, to be expressed as a linear combination of the matrices $\mathbf{I}, \mathbf{A}, \mathbf{A}^2, \cdots, \mathbf{A}^{n-1}$, where n is the order of $\mathbf{A}$.

Thus, for the matrix (220) considered above, we have the successive results

$$\begin{aligned} \mathbf{A}^2 &= 4\mathbf{A} - 3\mathbf{I}, \\ \mathbf{A}^3 &= 4\mathbf{A}^2 - 3\mathbf{A} = 4(4\mathbf{A} - 3\mathbf{I}) - 3\mathbf{A} = 13\mathbf{A} - 12\mathbf{I}, \end{aligned} \tag{222}$$

and so forth. In addition, we obtain the relation

$$\mathbf{A} - 4\mathbf{I} + 3\mathbf{A}^{-1} = \mathbf{0}.$$

Hence we deduce that

$$\mathbf{A}^{-1} = -\tfrac{1}{3}\mathbf{A} + \tfrac{4}{3}\mathbf{I},$$

and obtain successive *negative* integral powers of $\mathbf{A}$ by successive multiplications and simplifications.

A convenient determination of the constants of combination, in the case of a general polynomial, is afforded by a result next to be obtained under the assumption that all the characteristic numbers of $\mathbf{A}$ are *distinct*. In place of directly determining the constants involved in the representation $P(\mathbf{A}) = c_1\mathbf{A}^{n-1} + c_2\mathbf{A}^{n-2} + \cdots + c_n\mathbf{I}$, it is desirable for present purposes to assume the equivalent form

$$\begin{aligned} P(\mathbf{A}) = {} & C_1[(\mathbf{A} - \lambda_2\mathbf{I})(\mathbf{A} - \lambda_3\mathbf{I}) \cdots (\mathbf{A} - \lambda_n\mathbf{I})] \\ & + C_2[(\mathbf{A} - \lambda_1\mathbf{I})(\mathbf{A} - \lambda_3\mathbf{I}) \cdots (\mathbf{A} - \lambda_n\mathbf{I})] + \cdots \\ & + C_n[(\mathbf{A} - \lambda_1\mathbf{I})(\mathbf{A} - \lambda_2\mathbf{I}) \cdots (\mathbf{A} - \lambda_{n-1}\mathbf{I})], \end{aligned} \tag{223}$$

where each bracketed quantity, and hence also the complete right-hand side, is clearly a polynomial of degree $n - 1$ in $\mathbf{A}$. To determine the n C's, we postmultiply the equal members of (223) successively by each of the n characteristic vectors $\mathbf{u}_1, \cdots, \mathbf{u}_n$ of the matrix $\mathbf{A}$.

If both members are postmultiplied by $\mathbf{u}_k$, and use is made of the relation $\mathbf{A}\,\mathbf{u}_k = \lambda_k \mathbf{u}_k$, it is found that the coefficients of all C's except C_k then contain the factor $(\lambda_k - \lambda_k)$, and hence vanish. Thus there follows, after a simple calculation,

$$P(\mathbf{A})\,\mathbf{u}_k = C_k[(\lambda_k - \lambda_1)\cdots(\lambda_k - \lambda_{k-1})(\lambda_k - \lambda_{k+1})\cdots(\lambda_k - \lambda_n)]\mathbf{u}_k, \quad (224)$$

for $k = 1, 2, \cdots, n$. But reference to equation (214) then shows that the coefficient of $\mathbf{u}_k$ on the right must be equal to $P(\lambda_k)$. Thus *if the characteristic numbers of the matrix* $\mathbf{A}$ *are all distinct*, we obtain the result

$$C_k = \frac{P(\lambda_k)}{\prod_{r \neq k} (\lambda_k - \lambda_r)} \qquad (k = 1, 2, \cdots, n), \quad (225)$$

where the notation $\prod_{r \neq k}$ denotes the product of those factors for which r takes on the values 1 through n, excluding k. If this result is introduced into (223), the desired representation is obtained in the form

$$P(\mathbf{A}) = \sum_{k=1}^{n} P(\lambda_k) Z_k(\mathbf{A}), \quad (226)$$

with the convenient abbreviation

$$Z_k(\mathbf{A}) = \frac{\prod_{r \neq k} (\mathbf{A} - \lambda_r \mathbf{I})}{\prod_{r \neq k} (\lambda_k - \lambda_r)} \qquad (k = 1, 2, \cdots, n). \quad (227)$$

Cases in which certain characteristic numbers are repeated require special treatment.*

To verify this result in the case of the matrix (220), we notice that $\lambda_1 = 3$ and $\lambda_2 = 1$. To evaluate $P(\mathbf{A}) = \mathbf{A}^3$, we first calculate

$$Z_1(\mathbf{A}) = \frac{\mathbf{A} - \lambda_2 \mathbf{I}}{3 - 1} = \frac{1}{2}(\mathbf{A} - \mathbf{I}), \qquad Z_2(\mathbf{A}) = \frac{\mathbf{A} - \lambda_1 \mathbf{I}}{1 - 3} = -\frac{1}{2}(\mathbf{A} - 3\mathbf{I}).$$

Hence, with $P(3) = 27$ and $P(1) = 1$, there follows

$$\mathbf{A}^3 = \tfrac{27}{2}(\mathbf{A} - \mathbf{I}) - \tfrac{1}{2}(\mathbf{A} - 3\mathbf{I}) = 13\mathbf{A} - 12\mathbf{I},$$

in accordance with (222). The usefulness of (226) clearly would be better illustrated in the calculation of $\mathbf{A}^{100}$.

* See Reference 7. It can be shown that this representation (with appropriate modifications for repeated characteristic numbers) is valid for *any* square matrix $\mathbf{A}$.

It should be noticed that the matrices $Z_k(\mathbf{A})$ depend only on $\mathbf{A}$, and are *not* dependent upon the form of the polynomial P chosen. The result (226) is known as *Sylvester's formula.*

Having defined polynomial functions, we may next define other functions of $\mathbf{A}$ by *infinite series* such as

$$\sum_{m=0}^{\infty} \alpha_m \mathbf{A}^m = \lim_{M \to \infty} \sum_{m=0}^{M} \alpha_m \mathbf{A}^m, \tag{228}$$

for those matrices for which the indicated limit exists. We omit discussion of the convergence of such series. However, if $\mathbf{A}$ is of order n, it is clear that the sum of M terms of the series can be expressed as a polynomial of maximum degree $n - 1$ in $\mathbf{A}$, regardless of the value of M, in consequence of the preceding results. Thus we see that *if the series converges, the function represented by the series must also be so expressible, and hence must be determinable from* (226) *if the characteristic numbers of* $\mathbf{A}$ *are distinct.*

In particular, it can be shown that the series in the right-hand member of the relation

$$e^{\mathbf{A}} = \sum_{m=0}^{\infty} \frac{\mathbf{A}^m}{m!} \tag{229}$$

converges for *any* square matrix $\mathbf{A}$, and hence may serve to *define* $e^{\mathbf{A}}$. Suppose that $\mathbf{A}$ is a matrix of order *two*, with distinct characteristic numbers λ_1 and λ_2. Then (227) gives

$$Z_1(\mathbf{A}) = \frac{\mathbf{A} - \lambda_2 \mathbf{I}}{\lambda_1 - \lambda_2}, \qquad Z_2(\mathbf{A}) = \frac{\mathbf{A} - \lambda_1 \mathbf{I}}{\lambda_2 - \lambda_1},$$

and from (226) we obtain the evaluation

$$e^{\mathbf{A}} = \frac{1}{\lambda_1 - \lambda_2} [(e^{\lambda_1} - e^{\lambda_2})\mathbf{A} - (\lambda_2 e^{\lambda_1} - \lambda_1 e^{\lambda_2})\mathbf{I}]. \tag{230}$$

The corresponding evaluation when $\lambda_1 = \lambda_2$ can be obtained from this result as the limiting form when $\lambda_2 \to \lambda_1$ (see Problem 86).

1.23. Numerical solution of characteristic-value problems. In the process of dealing with a characteristic-value problem of the form

$$\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{x}, \tag{231}$$

it is necessary first to determine roots of the characteristic equation

$$|\mathbf{A} - \lambda\,\mathbf{I}| = 0, \tag{232}$$

and then, for each such value of λ, to obtain a nontrivial solution vector of (231). If $\mathbf{A}$ is of order n, equation (232) is an algebraic equation of the same degree in λ, and the numerical determination of the characteristic numbers

generally involves considerable labor when $n > 2$. Further, the actual *expansion* of (232) may be tedious in such cases.

In this section we outline a numerical iterative method which avoids these steps, and which is frequently useful in practice. This method is analogous to a method, associated with the names of *Vianello* and *Stodola*, which is applied to corresponding problems involving differential equations.*

Suppose first that the *dominant* characteristic number, that is, the characteristic number with largest magnitude, is required. To start the procedure, we choose an initial nonzero approximation to the corresponding characteristic vector, say $\mathbf{x}^{(1)}$. In the absence of advance knowledge as to the nature of this vector, we may, for example, start with the vector $\{0, 0, \cdots, 1\}$ or $\{1, 1, \cdots, 1\}$. This initial approximation is then introduced into the *left*-hand member of (231). If we then set

$$\mathbf{y}^{(1)} = \mathbf{A}\,\mathbf{x}^{(1)}, \tag{233}$$

the requirement that (231) be approximately satisfied becomes

$$\mathbf{y}^{(1)} \approx \lambda\,\mathbf{x}^{(1)}. \tag{234}$$

If the respective components of $\mathbf{x}^{(1)}$ and $\mathbf{y}^{(1)}$ are nearly in a constant ratio, we may expect that the approximation $\mathbf{x}^{(1)}$ is good, and that this ratio is an approximation to the true value of λ.

It is rather conventional to choose $\mathbf{x}^{(1)}$ in such a way that one component is *unity*, and to choose, as a first approximation to the dominant characteristic value of λ, the corresponding component of $\mathbf{y}^{(1)}$. A more efficient determination is outlined in the following section [see equations (250a,b)].

A convenient multiple of $\mathbf{y}^{(1)}$ is then taken as the next approximation $\mathbf{x}^{(2)}$, and the process is repeated until satisfactory agreement between successive approximations is obtained. As will be shown, in the case when $\mathbf{A}$ is real and symmetric, this method will lead inevitably to the *dominant* characteristic value of λ and to the corresponding characteristic vector, unless the vector $\mathbf{x}^{(1)}$ happens to be exactly orthogonal to that vector, except in the unusual case when the *negative* of the dominant characteristic number is *also* a characteristic number.

Analytically, if the rth approximation is denoted by $\mathbf{x}^{(r)}$, the iteration can be specified by the relations

$$\mathbf{y}^{(r)} = \mathbf{A}\,\mathbf{x}^{(r)}, \qquad \mathbf{x}^{(r+1)} = \alpha_r \mathbf{y}^{(r)} \qquad (r = 1, 2, \cdots), \tag{235a,b}$$

where α_r is a conveniently chosen nonzero multiplicative constant, and the assertion then is that, in general,

$$\mathbf{y}^{(r)} \sim \lambda_n \mathbf{x}^{(r)}, \qquad \mathbf{x}^{(r)} \to c\,\mathbf{e}_n \qquad (r \to \infty) \tag{236a,b}$$

* See Reference 9.

where λ_n is the characteristic number of $\mathbf{A}$ which is *of largest magnitude*, $\mathbf{e}_n$ is the corresponding unit characteristic vector, and c is a constant depending upon the choice of the α's. The introduction of the arbitrarily chosen α_r in the rth cycle permits one to take the length of $\mathbf{x}^{(r+1)}$ to be of the order of unity, and so to prevent the lengths of successive approximation vectors from growing unboundedly or tending to zero.

If the *smallest* characteristic value of λ is required, we may first transform (231) to the equation

$$\mathbf{x} = \lambda \mathbf{A}^{-1} \mathbf{x}.$$

With the notations

$$\mathbf{M} = \mathbf{A}^{-1}, \qquad \kappa = \frac{1}{\lambda} \tag{237a,b}$$

this equation takes the form

$$\mathbf{M}\mathbf{x} = \kappa \mathbf{x}. \tag{238}$$

The *largest* characteristic value of κ for this equation then can be determined by the iterative method, and its reciprocal is the *smallest* characteristic value of λ for (231).*

Analytically, if the rth approximation is again denoted by $\mathbf{x}^{(r)}$, and if we now denote $\mathbf{M}\mathbf{x}^{(r)} \equiv \mathbf{A}^{-1}\mathbf{x}^{(r)}$ by $\mathbf{y}^{(r)}$, it follows that also $\mathbf{x}^{(r)} = \mathbf{A}\mathbf{y}^{(r)}$, so that we may specify this iteration by the relations

$$\mathbf{x}^{(r)} = \mathbf{A}\mathbf{y}^{(r)}, \qquad \mathbf{x}^{(r+1)} = \alpha_r \mathbf{y}^{(r)} \qquad (r = 1, 2, \cdots) \tag{239a,b}$$

and the assertion in this case is that, in general,

$$\mathbf{y}^{(r)} \sim \frac{1}{\lambda_1}\mathbf{x}^{(r)}, \qquad \mathbf{x}^{(r)} \to c\,\mathbf{e}_1 \qquad (r \to \infty) \tag{240a,b}$$

where λ_1 is the characteristic number of $\mathbf{A}$ *of smallest magnitude* and $\mathbf{e}_1$ is the corresponding unit characteristic vector. As before, we begin by choosing the elements of the vector $\mathbf{x}^{(1)}$, but here we next determine the elements of $\mathbf{y}^{(1)}$ in such a way that $\mathbf{A}\mathbf{y}^{(1)} = \mathbf{x}^{(1)}$, whereas in approximating λ_n we determine $\mathbf{y}^{(1)}$ such that $\mathbf{y}^{(1)} = \mathbf{A}\mathbf{x}^{(1)}$. Clearly, the computation of the elements of the inverse matrix $\mathbf{M} \equiv \mathbf{A}^{-1}$ for the purpose of transforming (231) to (238) may be avoided at the expense of solving the simultaneous equations corresponding to (239a),

$$\left.\begin{aligned} a_{11}y_1^{(r)} + \cdots + a_{1n}y_n^{(r)} &= x_1^{(r)}, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{1n}y_1^{(r)} + \cdots + a_{nn}y_n^{(r)} &= x_n^{(r)} \end{aligned}\right\}, \tag{239a'}$$

by an appropriate numerical method in each cycle of the iteration.

* See also Problem 124.

This procedure clearly fails if **A** is *singular*, that is, if $\lambda = 0$ is a characteristic number of (231). A method which is useful in this case is presented in the following section (page 68).

To illustrate the basic procedure, we seek the largest characteristic value of λ for the system

$$\left.\begin{aligned} x_1 + x_2 + x_3 &= \lambda x_1, \\ x_1 + 2x_2 + 2x_3 &= \lambda x_2, \\ x_1 + 2x_2 + 3x_3 &= \lambda x_3 \end{aligned}\right\}. \tag{241}$$

With the initial approximation $\mathbf{x}^{(1)} = \{1, 1, 1\}$, there follows

$$\mathbf{y}^{(1)} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix} \begin{Bmatrix} 1 \\ 1 \\ 1 \end{Bmatrix} = \begin{Bmatrix} 3 \\ 5 \\ 6 \end{Bmatrix} = 6 \begin{Bmatrix} \frac{1}{2} \\ \frac{5}{6} \\ 1 \end{Bmatrix}. \tag{242}$$

If we determine λ such that the x_3 components of $\lambda\,\mathbf{x}^{(1)}$ and $\mathbf{y}^{(1)}$ are equal, we have $\lambda^{(1)} = 6$. Next, with $\mathbf{x}^{(2)} = \{\frac{1}{2}, \frac{5}{6}, 1\}$, there follows

$$\mathbf{y}^{(2)} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix} \begin{Bmatrix} \frac{1}{2} \\ \frac{5}{6} \\ 1 \end{Bmatrix} = \begin{Bmatrix} \frac{7}{3} \\ \frac{25}{6} \\ \frac{31}{6} \end{Bmatrix} = \frac{31}{6} \begin{Bmatrix} \frac{14}{31} \\ \frac{25}{31} \\ 1 \end{Bmatrix}. \tag{243}$$

The second approximation to the dominant characteristic number is then $\lambda^{(2)} = \frac{31}{6} \doteq 5.17$. The third step then gives

$$\mathbf{y}^{(3)} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix} \begin{Bmatrix} \frac{14}{31} \\ \frac{25}{31} \\ 1 \end{Bmatrix} = \begin{Bmatrix} \frac{70}{31} \\ \frac{126}{31} \\ \frac{157}{31} \end{Bmatrix} = \frac{157}{31} \begin{Bmatrix} \frac{70}{157} \\ \frac{126}{157} \\ 1 \end{Bmatrix} \tag{244}$$

and also $\lambda^{(3)} = \frac{157}{31} \doteq 5.06$. The ratios $x_1 : x_2 : x_3$ according to the four approximations are (1:1:1), (0.500:0.833:1), and (0.446:0.803:1). The next cycle leads to the value $\lambda^{(4)} \doteq 5.05$, and to the ratios 0.445:0.802:1, which hence may be expected to be accurate to three significant figures.

1.24. Additional techniques. In order to improve and extend the procedure just outlined, in the case when **A** is *real and symmetric*, it is desirable to consider the analytical basis of the precedure in that case.* For this purpose, we may suppose that $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$ are the true orthogonalized characteristic unit vectors of the problem (231), corresponding to the characteristic numbers $\lambda_1, \lambda_2, \cdots, \lambda_n$, arranged in increasing order of

* Certain other cases are considered in Sections 1.25 and 1.26.

magnitude. If the initial assumption $\mathbf{x}^{(1)}$ is imagined to be expressed in the form

$$\mathbf{x}^{(1)} = \sum_{k=1}^{n} c_k \mathbf{e}_k, \tag{245a}$$

then the vector $\mathbf{y}^{(1)} = \mathbf{A}\,\mathbf{x}^{(1)}$ accordingly must be given by

$$\mathbf{y}^{(1)} = \sum_{k=1}^{n} c_k \mathbf{A}\,\mathbf{e}_k = \sum_{k=1}^{n} \lambda_k c_k \mathbf{e}_k. \tag{245b}$$

Next, if a multiple of $\mathbf{y}^{(1)}$, say $\alpha_1 \mathbf{y}^{(1)}$, is taken to be $\mathbf{x}^{(2)}$, there then follows similarly

$$\mathbf{x}^{(2)} = \alpha_1 \sum_{k=1}^{n} \lambda_k c_k \mathbf{e}_k, \qquad \mathbf{y}^{(2)} = \alpha_1 \sum_{k=1}^{n} \lambda_k^2 c_k \mathbf{e}_k. \tag{246a,b}$$

More generally, after r steps we have

$$\mathbf{x}^{(r)} = \alpha'_{r-1} \sum_{k=1}^{n} \lambda_k^{r-1} c_k \mathbf{e}_k = \alpha'_{r-1} \lambda_n^{r-1} \sum_{k=1}^{n} \left(\frac{\lambda_k}{\lambda_n}\right)^{r-1} c_k \mathbf{e}_k$$

or

$$\mathbf{x}^{(r)} = \alpha'_{r-1} \lambda_n^{r-1} \left[c_n \mathbf{e}_n + \left(\frac{\lambda_{n-1}}{\lambda_n}\right)^{r-1} c_{n-1} \mathbf{e}_{n-1} + \cdots + \left(\frac{\lambda_1}{\lambda_n}\right)^{r-1} c_1 \mathbf{e}_1 \right] \tag{247a}$$

where $\alpha'_{r-1} = \alpha_1 \alpha_2 \cdots \alpha_{r-1}$, and, correspondingly,

$$\mathbf{y}^{(r)} = \alpha'_{r-1} \lambda_n^{r} \left[c_n \mathbf{e}_n + \left(\frac{\lambda_{n-1}}{\lambda_n}\right)^{r} c_{n-1} \mathbf{e}_{n-1} + \cdots + \left(\frac{\lambda_1}{\lambda_n}\right)^{r} c_1 \mathbf{e}_1 \right]. \tag{247b}$$

Since λ_n is the dominant characteristic number, the powers $(\lambda_k/\lambda_n)^r$ tend to zero when $k \neq n$ if $|\lambda_n| > |\lambda_{n-1}| \geqq \cdots$, and the expressions tend to multiples of $\mathbf{e}_n$ as r increases except in the very special case when the initial assumption $\mathbf{x}^{(1)}$ happens to be exactly orthogonal to $\mathbf{e}_n$, so that $c_n = 0$.

If λ_n is a multiple root of the characteristic equation, it is easily seen that the process will still lead to *one* corresponding characteristic vector. The case when λ_n and $-\lambda_n$ are both characteristic numbers requires special treatment.* However, in most practical cases the characteristic numbers are all nonnegative.

The *rate* of convergence of the method clearly depends upon the magnitude of the ratio of the two largest characteristic numbers. In case this ratio is near unity, and the convergence rate is slow, the matrix $\mathbf{A}$ may be first raised to an integral power p. The characteristic numbers of the new matrix $\mathbf{A}^p$ are then $\lambda_1^p, \cdots, \lambda_n^p$, and the ratio of the dominant and subdominant numbers is increased.

* It is apparent from (247a) that if $\lambda_{n-1} = -\lambda_n$ the sequence $\mathbf{x}^{(1)}, \mathbf{x}^{(3)}, \cdots$ converges to a multiple of $c_n\mathbf{e}_n + c_{n-1}\mathbf{e}_{n-1}$, whereas the sequence $\mathbf{x}^{(2)}, \mathbf{x}^{(4)}, \cdots$ converges to a multiple of $c_n\mathbf{e}_n - c_{n-1}\mathbf{e}_{n-1}$ (see Problem 96).

We may notice from (247a,b) that, if at any stage of the iteration the true vector $\mathbf{e}_n$ were known, the condition

$$(\mathbf{e}_n, \mathbf{y}^{(r)}) = \lambda(\mathbf{e}_n, \mathbf{x}^{(r)}) \tag{248}$$

would lead to the relation

$$\alpha'_{r-1}\lambda_n{}^r c_n = \lambda\alpha'_{r-1}\lambda_n{}^{r-1}c_n \qquad \text{or} \qquad \lambda = \lambda_n, \tag{249}$$

and hence would determine λ_n *exactly*. Clearly, any multiple of $\mathbf{e}_n$ would serve the same purpose. Thus it may be expected that a reasonably good approximation to λ_n would be obtained by replacing $\mathbf{e}_n$ by a convenient multiple of either the approximation $\mathbf{x}^{(r)}$ or the better approximation $\mathbf{y}^{(r)}$ in (248). This procedure gives the alternative formulas

$$(\mathbf{x}^{(r)}, \mathbf{y}^{(r)}) \approx \lambda_n(\mathbf{x}^{(r)}, \mathbf{x}^{(r)}) \tag{250a}$$

or

$$(\mathbf{y}^{(r)}, \mathbf{y}^{(r)}) \approx \lambda_n(\mathbf{x}^{(r)}, \mathbf{y}^{(r)}), \tag{250b}$$

of which the second is in general the more nearly accurate. It can be shown that the approximation given by (250a) is always conservative in absolute value (when $\mathbf{A}$ is real and symmetric). The same is true of that given by (250b) if the matrix $\mathbf{A}$ is also positive definite (see Problem 119).

We list in the following table the results of applying (A) the method of the preceding section, (B) the formula of (250a), and (C) the formula of (250b), to the illustrative example:

r	(A)	(B)	(C)
1	6.000	4.667	5.000
2	5.167	5.043	5.048
3	5.065	5.049	5.049
4	5.051	5.049	5.049

It may be seen that if (250a) or (250b) is used, the successive approximations to λ_n converge more rapidly than do the approximations to $\mathbf{e}_n$. This statement is generally true. Thus these formulas are useful in those cases when an accurate value of the dominant characteristic number is required, but comparable accuracy in the determination of the corresponding characteristic vector is not needed.

To obtain the *smallest* characteristic number of (241), we may write $\kappa = 1/\lambda$, resolve the equations in the form

$$\left.\begin{aligned} 2x_1 - x_2 &= \kappa x_1, \\ -x_1 + 2x_2 - x_3 &= \kappa x_2, \\ -x_2 + x_3 &= \kappa x_3 \end{aligned}\right\}, \tag{251}$$

and determine the largest characteristic value of κ by the preceding methods, or we may proceed equivalently by making use of (239) and (240).

Suppose now that *one* characteristic vector, say one which corresponds to a dominant characteristic number, is known *exactly*. Then for a *symmetric* matrix, all other characteristic vectors may be considered to be orthogonal to $\mathbf{e}_n$.* Hence, if we impose the constraint

$$(\mathbf{e}_n, \mathbf{x}) = 0 \tag{252}$$

on the problem (231), the resultant problem will possess those characteristic numbers and corresponding characteristic vectors which are in addition to λ_n and $\mathbf{e}_n$. But (252) permits one of the components, say x_r, to be expressed as a linear combination of the others. Hence we may eliminate x_r from the scalar equations corresponding to (231), disregard the rth resulting equation, and obtain a set of $n - 1$ equations involving only $n - 1$ components. The dominant characteristic number, and a corresponding characteristic vector, are then obtained as before, the component x_r being determined finally from (252).

Whereas the coefficient matrix associated with the new set of $n - 1$ equations is generally nonsymmetric, the convergence of the iterative method is assured in this case by results to be obtained in Section 1.26.

In particular, in the case when $|\mathbf{A}| = 0$ so that $\lambda = 0$ is a characteristic number of $\mathbf{A}$, we may replace $\mathbf{e}_n$ in (252) by the corresponding characteristic vector. Unless $\lambda = 0$ is of multiplicity greater than one, the corresponding reduced set of equations can then be inverted for the purpose of determining the smallest nonzero characteristic number. In the more general case, a number of unknowns equal to the multiplicity of the number $\lambda = 0$ must be eliminated in this way.

The procedure may be repeated until the solution is concluded or until only two components remain, at which stage the characteristic equation is quadratic in λ and the analysis can be conveniently completed without matrix iteration. Thus, if $\mathbf{A}$ is of order three, only one iterative process is needed. If $\mathbf{A}$ is of order four, we may conveniently determine the largest and smallest characteristic numbers and their corresponding vectors. The conditions $(\mathbf{e}_1, \mathbf{x}) = 0$ and $(\mathbf{e}_4, \mathbf{x}) = 0$ then permit the elimination of two components, and the reduction of the problem to one involving only the two remaining components.

In practice, the determination of the primary characteristic vector is only approximately effected. It is found that the numerical determination of a subdominant characteristic vector will often involve repeated subtraction of nearly equal quantities, particularly if the two relevant characteristic

* If the characteristic numbers are distinct, this *must* be so; otherwise, we may impose this condition without loss of generality.

numbers are nearly equal. In such cases, it may be necessary to calculate the components of the dominant characteristic vector to a degree of accuracy much higher than that required for the subdominant characteristic vector.

To illustrate the reduction in the preceding example, we notice that the dominant characteristic vector is given by $\{0.445, 0.802, 1\}$ to three significant figures. Hence (252) here becomes

$$0.445x_1 + 0.802x_2 + x_3 = 0. \tag{253}$$

If we eliminate x_3 between (253) and (241), and notice that the third equation is then a consequence of the first two (to the three significant figures retained), we obtain the reduced problem

$$\left.\begin{aligned} 0.555x_1 + 0.198x_2 &= \lambda x_1, \\ 0.110x_1 + 0.396x_2 &= \lambda x_2 \end{aligned}\right\}. \tag{254}$$

The dominant characteristic number of (254), and the two components of the corresponding characteristic vector, can then be obtained by matrix iteration, if this is desired, the component x_3 being determined in terms of them by (253). Otherwise, since the characteristic equation of (254) is quadratic, that equation can be solved by the quadratic formula, and the ratio of the x_1 and x_2 components of the corresponding characteristic vectors can be obtained directly.

✍ **1.25. Generalized characteristic-value problems.** In some applications we encounter characteristic-value problems of the more general form

$$\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{B}\,\mathbf{x}, \tag{255}$$

where $\mathbf{A}$ and $\mathbf{B}$ are real square matrices of order n. Such a problem reduces to the type considered previously when $\mathbf{B} = \mathbf{I}$. The characteristic equation corresponding to (255) is of the form

$$|\mathbf{A} - \lambda\,\mathbf{B}| = 0. \tag{256}$$

In the important practical cases in which both $\mathbf{A}$ and $\mathbf{B}$ are *symmetric*, so that $\mathbf{A}^T = \mathbf{A}$ and $\mathbf{B}^T = \mathbf{B}$, we next establish a useful generalization of the results of Section 1.12. If λ_1 and λ_2 are distinct characteristic numbers corresponding, respectively, to the characteristic vectors $\mathbf{u}_1$ and $\mathbf{u}_2$, there follows

$$\mathbf{A}\,\mathbf{u}_1 = \lambda_1\mathbf{B}\,\mathbf{u}_1, \qquad \mathbf{A}\,\mathbf{u}_2 = \lambda_2\mathbf{B}\,\mathbf{u}_2$$

and hence also

$$(\mathbf{A}\,\mathbf{u}_1)^T\,\mathbf{u}_2 = \lambda_1(\mathbf{B}\,\mathbf{u}_1)^T\,\mathbf{u}_2, \qquad \mathbf{u}_1{}^T\,\mathbf{A}\,\mathbf{u}_2 = \lambda_2\mathbf{u}_1{}^T\,\mathbf{B}\,\mathbf{u}_2,$$

or, making use of the symmetry in $\mathbf{A}$ and $\mathbf{B}$,

$$\mathbf{u}_1{}^T\,\mathbf{A}\,\mathbf{u}_2 = \lambda_1\mathbf{u}_1{}^T\,\mathbf{B}\,\mathbf{u}_2, \qquad \mathbf{u}_1{}^T\,\mathbf{A}\,\mathbf{u}_2 = \lambda_2\mathbf{u}_1{}^T\,\mathbf{B}\,\mathbf{u}_2. \tag{257}$$

By subtracting the first equation from the second in (257), we then obtain the relation

$$(\lambda_2 - \lambda_1)\mathbf{u}_1{}^T \mathbf{B}\, \mathbf{u}_2 = 0. \tag{258}$$

Thus, since $\lambda_1 \neq \lambda_2$ by assumption, we conclude that $\mathbf{u}_1{}^T \mathbf{B}\, \mathbf{u}_2 = 0$. That is, *if* $\mathbf{u}_1$ *and* $\mathbf{u}_2$ *are characteristic vectors, corresponding to two distinct characteristic numbers of the problem* $\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{B}\,\mathbf{x}$, *where* $\mathbf{A}$ *and* $\mathbf{B}$ *are symmetric, there follows*

$$\mathbf{u}_1{}^T \mathbf{B}\, \mathbf{u}_2 = 0. \tag{259}$$

It is convenient to speak of the left-hand member of (259) as the *scalar product of* $\mathbf{u}_1$ *and* $\mathbf{u}_2$ *relative to* $\mathbf{B}$, and to say that when (259) is satisfied *the vectors* $\mathbf{u}_1$ *and* $\mathbf{u}_2$ *are orthogonal relative to* $\mathbf{B}$. The ordinary type of orthogonality is thus relative to the *unit matrix* $\mathbf{I}$.

In consequence of (259) and (257), we deduce that *the vectors* $\mathbf{u}_1$ *and* $\mathbf{u}_2$ *are also orthogonal relative to the matrix* $\mathbf{A}$.

The left-hand member of (259) is conveniently denoted by $(\mathbf{u}_1, \mathbf{u}_2)_\mathbf{B}$. More generally, we write

$$(\mathbf{u}, \mathbf{v})_\mathbf{B} \equiv \mathbf{u}^T \mathbf{B}\, \mathbf{v} = \mathbf{v}^T \mathbf{B}\, \mathbf{u} \tag{260}$$

for the scalar product of $\mathbf{u}$ and $\mathbf{v}$ relative to a *symmetric* matrix $\mathbf{B}$. In particular, when $\mathbf{v} = \mathbf{u}$ we define the product

$$l_\mathbf{B}{}^2(\mathbf{u}) \equiv (\mathbf{u}, \mathbf{u})_\mathbf{B} \equiv \mathbf{u}^T \mathbf{B}\, \mathbf{u} \tag{261}$$

to be the square of the *generalized length* of $\mathbf{u}$, relative to $\mathbf{B}$. In order that this quantity be necessarily *positive* when $\mathbf{u}$ is real, except only when $\mathbf{u}$ is a *zero vector*, the matrix $\mathbf{B}$ must be *positive definite*. This is the case which most frequently arises in practice.*

In the remainder of this section, we assume that $\mathbf{A}$ *is real and symmetric and* $\mathbf{B}$ *real, symmetric, and positive definite.* In particular, this implies that $\mathbf{B}$ is *nonsingular*. The generalized length of a real vector, relative to $\mathbf{B}$, is then real and positive unless the vector is a zero vector, in which case its generalized length is zero.

Further it is easily shown that *when* $\mathbf{B}$ *is real and positive definite, any set of nonzero real vectors which are mutually orthogonal relative to* $\mathbf{B}$ *is a linearly independent set* (see Problem 103).

By a method analogous to that used in Section 1.12, it is then easily shown that *the characteristic numbers of* (255) *are real.* Further, by an argument similar to that used in Section 1.18, it can be shown that *to a characteristic number of multiplicity s there correspond s linearly independent characteristic vectors.* Then, by methods completely analogous to those of

* Usually at least *one* of the matrices A and B is positive definite. If A is positive definite, we may replace λ by $1/\lambda'$ and interchange the roles of A and B throughout this section.

Section 1.13, this set can be orthogonalized relative to **B**, and normalized in such a way that each vector possesses generalized length unity. It is seen that the condition $|\mathbf{B}| \neq 0$ guarantees that the characteristic equation (256) be of degree n. Hence, in the case under consideration, we may always obtain a set of n mutually orthogonal unit characteristic vectors $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$, such that

$$(\mathbf{e}_i, \mathbf{e}_j)_{\mathbf{B}} = \delta_{ij}. \tag{262}$$

An *orthonormal modal matrix* **M**, associated with (255), may now be defined as the matrix having the components of the kth vector of the set as the elements of its kth *column*. Then, in consequence of the relation

$$\mathbf{A}\,\mathbf{e}_i = \lambda_i \mathbf{B}\,\mathbf{e}_i \qquad (i = 1, 2, \cdots, n),$$

there follows

$$\mathbf{A}\,\mathbf{M} = \mathbf{B}\,\mathbf{M}\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} \equiv \mathbf{B}\,\mathbf{M}\,\mathbf{D}, \tag{263}$$

and also, by virtue of (262),

$$\mathbf{M}^T \mathbf{B}\,\mathbf{M} = \mathbf{I}. \tag{264}$$

(See Problem 41.)

We may now verify the fact that, with the change of variables

$$\mathbf{x} = \mathbf{M}\,\boldsymbol{\alpha}, \tag{265}$$

the *two* quadratic forms

$$A = \mathbf{x}^T \mathbf{A}\,\mathbf{x}, \qquad B = \mathbf{x}^T \mathbf{B}\,\mathbf{x} \tag{266a,b}$$

are reduced *simultaneously* to the canonical forms

$$A = \boldsymbol{\alpha}^T \mathbf{D}\,\boldsymbol{\alpha} = \lambda_1 \alpha_1^2 + \lambda_2 \alpha_2^2 + \cdots + \lambda_n \alpha_n^2, \tag{267a}$$

$$B = \boldsymbol{\alpha}^T \boldsymbol{\alpha} = \alpha_1^2 + \alpha_2^2 + \cdots + \alpha_n^2. \tag{267b}$$

For the substitution of (265) into (266b), and the use of (264), gives immediately

$$B = \boldsymbol{\alpha}^T \mathbf{M}^T \mathbf{B}\,\mathbf{M}\,\boldsymbol{\alpha} = \boldsymbol{\alpha}^T \boldsymbol{\alpha},$$

in accordance with (267b), whereas the substitution of (265) into (266a) gives

$$A = \boldsymbol{\alpha}^T \mathbf{M}^T \mathbf{A}\,\mathbf{M}\,\boldsymbol{\alpha}$$

and the use of (263) and (264) leads to the result

$$A = \boldsymbol{\alpha}^T \mathbf{M}^T \mathbf{B}\,\mathbf{M}\,\mathbf{D}\,\boldsymbol{\alpha} = \boldsymbol{\alpha}^T \mathbf{D}\,\boldsymbol{\alpha},$$

in accordance with (267a).

From (264) it follows that

$$|\mathbf{M}| = \pm \frac{1}{\sqrt{|\mathbf{B}|}},$$

so that the transformation (265) is *nonsingular*. Further, since (264) leads to the relation $\mathbf{M}^{-1} = \mathbf{M}^T \mathbf{B}$, the inversion of (265) may be conveniently effected by use of the equation

$$\boldsymbol{\alpha} = \mathbf{M}^T \mathbf{B}\, \mathbf{x}. \tag{268}$$

Equations (267a,b) are identical with equations (169) and (171) of Section 1.19. It is important to notice that the coefficients λ_i in (267) are the roots of the equation $|\mathbf{A} - \lambda \mathbf{B}| = 0$, and hence are *real*, under the present restrictions on **A** and **B**.

If the matrices **A** *and* **B** *are both positive definite, the characteristic numbers* λ_i *are also necessarily positive.* This result is established by noticing that the relation

$$\mathbf{A}\, \mathbf{e}_i = \lambda_i \mathbf{B}\, \mathbf{e}_i$$

implies the relation

$$\mathbf{e}_i{}^T \mathbf{A}\, \mathbf{e}_i = \lambda_i \mathbf{e}_i{}^T \mathbf{B}\, \mathbf{e}_i.$$

Since both $\mathbf{e}_i{}^T \mathbf{A}\, \mathbf{e}_i$ and $\mathbf{e}_i{}^T \mathbf{B}\, \mathbf{e}_i$ are positive when **A** and **B** are positive definite (and $\mathbf{e}_i \neq \mathbf{0}$), the same is true of λ_i.

The preceding results will be of importance in Section 2.14 of the following chapter.

Since the vectors $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$ are *linearly independent* (see Problem 103), it follows that any vector **v** in n-space can be expressed as a linear combination of these vectors, of the form

$$\mathbf{v} = \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2 + \cdots + \alpha_n \mathbf{e}_n = \sum_{k=1}^{n} \alpha_k \mathbf{e}_k. \tag{269}$$

In order to evaluate any coefficient α_r, we merely form the generalized scalar product of $\mathbf{e}_r$ into both sides of (269), and use (262) to obtain the result

$$\alpha_r = (\mathbf{e}_r, \mathbf{v})_{\mathbf{B}} \equiv \mathbf{e}_r{}^T \mathbf{B}\, \mathbf{v} \qquad (r = 1, 2, \cdots, n). \tag{270}$$

The case of most common occurrence in practice is that in which **B** is a *diagonal matrix* **G**, say

$$\mathbf{G} = \begin{bmatrix} g_1 & 0 & \cdots & 0 \\ 0 & g_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & g_n \end{bmatrix}, \tag{271}$$

so that the equations $\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{G}\,\mathbf{x}$ take the special form

$$\left.\begin{aligned} &a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = \lambda g_1 x_1, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = \lambda g_n x_n \end{aligned}\right\}, \tag{272}$$

where $a_{ji} = a_{ij}$.

The generalized scalar product $(\mathbf{x}, \mathbf{y})_G$ then takes the form

$$(\mathbf{x}, \mathbf{y})_G = g_1 x_1 y_1 + g_2 x_2 y_2 + \cdots + g_n x_n y_n, \tag{273}$$

while the generalized length of $\mathbf{x}$ is given by

$$l_G^2(\mathbf{x}) = (\mathbf{x}, \mathbf{x})_G = g_1 x_1^2 + g_2 x_2^2 + \cdots + g_n x_n^2. \tag{274}$$

The condition that $\mathbf{G}$ be positive definite requires that the diagonal elements be positive:

$$g_i > 0 \qquad (i = 1, 2, \cdots, n). \tag{275}$$

It may be noticed that in certain cases a set of equations of the matrix form $\mathbf{A}'\,\mathbf{x} = \lambda\,\mathbf{x}$, where $\mathbf{A}'$ is a *nonsymmetric* square matrix, can be reduced to a set of the matrix form $\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{G}\,\mathbf{x}$, where $\mathbf{A}$ is symmetric and $\mathbf{G}$ is a diagonal matrix with positive diagonal elements, by multiplying the ith equation of the original set by a suitably chosen *positive* constant g_i. When $\mathbf{A}'$ is of order *two*, this reduction is clearly always possible if $a'_{12}a'_{21} > 0$; it is possible in other cases only when the coefficients satisfy certain compatibility conditions. If and only if such a reduction is possible, $\mathbf{A}'$ can be expressed as a product $\mathbf{D}\,\mathbf{A}$, where $\mathbf{D} \equiv \mathbf{G}^{-1}$ is a *diagonal* matrix with positive diagonal elements, and $\mathbf{A}$ is symmetric.

To conclude this section, we indicate the extension of the numerical methods of the preceding sections to the treatment of a characteristic-value problem of the form

$$\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{B}\,\mathbf{x}, \tag{276}$$

where again $\mathbf{A}$ is a symmetric matrix of order n, and $\mathbf{B}$ is a positive definite, symmetric, matrix of the same order.

Since, by assumption, $\mathbf{B}$ is nonsingular, equation (276) can be reduced to the form

$$\mathbf{B}^{-1}\,\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{x}, \tag{277}$$

which is of the type considered previously. However, the matrix $\mathbf{B}^{-1}\,\mathbf{A}$ will now *not* be symmetric, in general. In the case of (272) the reduction to the form (277) involves only division of both sides of the ith equation by g_i.

Whether or not the transformation of (276) to (277) is effected, we may define sequences of vectors $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots$ and $\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \cdots$ by the relations

$$\mathbf{A}\,\mathbf{x}^{(r)} = \mathbf{B}\,\mathbf{y}^{(r)}, \qquad \mathbf{x}^{(r+1)} = \alpha_r \mathbf{y}^{(r)} \qquad (r = 1, 2, \cdots), \tag{278a,b}$$

where α_r is a conveniently chosen nonzero constant and where $\mathbf{x}^{(1)}$ is to be *chosen* to start the iteration, in the hope that there will follow

$$\mathbf{y}^{(r)} \sim \lambda_n \mathbf{x}^{(r)}, \qquad \mathbf{x}^{(r)} \to c\,\mathbf{e}_n \qquad (r \to \infty) \tag{279a,b}$$

in general, where λ_n is the relevant characteristic number of *largest* magnitude. Here the vector $\mathbf{y}^{(r)}$ is to be obtained from $\mathbf{x}^{(r)}$, in accordance with (278a), either by making use of the matrix $\mathbf{B}^{-1}$ or by a numerical solution of the associated set of scalar equations in each iteration.

In order to investigate the convergence of the iterative procedure in this case, let the normalized characteristic unit vectors be denoted by $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$, corresponding, respectively, to $\lambda_1, \lambda_2, \cdots, \lambda_n$, so that

$$\mathbf{A}\,\mathbf{e}_r = \lambda_r \mathbf{B}\,\mathbf{e}_r. \tag{280}$$

The initial approximation $\mathbf{x}^{(1)}$ can be imagined to be expressed as a linear combination of these vectors, in the form

$$\mathbf{x}^{(1)} = \sum_{k=1}^{n} c_k \mathbf{e}_k. \tag{281a}$$

There then follows

$$\mathbf{y}^{(1)} \equiv \mathbf{B}^{-1}\mathbf{A}\,\mathbf{x}^{(1)} = \sum_{k=1}^{n} c_k \mathbf{B}^{-1}\mathbf{A}\,\mathbf{e}_k = \sum_{k=1}^{n} c_k \mathbf{B}^{-1}\lambda_k \mathbf{B}\,\mathbf{e}_k$$

or

$$\mathbf{y}^{(1)} \equiv \mathbf{B}^{-1}\mathbf{A}\,\mathbf{x}^{(1)} = \sum_{k=1}^{n} \lambda_k c_k \mathbf{e}_k. \tag{281b}$$

By comparing (281a,b) with (245a,b) of the preceding section, we see that the arguments presented in that section again apply here, to show that successive approximations will indeed converge to a multiple of the dominant vector $\mathbf{e}_n$ when $|\lambda_n| > |\lambda_{n-1}| \geqq \cdots$, if $c_n \neq 0$.

In this case, however, it is seen that the requirement

$$(\mathbf{e}_n, \mathbf{y}^{(r)})_{\mathbf{B}} = \lambda(\mathbf{e}_n, \mathbf{x}^{(r)})_{\mathbf{B}},$$

in place of (248), would give $\lambda = \lambda_n$ exactly. Hence (250a,b) here should be modified to the alternative conditions

$$(\mathbf{x}^{(r)}, \mathbf{y}^{(r)})_{\mathbf{B}} \approx \lambda_n (\mathbf{x}^{(r)}, \mathbf{x}^{(r)})_{\mathbf{B}} \tag{282a}$$

or, better,

$$(\mathbf{y}^{(r)}, \mathbf{y}^{(r)})_{\mathbf{B}} \approx \lambda_n (\mathbf{x}^{(r)}, \mathbf{y}^{(r)})_{\mathbf{B}}. \tag{282b}$$

Similarly, equation (252) must be replaced by the relation

$$(\mathbf{e}_n, \mathbf{x})_{\mathbf{B}} = 0, \tag{283}$$

which permits reduction of the order of the system when one characteristic vector has been obtained.

The same statements apply to the inversion of (277),

$$\frac{1}{\lambda}\mathbf{x} = \mathbf{A}^{-1}\mathbf{B}\,\mathbf{x},$$

which may be used in determining the smallest characteristic value of λ when $\mathbf{A}$ is nonsingular.* Here the relations (278) and (279) are to be replaced by the relations

$$\mathbf{A}\,\mathbf{y}^{(r)} = \mathbf{B}\,\mathbf{x}^{(r)}, \qquad \mathbf{x}^{(r+1)} = \alpha_r \mathbf{y}^{(r)} \qquad (r = 1, 2, \cdots) \qquad (284a,b)$$

and

$$\mathbf{y}^{(r)} \sim \frac{1}{\lambda_1}\mathbf{x}^{(r)}, \qquad \mathbf{x}^{(r)} \to c\,\mathbf{e}_1 \qquad (r \to \infty), \qquad (285a,b)$$

where the iteration is to be initiated by choosing $\mathbf{x}^{(1)}$ and determining $\mathbf{y}^{(1)}$, either by making use of $\mathbf{A}^{-1}$ or by a direct numerical method.

1.26. Characteristic numbers of nonsymmetric matrices. Whereas the characteristic equation $|\mathbf{A} - \lambda\,\mathbf{I}| = 0$ of a nonsymmetric square matrix $\mathbf{A}$ of order n is of degree n, we have seen that when the roots of this equation are not *distinct* the total number of linearly independent characteristic vectors may be less than n. In the present section we exclude the exceptional cases, which rarely occur in practice, and suppose that the n characteristic numbers of $\mathbf{A}$ are *real and distinct*. The corresponding characteristic vectors are then linearly independent.

In order to establish this fact, we assume the contrary and deduce a contradiction. Suppose that the characteristic numbers $\lambda_1, \cdots, \lambda_n$ are all distinct, and denote the corresponding characteristic vectors by $\mathbf{u}_1, \cdots, \mathbf{u}_n$. We then have the relations $\mathbf{A}\,\mathbf{u}_i = \lambda_i \mathbf{u}_i$ for $i = 1, 2, \cdots, n$. *Assume* that the first r characteristic vectors are linearly independent, but that

$$\mathbf{u}_{r+1} = \sum_{k=1}^{r} c_k \mathbf{u}_k,$$

where at least one c_k is not zero since $\mathbf{u}_{r+1} \neq \mathbf{0}$. By premultiplying the equal members of this relation by $\mathbf{A}$, there then follows

$$\lambda_{r+1}\mathbf{u}_{r+1} = \sum_{k=1}^{r} c_k \lambda_k \mathbf{u}_k,$$

and hence also, by comparing these relations,

$$\sum_{k=1}^{r} c_k(\lambda_{r+1} - \lambda_k)\mathbf{u}_k = \mathbf{0}.$$

But, since $\mathbf{u}_1, \cdots, \mathbf{u}_r$ are linearly independent, the coefficient of *each* $\mathbf{u}_k$ must vanish. Since at least one c_k is not zero, at least one λ_k must equal λ_{r+1}, in contradiction with the hypothesis that the λ's are distinct.

In correspondence with the characteristic-value problem

$$\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{x}, \qquad (286)$$

* See also Problem 125.

we may consider the problem

$$\mathbf{A}^T \mathbf{x}' = \lambda \mathbf{x}', \tag{287}$$

associated with the *transpose* of **A**. By virtue of the fact that the two matrices $\mathbf{A} - \lambda \mathbf{I}$ and $\mathbf{A}^T - \lambda \mathbf{I}$ differ only in that rows and columns are interchanged, their determinants possess the same expansion, so that (286) *and* (287) *possess the same characteristic numbers.* Let λ_1 and λ_2 denote any two distinct characteristic numbers, and let corresponding solutions of (286) and (287) be denoted by $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_1'$, $\mathbf{u}_2'$, respectively. We then have the relations

$$\mathbf{A}\,\mathbf{u}_1 = \lambda_1 \mathbf{u}_1, \qquad \mathbf{A}^T \mathbf{u}_2' = \lambda_2 \mathbf{u}_2',$$

from which there follows

$$(\lambda_2 - \lambda_1)\mathbf{u}_1{}^T \mathbf{u}_2' = 0. \tag{288}$$

Hence we conclude that *any characteristic vector of* (286) *is orthogonal to any characteristic vector of* (287) *which corresponds to a different characteristic number:*

$$(\mathbf{u}_i, \mathbf{u}_j') = 0 \qquad (\lambda_i \neq \lambda_j). \tag{289}$$

This property permits the generalization of the methods of Sections 1.23 and 1.24 to the more general case considered here. While the problems considered in Section 1.25 are included in this generalization, the methods given in that section are usually preferable when they are applicable.

In particular, it is seen that the coefficients in the representation

$$\mathbf{v} = \sum_{k=1}^{n} \alpha_k \mathbf{u}_k \tag{290a}$$

are determined by forming the scalar product of $\mathbf{u}_k'$ with the two members of this equation, in the form

$$\alpha_k(\mathbf{u}_k, \mathbf{u}_k') = (\mathbf{v}, \mathbf{u}_k'). \tag{290b}$$

A development analogous to that of Section 1.24 then shows that the matrix iteration procedure again converges in this case to the characteristic vector corresponding to the dominant characteristic number of **A**. In fact, such a development shows that the convergence of this procedure is insured if the matrix **A** possesses n linearly independent characteristic vectors, and only *real* characteristic numbers (which need not be *distinct*).* While formulas analogous to (250a,b) can be devised for more accurate estimates of λ_n, their use involves a considerable increase in calculation.

* By a somewhat more involved analysis, which may be based on the generalized Sylvester formula mentioned in Section 1.22 (cf. Problem 89), it can be shown that the iterative procedure converges to the dominant characteristic number of any real matrix *if that number is real, and if no unequal characteristic number has equal absolute value.* If the dominant number is *repeated*, however, the rate of convergence is not always *exponential* and the procedure is often not practical unless the multiplicity of that number is somehow known in advance. Modifications (similar to those of Problem 96) which are useful in this special situation, as well as in the case when the dominant numbers are conjugate complex, are given in Reference 7.

The essential modification in procedure is involved in the calculation of subdominant characteristic quantities. In the more general case considered here, the constraint condition (252) must be replaced by the equation

$$(\mathbf{u}_n', \mathbf{x}) = 0. \tag{291}$$

Thus after the (approximate) determination of the dominant characteristic number λ_n, and the corresponding vector solution $\mathbf{u}_n$, a vector $\mathbf{u}_n'$ satisfying the related equation $\mathbf{A}^T \mathbf{x}' = \lambda_n \mathbf{x}'$ must be determined (see Problem 108). The constraint (291), which corresponds to the fact that all other characteristic vectors of $\mathbf{A}$ are orthogonal to $\mathbf{u}_n'$, then permits the elimination of one of the unknowns in the system of equations (and the neglect of one of the resulting equations) so that the order of the system is reduced by unity.

It may be noticed that the relation

$$\mathbf{A}^T \mathbf{u}_r' = \lambda_r \mathbf{u}_r'$$

can also be written in the form

$$\mathbf{u}_r'^T \mathbf{A} = \lambda_r \mathbf{u}_r'^T \tag{292}$$

Thus, if $\mathbf{u}_r'$ is a characteristic vector of $\mathbf{A}^T$, corresponding to λ_r, then the *row vector* $\mathbf{u}_r'^T$ is such that *post*multiplication by the matrix $\mathbf{A}$ multiplies $\mathbf{u}_r'^T$ by the scalar λ_r, whereas the *column vector* $\mathbf{u}_r$ is such that *pre*multiplication by $\mathbf{A}$ multiplies $\mathbf{u}_r$ by λ_r.

If, for *any* $n \times n$ matrix $\mathbf{A}$ with n independent characteristic vectors, a *modal matrix* $\mathbf{M}$ is formed, in such a way that the components of n successive linearly independent characteristic vectors of $\mathbf{A}$ comprise successive columns of $\mathbf{M}$, the matrix $\mathbf{A}$ can be *diagonalized* by the *similarity* transformation $\mathbf{M}^{-1} \mathbf{A} \mathbf{M}$, the resultant diagonal elements being the characteristic numbers of $\mathbf{A}$ (see Problem 109). However, in consequence of the fact that the n characteristic vectors are generally not orthogonal, it follows that $\mathbf{M}^{-1} \neq \mathbf{M}^T$, in general, so that the matrix $\mathbf{M}$ generally is *not* an *orthogonal matrix*.

If certain characteristic numbers of a nonsymmetric and non-Hermitian matrix are repeated, there may be less than n linearly independent characteristic vectors, so that complete diagonalization in this way is impossible. In any case, it can be shown that *any* square matrix can be transformed by a *similarity* transformation (which is not necessarily orthogonal) to a *canonical matrix* with the following properties:

1. All elements *below* the principal diagonal are zero.
2. The diagonal elements are the characteristic numbers of the matrix.
3. All elements *above* the principal diagonal are zero *except* possibly those elements which are adjacent to *two* equal diagonal elements.
4. The latter elements are each either zero or unity.

A matrix having these four properties is known as a *Jordan canonical matrix*.

In illustration, for a matrix of order five for which $\lambda_1 = \lambda_2 = \lambda_3$ and $\lambda_4 = \lambda_5$, but $\lambda_1 \neq \lambda_4$, this canonical form would be

$$\begin{bmatrix} \lambda_1 & \alpha_1 & 0 & 0 & 0 \\ 0 & \lambda_1 & \alpha_2 & 0 & 0 \\ 0 & 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & 0 & \lambda_4 & \alpha_3 \\ 0 & 0 & 0 & 0 & \lambda_4 \end{bmatrix},$$

where each of the elements α_1, α_2, and α_3 is either unity or zero, according as λ_1 corresponds to one, two, or three independent characteristic vectors, and λ_4 to one or two independent characteristic vectors. Reductions to certain other standard forms have also been studied.*

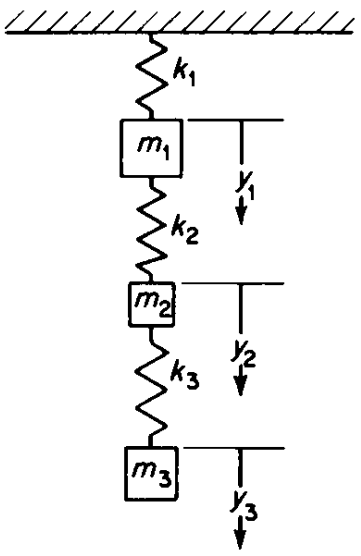

FIGURE 1.1

To conclude this section, we repeat the following comments made in Section 1.22:

The Cayley-Hamilton theorem [see equation (219)] *is true for any square matrix* **A**. *Sylvester's formula* [see equations (226) and (227)] *is valid for any square matrix* **A** *with no repeated characteristic numbers, and accordingly can be applied to the evaluation of any function* $f(\mathbf{A})$ *which is representable as a convergent series of powers of* **A**.

1.27. A physical application. Several applications of the preceding methods will be found in the chapters which follow. In this section, we present one such application to a mechanical problem.

We consider the problem of determining the natural modes of free vibration of the mechanical system indicated in Figure 1.1, in which the masses m_1, m_2, and m_3 are connected in series to a fixed support, by linear springs with spring constants k_1, k_2, and k_3. The effects of viscous damping are neglected. If we denote the displacements of the respective masses from their equilibrium positions by $y_1(t)$, $y_2(t)$, and $y_3(t)$, respectively, the differential equations of motion are of the form

$$\left.\begin{aligned} m_1 \frac{d^2y_1}{dt^2} &= k_2(y_2 - y_1) - k_1y_1 = -(k_1 + k_2)y_1 + k_2y_2, \\ m_2 \frac{d^2y_2}{dt^2} &= k_3(y_3 - y_2) - k_2(y_2 - y_1) = k_2y_1 - (k_2 + k_3)y_2 + k_3y_3. \\ m^3 \frac{d^2y_3}{dt^2} &= -k_3(y_3 - y_2) = k_3y_2 - k_3y_3 \end{aligned}\right\}. \quad (239)$$

* See Reference 13.

The *natural modes* of vibration are those in which the masses oscillate in phase with a common frequency, and hence the displacements are specified by equations of the form

$$\left.\begin{aligned} y_1(t) &= x_1 \sin(\omega t + \alpha), \\ y_2(t) &= x_2 \sin(\omega t + \alpha), \\ y_3(t) &= x_3 \sin(\omega t + \alpha) \end{aligned}\right\}, \tag{294}$$

where the amplitudes x_1, x_2, and x_3 and the common circular frequency ω are to be determined. By introducing (294) into (293), and canceling the common resultant time factors, we obtain the equations

$$\left.\begin{aligned} (k_1 + k_2)x_1 \qquad - k_2 x_2 \qquad &= m_1\omega^2 x_1, \\ -k_2 x_1 + (k_2 + k_3)x_2 - k_3 x_3 &= m_2\omega^2 x_2, \\ -k_3 x_2 + k_3 x_3 &= m_3\omega^2 x_3 \end{aligned}\right\}. \tag{295}$$

It should be noticed that the matrix of the coefficients in the left-hand members is *symmetric*. Also it is found that the "discriminants" Δ_m defined in Section 1.20 are of the form

$$\left.\begin{aligned} \Delta_1 &= k_1 + k_2, \\ \Delta_2 &= k_1k_2 + k_2k_3 + k_3k_1, \\ \Delta_3 &= k_1k_2k_3 \end{aligned}\right\}, \tag{296}$$

so that the matrix of coefficients is also *positive definite* when the spring constants are positive.

In the special case when $k_1 = k_2 = k_3 \equiv k$, and $m_1 = m_2 = m_3 \equiv m$, these equations reduce to equations (251) of Section 1.24 if we set

$$\kappa = \frac{1}{\lambda} = \frac{m\omega^2}{k}. \tag{297}$$

Hence the characteristic values of λ discussed in the example of that section are inversely proportional to the squares of the natural frequencies of the physical system under consideration, and the components of the characteristic vectors are in the same ratio as the three amplitudes x_1, x_2, and x_3 in a corresponding mode of vibration.

In the *fundamental mode*, corresponding to the *smallest* natural frequency, and hence to the *dominant* characteristic value of λ as defined by (297), the circular frequency is hence given by

$$\frac{m\omega_1^2}{k} = \frac{1}{5.05}: \qquad \omega_1 = 0.445\sqrt{\frac{k}{m}}.$$

Here the three masses all move in the same direction, the respective displacements from equilibrium at any instant being in the ratio 0.445:0.802:1.

By completing the analysis indicated in Section 1.24, we find that in the *second* mode there follows

$$\omega_2 = 1.247\sqrt{\frac{k}{m}}.$$

In this mode the first two masses move in the same direction, whereas the third mass moves in the opposite direction, the displacements being in the ratio $-1.247:-0.555:1$. In the *third* mode there follows

$$\omega_3 = 1.802\sqrt{\frac{k}{m}}.$$

Here the first and third masses move in the same direction, and the second mass in the opposite direction, the displacements being in the ratio $1.802:-2.247:1$.

The most general motion of the system, possible in the absence of externally applied forces, is then a superposition of the three modes just described, in which the phase angle α of equation (294) may take on different values in the individual modes.

If the three masses are unequal, equations (295) are of the form of equations (272), with g_i proportional to m_i. To illustrate the treatment of this case, we suppose that

$$k_1 = k_2 = k_3 \equiv k, \qquad m_1 = m_2 \equiv m, \qquad m_3 = 2m,$$

so that equations (295) become

$$\left.\begin{aligned} 2x_1 - x_2 &= \frac{m\omega^2}{k}x_1, \\ -x_1 + 2x_2 - x_3 &= \frac{m\omega^2}{k}x_2, \\ -x_2 + x_3 &= 2\frac{m\omega^2}{k}x_3 \end{aligned}\right\}.$$

In this case we may write

$$g_1 = 1, \qquad g_2 = 1, \qquad g_3 = 2.$$

In order to determine the *fundamental* mode *directly*, we may first *invert* these equations in the form

$$\left.\begin{aligned} x_1 + x_2 + 2x_3 &= \lambda x_1, \\ x_1 + 2x_2 + 4x_3 &= \lambda x_2, \\ x_1 + 2x_2 + 6x_3 &= \lambda x_3 \end{aligned}\right\},$$

where

$$\lambda = \frac{k}{m\omega^2}.$$

Except for refined successive estimates of $\lambda_3 = k/(m\omega_1^2)$, the calculation proceeds exactly as before. The results of successive steps are tabulated below to three significant figures:

	$\mathbf{x}^{(1)}$	$\mathbf{y}^{(1)}$	$\mathbf{x}^{(2)}$	$\mathbf{y}^{(2)}$	$\mathbf{x}^{(3)}$	$\mathbf{y}^{(3)}$	$\mathbf{x}^{(4)}$	$\mathbf{y}^{(4)}$	$\mathbf{x}^{(5)}$
x_1	1	4	0.444	3.22	0.402	3.15	0.399	3.15	0.399
x_2	1	7	0.777	6.00	0.750	5.90	0.747	5.89	0.747
x_3	1	9	1	8.00	1	7.90	1	7.89	1

Thus, after four cycles, the modal column $\{0.399, 0.747, 1\}$ is repeated. The dominant characteristic value of λ is seen to be 7.89, so that the fundamental circular frequency of the physical system is

$$\omega_1 = 0.356\sqrt{\frac{k}{m}}.$$

If only this value were of interest, and accurate values of the corresponding mode components were not required, the use of either (282a), in the form

$$\lambda(x_1^2 + x_2^2 + 2x_3^2) \approx (x_1y_1 + x_2y_2 + 2x_3y_3),$$

or (282b), in the form

$$\lambda(x_1y_1 + x_2y_2 + 2x_3y_3) \approx (y_1^2 + y_2^2 + 2y_3^2),$$

would yield the above result for the dominant value of λ after only two cycles.

If the remaining modes are required, the orthogonality relation (283) becomes

$$0.399x_1 + 0.747x_2 + 2.000x_3 = 0,$$

and permits the reduction of the order of the system to two.

1.28. Function space. In this section, we develop certain analogies between vector space and so-called "function space" and point out certain essential difficulties involved in the treatment of the latter.

If, in three-dimensional space, we consider any two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ which are not scalar multiples of each other, we see that the totality of all vectors of the form $c_1\mathbf{u} + c_2\mathbf{v}$ comprises a double infinity of vectors, namely, all vectors in that space which are parallel to the plane of $\mathbf{u}$ and $\mathbf{v}$. If $\mathbf{w}$ is any third nonzero vector which is not parallel to the plane of $\mathbf{u}$ and $\mathbf{v}$, then *all* vectors in that space are comprised in the representation $c_1\mathbf{u} + c_2\mathbf{v} + c_3\mathbf{w}$. More generally, in the language of linear vector spaces, we say that "any n linearly independent vectors form a basis in n-dimensional space."

Similarly, if we consider two functions $f(x)$ and $g(x)$, defined over an interval (a, b) and not multiples of each other over that interval, those functions which are of the form $c_1 f(x) + c_2 g(x)$ comprise a double infinity of functions. The question then arises as to the possibility of choosing a more comprehensive set of basic functions such that *any* function, satisfying appropriate regularity conditions, can be expressed as a linear combination of those functions over a certain given interval, that is, as to the possibility of choosing a "basis" in a "function space" associated with that interval. Certainly any such set of functions must have infinitely many members; that is, such a "function space" comprises infinitely many dimensions. Also, as in a vector space, we would expect the choice to be by no means a *unique* one.

In a *vector* space of n dimensions, great advantage is attained by choosing as a basis a set of n mutually *orthogonal* nonzero vectors, that is, a set of nonzero vectors such that the *scalar product* of any two distinct vectors in the set is zero. This fact suggests that we introduce an analogous definition of the scalar product of two *functions*, relative to the interval under consideration.

For real-valued functions, to which attention will be restricted here, it is found that a particularly useful definition is of the form

$$(f, g) = \int_a^b fg \, dx. \tag{298}$$

This definition is a natural generalization of the vector definition

$$(\mathbf{u}, \mathbf{v}) = \sum_{k=1}^{n} u_k v_k$$

when the dimension of the space and the number of components involved become infinitely large.

Thus (assuming here and henceforth that the functions involved are such that the integrals involved *exist*) we are led to say that *two functions* $f(x)$ *and* $g(x)$ *are orthogonal over an interval* (a, b) *if the integral* $\int_a^b fg \, dx$ *vanishes.*

In particular, when $f = g$ we may think of the number $\int_a^b f^2 \, dx$ as the "square of the length" of $f(x)$ in the function space associated with the interval (a, b). It is more conventional to speak of this quantity as the square of the *norm* of f, and to write

$$\text{norm } f \equiv \|f\| = (f, f)^{1/2} = \left(\int_a^b f^2 \, dx \right)^{1/2}. \tag{299}$$

A function whose norm is *unity* is said to be *normalized*, and is seen to be analogous to a *unit vector* in vector space.*

* In some references, the norm of f is defined as the quantity (f, f) itself.

We notice that if the norm of f is *zero*, then the integral of the non-negative function f^2 over the interval (a, b) must vanish. This means that $f(x)$ cannot differ from zero over any range of positive length in (a, b). In particular, if f is *continuous* everywhere in (a, b), and has a zero norm over that interval, then f must *vanish everywhere* in (a, b). However, it is clear that if $f(x)$ were zero everywhere except at a finite number of points in (a, b), the integral of f^2 over that interval would still vanish. It is convenient to speak of a function $f(x)$ for which $\int_a^b f^2\,dx = 0$ as a *trivial function*, and to say that such a function vanishes "almost everywhere" in (a, b).*

A set of n functions may be said to be *linearly independent* in (a, b) if *no* linear combination of those functions, with at least one nonvanishing coefficient, is a trivial function over that interval. Given any such set of real functions $f_k(x)$, for each of which the integral $\int_a^b f_k^2\,dx$ exists and is nonzero, we can determine a set of n new functions $\phi_k(x)$, each of which is a linear combination of certain of the f's, such that the ϕ's are mutually orthogonal and normalized in (a, b). The procedure is completely analogous to the Gram-Schmidt procedure of Section 1.13. We call such a set of functions an *orthonormal* set. Any two functions of the set then have the property that

$$(\phi_i, \phi_j) \equiv \int_a^b \phi_i \phi_j\,dx = \delta_{ij}, \tag{300}$$

where δ_{ij} is the Kronecker delta of equation (37).

Now for any (sufficiently regular) function $f(x)$ defined in (a, b) we may evaluate the scalar product of that function with each function ϕ_k:

$$c_k = (f, \phi_k) = \int_a^b f\phi_k\,dx. \tag{301}$$

The functions ϕ_k are analogous to a set of n mutually orthogonal unit vectors in space, and we may think of the numbers $c_1, c_2, \cdots, c_n$ as the scalar components of $f(x)$ relative to those functions. We refer to these numbers as the *Fourier constants* of $f(x)$ relative to the functions $\phi_k(x)$ in (a, b).

There then exists an n-fold infinity of functions which can be generated as linear combinations of the n ϕ's. If for any such function $F(x)$ we write

$$F(x) = \sum_{k=1}^{n} a_k \phi_k(x) \qquad (a < x < b), \tag{302}$$

* For a more satisfactory definition of this term it is desirable to consider an extension of the usual concepts of integration. The points at which a trivial function differs from zero may be *infinite* in number, so long as they are not "densely" distributed.

each coefficient a_r can be determined by forming the scalar product of ϕ_r with both members of (302), and using (300) to obtain the result

$$a_r = (F, \phi_r). \tag{303}$$

Thus the coefficient a_k in (302) is the scalar component of $F(x)$ relative to $\phi_k(x)$.

For a more general real function $f(x)$, we may assume an *approximation* of the form

$$f(x) \approx \sum_{k=1}^{n} a_k \phi_k(x) \qquad (a < x < b), \tag{304}$$

and determine the coefficients a_k in such a way that the norm of the difference between the two members of (304) over (a, b) is as small as possible, and hence also such that

$$\Delta_n \equiv \left\| f(x) - \sum_{k=1}^{n} a_k \phi_k(x) \right\|^2 = \int_a^b \left[f(x) - \sum_{k=1}^{n} a_k \phi_k(x) \right]^2 dx = \min. \tag{305}$$

The approximation to be obtained, over the interval (a, b), is thus the best possible in the "least squares" sense.

If we think of a function $f(x)$ as a "vector" in function space, extending from an origin O to a "point" P in that space (see Problems 126–131), we can interpret (305) as choosing, from all points which can be attained by vectors of the form $\sum_{k=1}^{n} a_k \phi_k(x)$, that point whose "distance" from P is as small as possible.

Equation (305) is equivalent to the requirement that the expression

$$\Delta_n \equiv \int_a^b f^2 \, dx - 2 \sum_{k=1}^{n} a_k \int_a^b f \phi_k \, dx + \int_a^b \left[\sum_{k=1}^{n} a_k \phi_k(x) \right]^2 dx$$

take on a minimum value. But since the functions ϕ_k are orthonormal it follows that only squared terms in the last integrand have nonzero integrals. Hence, with the notation of (301), we obtain the result

$$\Delta_n = \int_a^b f^2 \, dx - 2 \sum_{k=1}^{n} a_k c_k + \sum_{k=1}^{n} a_k^2,$$

which can be put in the more convenient form

$$\Delta_n = \int_a^b f^2 \, dx - \sum_{k=1}^{n} c_k^2 + \sum_{k=1}^{n} (c_k - a_k)^2. \tag{306}$$

From this result it is clear that, since f and the c's are fixed, Δ_n takes a minimum value when the coefficients a_k are chosen such that

$$a_k = c_k. \tag{307}$$

Thus it follows that *the best approximation* (304) *in the least-squares sense is obtained when* a_k *is taken as the Fourier constant of* $f(x)$ *relative to* $\phi_k(x)$ *over* (a, b).

The squared norm of the deviation between $f(x)$ and its *best* n-term approximation of the form (304) is then obtained, by introducing (307) into (306), in the form

$$\left\| f(x) - \sum_{k=1}^{n} c_k \phi_k(x) \right\|^2_{\min} = \int_a^b f^2\, dx - \sum_{k=1}^{n} c_k^2. \tag{308}$$

From the definition (305) it follows that (308) cannot be *negative*; that is, we must have

$$\int_a^b f^2\, dx - \sum_{k=1}^{n} c_k^2 \geqq 0. \tag{309}$$

This relation is known as *Bessel's inequality*.

Suppose now that the dimension n of the orthonormal set $\phi_1, \phi_2, \cdots, \phi_n$ is increased without limit. The positive series in (309) must increase with n (unless the corresponding c's vanish) so that the norm of the error involved *decreases*, but since the sum cannot become greater than the fixed number $\int_a^b f^2\, dx$, we conclude that the series $\sum_1^\infty c_k^2$ always *converges* to some positive number not greater than $\int_a^b f^2\, dx$. However, there is no assurance that the limit to which this series converges will actually *equal* the value of this integral, so that the right-hand member of (308) then will tend to zero as n increases. That is, it is *not* sufficient merely to have a set of *infinitely many* mutually orthogonal functions.

In illustration, we may recall that the functions $\cos(k\pi x/l)$ $(k = 1, 2, 3, \cdots)$ constitute an orthogonal set of functions over the interval $(0, l)$; that is, we have the relation

$$\int_0^l \cos\frac{r\pi x}{l} \cos\frac{s\pi x}{l}\, dx = 0 \qquad (r \neq s).$$

The norm of each function over $(0, l)$ is $\sqrt{l/2}$, so that the functions

$$\phi_k(x) = \sqrt{\frac{2}{l}} \cos\frac{k\pi x}{l} \qquad (k = 1, 2, \cdots) \tag{310}$$

form an infinite orthonormal set over $(0, l)$. However, for the simple function $f(x) = 1$, the relevant Fourier constants are all *zeros*, since here

$$c_k = \sqrt{\frac{2}{l}} \int_0^l 1 \cdot \cos\frac{k\pi x}{l}\, dx = 0 \qquad (k = 1, 2, \cdots).$$

Hence, in this case the right-hand member of (308) is constantly equal to l, regardless of the value of n.

In a vector space of n dimensions, if we construct a set of n mutually orthogonal unit vectors, then the possibility of expressing any other vector as a linear combination of these vectors is a consequence of the fact that no other vector can be linearly independent of them; that is, there exists no vector in that space, other than the zero vector, which is simultaneously orthogonal to these n vectors. However, in function space (of infinitely many dimensions) the difficulty consists of the fact that a function may simultaneously be orthogonal to an *infinite number* of mutually orthogonal functions. Thus, in the above case, the function $f(x) = 1$ is orthogonal to *all* the functions in the set (310) over the interval $(0, l)$. However, it can be shown that any function which has this property differs from a constant by a *trivial* function, so that for the extended set

$$\sqrt{\frac{1}{l}}, \sqrt{\frac{2}{l}}\cos\frac{\pi x}{l}, \sqrt{\frac{2}{l}}\cos\frac{2\pi x}{l}, \cdots, \sqrt{\frac{2}{l}}\cos\frac{n\pi x}{l}, \cdots \tag{311}$$

there is no nontrivial function whose Fourier constants *all* vanish. Such a set of orthogonal functions is said to be *complete.*

It is easily verified that the set (311) is also orthogonal (but not normalized) over the larger interval $(-l, l)$. However, it is obvious that *any odd function* of x [for which $f(-x) = -f(x)$] will possess zero Fourier constants relative to this set, over that interval. To complete the set, it is found to be sufficient to add the functions

$$\sqrt{\frac{2}{l}}\sin\frac{\pi x}{l}, \sqrt{\frac{2}{l}}\sin\frac{2\pi x}{l}, \cdots, \sqrt{\frac{2}{l}}\sin\frac{n\pi x}{l}, \cdots. \tag{312}$$

Either of the sets (311) and (312) is complete over $(0, l)$, while the combination of the two sets is complete over $(-l, l)$ or, as a matter of fact, over *any* interval of length $2l$. These results are consequences of the known theory of *Fourier series.*

It can be shown that if the set of functions $\phi_1, \phi_2, \cdots, \phi_n, \cdots$ is *complete* in (a, b), then the right-hand member of (308) *does* indeed tend to zero for any function $f(x)$ which is of integrable square over that interval, so that

$$\sum_{k=1}^{\infty} c_k^2 = \int_a^b f^2\, dx \tag{313}$$

in that case. The proof of this relation, known as *Parseval's equality*, is involved. Furthermore, it is difficult in practice to *establish* the completeness of a given infinite orthogonal set of functions. For this reason, no attempt is made here to pursue the general theory.

However, it is important to realize that one further difficulty exists. Even though we prove that the right-hand member of (308) tends to zero

as n increases, so that

$$\lim_{n\to\infty}\int_a^b\left[f(x)-\sum_{k=1}^{n}c_k\phi_k(x)\right]^2dx=0, \tag{314}$$

we cannot then conclude that the integrand tends to zero everywhere in (a, b). That is, there may be no *specific* value of x in (a, b) for which we are then certain that the statement

$$f(x)=\lim_{n\to\infty}\sum_{k=1}^{n}c_k\phi_k(x)$$

is true. We know only that the mean square error in (a, b) tends to zero, and we say accordingly that if (314) is true then the series *converges in the mean* to $f(x)$.

In this case it is conventional to write

$$f(x)=\text{l.i.m.}_{n\to\infty}\sum_{k=1}^{n}c_k\phi_k(x),$$

where "l.i.m." is to be read "limit in the mean" and is to be carefully distinguished from the abbreviation "lim" which corresponds to the more usual limiting process.

However, if $f(x)$ is *continuous* throughout the interval (a, b), and if we can prove that the series $\sum_1^\infty c_k\phi_k(x)$ also represents a continuous function over that interval,* then the difference between these two functions is a continuous function with zero norm, and hence is indeed zero *everywhere* in (a, b), so that the series then converges to $f(x)$ in the true sense at each point of (a, b). Unfortunately, these conditions frequently are not fulfilled in practical cases.

While the knowledge that a series represents a function which differs from $f(x)$ in (a, b) by a trivial function is often all that is required (for such purposes as *integration*), it is nevertheless frequently desirable to determine whether or not the series actually represents $f(x)$ *at a given point*. The treatment of problems of this type is again beyond the scope of the present work.

The problems just discussed have been satisfactorily solved, in the mathematical literature, for a very large class of sets of orthogonal functions which frequently arise in practice. Certain known results are summarized, for convenient reference, in the following section.

It should be pointed out first that, in analogy with the corresponding situation in vector space, it is often desirable to modify somewhat the definition of the *norm* of a function. In particular, if $f(x)$ is a *complex*

* This will be the case, in particular, if the functions ϕ_k are continuous and if the series converges *uniformly* in the interval (a, b).

function of a real variable x, of the form $u(x) + i\,v(x)$, the norm of f is usually defined to be the nonnegative *real* quantity

$$\|f\|_H = (\bar{f}, f)^{1/2} = \left(\int_a^b \bar{f} f\, dx\right)^{1/2}, \tag{315}$$

where a bar indicates that the complex conjugate is to be taken. We speak of (315) as the *Hermitian norm* of f. The Hermitian scalar product of two complex functions f and g is then defined to be one of the two different quantities $(\bar{f}, g)$ and $(f, \bar{g})$, these two quantities being complex conjugates. In particular, f and g are said to be *orthogonal in the Hermitian sense* if

$$(\bar{f}, g) = (f, \bar{g}) = 0. \tag{316}$$

In problems analogous to those discussed in Section 1.25, but involving *differential* equations, sets of real functions are often generated for which the members $\phi_1, \phi_2, \cdots, \phi_n, \cdots$ are not orthogonal in the sense of (300), but for which a relation holds of the form

$$\int_a^b r(x)\phi_i(x)\phi_j(x)\, dx = 0 \qquad (i \neq j), \tag{317}$$

where $r(x)$ is real, nontrivial, and *nonnegative* in (a, b).

We may define the left-hand member of (317) to be the *generalized* or *weighted* scalar product of ϕ_i and ϕ_j, relative to the *weighting function* $r(x)$. Finally, the norm of any function f relative to the function r is defined to be

$$\|f\|_r \equiv \left(\int_a^b r f^2\, dx\right)^{1/2}, \tag{318}$$

and a function with *unit* norm, so defined, is said to be *normalized* relative to $r(x)$. The weighted scalar product of f and g is conveniently indicated by the notation

$$(f, g)_r \equiv \int_a^b rfg\, dx. \tag{319}$$

✶ **1.29. Sturm-Liouville problems.** In this section we summarize briefly certain known results concerning sets of orthogonal functions generated by certain types of boundary-value problems involving *linear ordinary differential equations.*

A problem which consists of a homogeneous linear differential equation of the form

$$\frac{d}{dx}\left(p\frac{dy}{dx}\right) + qy + \lambda r y = 0, \tag{320}$$

together with *homogeneous boundary conditions* of a rather general type, prescribed at the end points of an interval (a, b), generally possesses a

nontrivial solution only if the parameter λ is assigned one of a certain set of permissible values. For such a value of λ, say $\lambda = \lambda_k$, the conditions of the problem are satisfied by an expression of the form $y = C\phi_k(x)$ where C is an arbitrary constant. The permissible values of λ are known as its *characteristic values* (or "eigenvalues") and the corresponding functions $\phi_k(x)$, which then satisfy the conditions of the problem when $\lambda = \lambda_k$, are known as the *characteristic functions* (or "eigenfunctions").

In most cases occurring in practice, the functions $p(x)$ and $r(x)$ are positive in the interval (a, b), except possibly at one or both of the end points.

If we define a *linear differential operator* of second order by the formal equation

$$\mathscr{L} = \frac{d}{dx}\left(p\frac{d}{dx}\right) + q, \tag{321}$$

the differential equation (320) takes the operational form

$$\mathscr{L}y + \lambda r y = 0, \tag{322}$$

and is seen to be analogous to equation (255) of Section 1.25.

We show next that, when suitable homogeneous boundary conditions are prescribed at the ends of an interval (a, b), the characteristic functions of the resulting problem have properties analogous to those discussed in Section 1.25. For this purpose, let $\phi_i(x)$ and $\phi_j(x)$ be two characteristic functions, satisfying the conditions of the problem in correspondence with *distinct* characteristic numbers λ_i and λ_j. We then have the relations

$$\frac{d}{dx}\left(p\frac{d\phi_i}{dx}\right) + q\phi_i + \lambda_i r\phi_i = 0 \tag{323a}$$

and

$$\frac{d}{dx}\left(p\frac{d\phi_j}{dx}\right) + q\phi_j + \lambda_j r\phi_j = 0. \tag{323b}$$

If we multiply (323a) by ϕ_j and (323b) by ϕ_i, and subtract the resultant equations from each other, there follows

$$\begin{aligned}(\lambda_j - \lambda_i)r\phi_i\phi_j &= \phi_j\frac{d}{dx}\left(p\frac{d\phi_i}{dx}\right) - \phi_i\frac{d}{dx}\left(p\frac{d\phi_j}{dx}\right)\\ &= \frac{d}{dx}\left[p\left(\phi_j\frac{d\phi_i}{dx} - \phi_i\frac{d\phi_j}{dx}\right)\right],\end{aligned} \tag{324}$$

and the result of integrating both members of (324) over the interval (a, b) takes the form

$$(\lambda_j - \lambda_i)\int_a^b r\phi_i\phi_j\,dx = \left[p\left(\phi_j\frac{d\phi_i}{dx} - \phi_i\frac{d\phi_j}{dx}\right)\right]_a^b. \tag{325}$$

Thus, since we have assumed that $\lambda_j \neq \lambda_i$, we conclude that if the specified boundary conditions require the right-hand member of (325) to vanish, then *the characteristic functions ϕ_i and ϕ_j are orthogonal relative to the weighting function $r(x)$:*

$$(\phi_i, \phi_j)_r \equiv \int_a^b r\phi_i\phi_j \, dx = 0 \qquad (\lambda_i \neq \lambda_j). \tag{326}$$

Appropriate boundary conditions which may be seen to give rise to this situation include the following:

1. At each end of the interval we may require that either y or dy/dx or a linear combination $\alpha y + \beta \, dy/dx$ *vanish.*

2. If it happens that $p(x)$ vanishes at $x = a$ or at $x = b$, we may require instead merely that y and dy/dx remain finite at that point, and impose one of the conditions 1 at the other point.

3. If it happens that $p(b) = p(a)$, we may require merely that $y(b) = y(a)$ and $y'(b) = y'(a)$.

In most practical cases [in particular, if p, q, and r are *regular** and both p and r *positive* throughout (a, b)], when the interval (a, b) is of *finite length* it is found that in each of the listed cases there exists an *infinite* set of distinct characteristic numbers $\lambda_1, \lambda_2, \cdots, \lambda_n, \cdots$. If also the function $q(x)$ is *nonpositive* in (a, b), and if

$$[p\phi_i\phi_i']_a^b \leqq 0, \tag{327}$$

the λ's are all *nonnegative.* Furthermore, except in the case of the periodicity condition 3, to each characteristic number there corresponds *one and only one* characteristic function, an arbitrary multiple of which satisfies all the specified conditions when λ is assigned the appropriate characteristic value. In case 3, *two* linearly independent characteristic functions generally correspond to each characteristic number. Such pairs of functions then can be orthogonalized, if this is desirable, by the Gram-Schmidt procedure.

A problem of the general type just considered is known as a *Sturm-Liouville problem.*

The importance of such problems stems from the known fact that the sets of orthogonal functions generated by these problems generally are *complete*, in the sense of the preceding section, and further, that a positive statement can be made in such cases concerning actual *convergence* of the series representation of a sufficiently well-behaved function $f(x)$ to the value of the function *at all points where $f(x)$ is continuous.*

In actual practice, it is often inconvenient to *normalize* the characteristic functions (so that their norm relative to r is unity). In such cases, the

* A function $f(x)$ is said to be *regular* at $x = x_0$ if it can be represented by a convergent power series over an interval including x_0.

coefficients in a series representation

$$f(x) = \sum_{k=1}^{\infty} C_k \phi_k(x) \qquad (a < x < b) \tag{328}$$

are given by the formula

$$C_i \int_a^b r\phi_i^2 \, dx = \int_a^b rf\phi_i \, dx \tag{329a}$$

or, symbolically,

$$C_i \, \|\phi_i\|_r^2 = (f, \phi_i)_r. \tag{329b}$$

This result is obtained formally by multiplying both sides of (328) by the product $r\phi_i$, integrating the results term-by-term over (a, b), and taking into account the orthogonality of the characteristic functions relative to the weighting function $r(x)$. We notice that (329b) reduces to the obvious generalization of (301) when $\|\phi_i\|_r = 1$. The theorem to which reference was made above can then be stated as follows:

Let the functions $p(x)$, $q(x)$, and $r(x)$ in (320) *be regular in the finite interval (a, b), let $p(x)$ and $r(x)$ be positive in that interval, including the end points, and let the homogeneous end conditions be such that* (327) *is satisfied. Then, if $f(x)$ is bounded and piecewise differentiable in (a, b), the series* (328) *converges to $f(x)$ at all points inside that interval where $f(x)$ is continuous,* and to the mean value $\frac{1}{2}[f(x+) + f(x-)]$ at any point where a finite jump occurs.*

While the stated conclusions follow also under even milder restrictions on $f(x)$, the conditions given here are satisfied by most functions arising in practice.

To illustrate this result, we may consider the differential equation

$$\frac{d^2y}{dx^2} + \lambda y = 0, \tag{330}$$

which is the special case of (320) in which $p(x) = r(x) = 1$ and $q(x) = 0$. If we consider the interval $(0, l)$, and impose the boundary conditions

$$y(0) = 0, \qquad y(l) = 0, \tag{331}$$

it is easily verified that the characteristic values of λ, for which this problem possesses a solution other than the trivial solution $y(x) \equiv 0$, are of the form $\lambda_k = k^2\pi^2/l^2$, where k is any positive integer, and that the corresponding

* The convergence is absolute and uniform in any *interior* subinterval which does not include a point of discontinuity as an interior or end point.

characteristic functions are given by

$$\phi_k(x) = \sin \frac{k\pi x}{l} \qquad (k = 1, 2, \cdots).$$

Thus we obtain in this way a derivation of the *Fourier sine-series* representation

$$f(x) = \sum_{k=1}^{\infty} C_k \sin \frac{k\pi x}{l} \qquad (0 < x < l), \tag{332}$$

where, with $r(x) = 1$, equation (329) determines the coefficients in the form

$$C_k = \frac{2}{l} \int_0^l f(x) \sin \frac{k\pi x}{l}\, dx. \tag{333}$$

In a similar way, the conditions $y'(0) = y'(l) = 0$ associated with (330) give rise to the Fourier *cosine-series* representation, while the periodicity conditions $y(-l) = y(l)$ and $y'(-l) = y'(l)$, relevant to the interval $(-l, l)$, lead to the *general* Fourier series representation over that interval, involving both sines and cosines of period $2l$.

By considering other appropriate special forms of (320), expansions in terms of *Bessel functions*, *Legendre polynomials*, and so forth, may be established. In certain of these cases the coefficient functions p, q, and r do not satisfy the requirements specified in the preceding theorem, but it has been found that the conclusions of the theorem are still valid.

Elementary discussions of such developments may be found in Reference 9 (Chapter 5). For more detailed treatments of these topics, References 2 and 12 are suggested.

In those cases when the interval (a, b) is of *infinite* length, or when other conditions of the stated theorem are violated, it frequently happens that the characteristic values of λ are no longer *discretely* distributed, but that all values of λ in some *continuous* range are characteristic values. In such cases, the superposition of characteristic functions is accomplished by *integration*, rather than summation. In particular, for the problems discussed relative to equation (330), it is found that *all positive values of* λ are characteristic values when the fundamental interval is of infinite length, and one is led to the *Fourier integral* representations. In certain other exceptional cases the characteristic values again may be discretely distributed, or there may be both *continuously* distributed and *discretely* distributed characteristic values of λ.

Finally, we remark that the preceding discussion can be generalized to apply to characteristic functions of boundary-value problems governed by certain linear ordinary differential equations of higher order, as well as to characteristic functions of two or more variables associated with certain *partial* differential equations. Similar problems are related to linear *integral equations* and to linear *difference equations*.

REFERENCES

1. Birkhoff, G., and S. Maclane: *A Survey of Modern Algebra*, The Macmillan Company, New York, 1950.
2. Courant, R., and D. Hilbert: *Methods of Mathematical Physics*, Interscience Publishers, Inc., New York, 1953.
3. Crout, P. D.: "A Short Method for Evaluating Determinants and Solving Systems of Linear Equations with Real or Complex Coefficients," *Trans. AIEE*, Vol. 60, pp. 1235–1241 (1941).
4. Dwyer, Paul S.: *Linear Computations*, John Wiley & Sons, Inc., New York, 1951.
5. Faddeeva, V. N. (Translated from Russian by Curtis D. Benster): *Computational Methods of Linear Algebra*, Dover Publications, Inc., New York, 1959.
6. Forsythe, George E.: "Solving Linear Algebraic Equations Can Be Interesting," *Bull. Amer. Math. Soc.*, Vol. 59, pp. 299–329 (1953).
7. Frazer, R. A., W. J. Duncan, and A. R. Collar: *Elementary Matrices*, Cambridge University Press, London, 1960.
8. Halmos, P. R.: *Finite Dimensional Vector Spaces*, D. Van Nostrand Company, Inc., Princeton, N.J., 1958.
9. Hildebrand, F. B.: *Advanced Calculus for Applications*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1962.
10. Hoffman, Kenneth, and Ray Kunze: *Linear Algebra*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1961.
11. Perlis, Sam: *Theory of Matrices*, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1952.
12. Titchmarsh, E. C.: *Eigenfunction Expansions*, The Clarendon Press, Oxford, 1946.
13. Turnbull, H. W., and A. C. Aitken: *An Introduction to the Theory of Canonical Matrices*, Blackie and Son, Ltd., London, 1932.
14. Zurmühl, R.: *Matrizen*, Springer-Verlag, Berlin, 1950.

PROBLEMS

Sections 1.1, 1.2.

1. Illustrate the use of the Gauss-Jordan reduction in obtaining the general solution of each of the following sets of equations:

$$\text{(a)}\quad \begin{aligned} x_1 + 2x_2 + 2x_3 &= 1,\\ 2x_1 + 2x_2 + 3x_3 &= 3,\\ x_1 - x_2 + 3x_3 &= 5. \end{aligned} \qquad \text{(b)}\quad \begin{aligned} 2x_1 + x_3 &= 4,\\ x_1 - 2x_2 + 2x_3 &= 7,\\ 3x_1 + 2x_2 &= 1. \end{aligned}$$

$$\text{(c)}\quad \begin{aligned} 2x_1 - x_2 &= 6,\\ -x_1 + 3x_2 - 2x_3 &= 1,\\ -2x_2 + 4x_3 - 3x_4 &= -2,\\ -3x_3 + 5x_4 &= 1. \end{aligned}$$

Section 1.3.

2. Evaluate the following matrix products:

$$\text{(a)}\ \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix}. \qquad \text{(b)}\ \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}\begin{bmatrix} 6 & -2 \\ -3 & 1 \end{bmatrix}.$$

$$\text{(c)}\quad [a_1 \quad a_2 \quad \cdots \quad a_n]\begin{Bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{Bmatrix}. \qquad \text{(d)}\quad \begin{Bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \end{Bmatrix}[a_1 \quad a_2 \quad \cdots \quad a_n].$$

$$\text{(e)}\ \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}. \qquad \text{(f)}\ \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix}.$$

3. If the product **A**(**B C**) is defined, show that it is of the form

$$[a_{ir}]([b_{rs}][c_{sj}]) = \left[\sum_r \sum_s a_{ir}b_{rs}c_{sj}\right]$$

and deduce that then **A**(**B C**) = (**A B**)**C**.

4. If **A B** = **C**, show that the columns of **C** are linear combinations of the columns of **A** and that the rows of **C** are linear combinations of the rows of **B**. [Show, for example, that

$$\text{(col. } i \text{ of C)} = b_{1i}\text{(col. 1 of A)} + b_{2i}\text{(col. 2 of A)} + \cdots.]$$

5. If **A** and **B** are $n \times n$ matrices, when is it true that

$$(\mathbf{A} + \mathbf{B})(\mathbf{A} - \mathbf{B}) = \mathbf{A}^2 - \mathbf{B}^2?$$

Give an example in which this relation does *not* hold.

6. Find the most general 2×2 matrix **A** such that $\mathbf{A}^2 = \mathbf{0}$.

7. Prove that, if two square matrices of order three are both symmetrically partitioned as in the text on page 9, then these matrices may be correctly multiplied by treating the submatrices as single elements.

8. It is required to determine values of the function

$$\Phi(x) = \int_a^b K(x, \xi)f(\xi)\, d\xi,$$

at the n points $x_1, x_2, \cdots, x_n$. Show that, if in each case the integral is approximated by the use of Simpson's rule, as a linear combination of the ordinates at N equally spaced points $\xi_1 = a, \xi_2, \cdots, \xi_{N-1}, \xi_N = b$, where N is odd, the calculations can be arranged in the matrix form

$$\begin{Bmatrix} \Phi_1 \\ \Phi_2 \\ \cdot \\ \cdot \\ \cdot \\ \Phi_n \end{Bmatrix} \approx \frac{b-a}{3N-3} \begin{bmatrix} K_{11} & K_{12} & \cdots & K_{1N} \\ K_{21} & K_{22} & \cdots & K_{2N} \\ K_{31} & K_{32} & \cdots & K_{3N} \\ \cdots & \cdots & \cdots & \cdots \\ K_{n1} & K_{n2} & \cdots & K_{nN} \end{bmatrix} \begin{Bmatrix} f_1 \\ 4f_2 \\ 2f_3 \\ \cdot \\ \cdot \\ \cdot \\ f_N \end{Bmatrix},$$

where $\Phi_i \equiv \Phi(x_i)$, $K_{ij} \equiv K(x_i, \xi_j)$, and $f_j \equiv f(\xi_j)$.

9. Apply the procedure of Problem 8 to the approximate evaluation of the integral

$$\Phi(x) = \int_0^1 \sqrt{x^2 + \xi^2} \sin \pi\xi \, d\xi,$$

for $x = 0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$, and 1, with $N = 5$. Retain three significant figures in the calculation.

Section 1.4.

10. Prove, by direct expansion or otherwise, that

$$|\mathbf{A}\,\mathbf{B}| = |\mathbf{A}|\,|\mathbf{B}|$$

when **A** and **B** are square matrices of order two.

11. Determine those values λ for which the following set of equations may possess a nontrivial solution:

$$\begin{aligned} 3x_1 + x_2 - \lambda x_3 &= 0, \\ 4x_1 - 2x_2 - 3x_3 &= 0, \\ 2\lambda x_1 + 4x_2 + \lambda x_3 &= 0. \end{aligned}$$

For each permissible value of λ, determine the most general solution.

12. Show that the equation of the straight line $ax + by + c = 0$ which passes through the points (x_1, y_1) and (x_2, y_2) can be written in the form

$$\begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = 0.$$

13. Express the requirement, that four points (x_i, y_i) $(i = 1, 2, 3, 4)$ lie simultaneously on a conic of the form $ax^2 + bxy + cy^2 + d = 0$, in terms of the vanishing of a determinant.

Section 1.5.

14. If **A** and **B** commute, prove that $\mathbf{A}^T$ and $\mathbf{B}^T$ also commute.

15. A *symmetric* matrix $\mathbf{A} = [a_{ij}]$ is a square matrix for which $a_{ji} = a_{ij}$.

(a) Show that $\mathbf{A}^T = \mathbf{A}$ if and only if **A** is symmetric.

(b) Let **A** and **B** represent symmetric matrices of order n. Prove that **A B** is also symmetric if and only if **A** and **B** are commutative.

16. Verify that, if **A** and **B** are square matrices of order two, there follows Adj (**A B**) = (Adj **B**)(Adj **A**).

17. Let **A** and **B** represent diagonal matrices of order n.

(a) Prove that **A B** is also a diagonal matrix.

(b) Prove that **B A** = **A B**.

18. Evaluate each of the following sums:

(a) $\sum_{k=1}^{n} \delta_{ik}\delta_{kj}$.

(b) $\sum_{k=1}^{n} d_k\delta_{ik}\delta_{kj}$,

(c) $\sum_{k=1}^{n}\sum_{l=1}^{n} d_k\delta_{ik}\delta_{kl}\delta_{lj}$.

Section 1.6.

19. (a) If **A B** = **0** and **A** is nonsingular, prove that **B** = **0**.

(b) If **A B** = **0** and **B** is nonsingular, prove that **A** = **0**.

20. If **A B** = **A C**, where **A** is a square matrix, when does it necessarily follow that **B** = **C**? Give an example in which this conclusion does *not* follow.

21. Determine the elements of $\mathbf{A}^T$, Adj **A**, and $\mathbf{A}^{-1}$ when

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

22. Determine the elements of the matrix **M** such that **A M B** = **C** when

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

23. If $\mathbf{D} = [d_i\delta_{ij}]$ is a nonsingular diagonal matrix, prove that its inverse is given by

$$\mathbf{D}^{-1} = \left[\frac{1}{d_i}\delta_{ij}\right].$$

24. (a) Prove that A(Adj A) = **0** if **A** is singular, and illustrate by an example.

(b) Prove that $|\text{Adj A}| = |\mathbf{A}|^{n-1}$ (**A** of order n) and illustrate by an example.

25. If $|\mathbf{A}| \neq 0$, prove that $|\mathbf{A}^{-1}| = |\mathbf{A}|^{-1}$.

26. If **A** and **B** commute and **B** is nonsingular, prove that **A** and $\mathbf{B}^{-1}$ also commute.

27. For any matrix **A**, a matrix **P** is said to be a *left inverse* of **A** if $\mathbf{P}\,\mathbf{A} = \mathbf{I}$ and a matrix **Q** is said to be a *right inverse* of **A** if $\mathbf{A}\,\mathbf{Q} = \mathbf{I}$. Show that the matrix

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

has a two-parameter set of left inverses but no right inverse, whereas $\mathbf{A}^T$ has a two-parameter set of right inverses but no left inverse. [It is true (see Problem 31) that **A** cannot have *both* right and left inverses unless **A** is nonsingular.]

Section 1.7.

28. If **A** is an $m \times n$ matrix and if rank $\mathbf{A} \geqq m$, prove that rank $\mathbf{A} = m$.

29. Use the result of Problem 4 to prove that

$$\text{rank } (\mathbf{A}\,\mathbf{B}) \leqq \min (\text{rank } \mathbf{A}, \text{rank } \mathbf{B}).$$

30. (a) If $a_{ij} = r_i s_j$, prove that $\mathbf{A} = [a_{ij}]$ is of rank one or zero.

(b) If $\mathbf{A} = [a_{ij}]$ is of rank one, prove that a_{ij} can be written as $r_i s_j$. [Such a matrix is called a *dyad*.]

31. Suppose that **A** is an $m \times n$ matrix. Prove that if **A** has a left inverse **P** (see Problem 27) then the rank of **A** is n, and also that if **A** has a right inverse **Q** then the rank of **A** is m. [Use the result of Problem 29, noticing that **P** and **Q** are $n \times m$.] Hence deduce that *if* **A** *has both a left inverse and a right inverse then* **A** *must be nonsingular.*

Section 1.8.

32. If **A** is an $m \times n$ matrix, show that each of the three elementary operations on *rows* of **A** can be accomplished by *pre*multiplying **A** by a matrix **P**, where **P** is formed by performing that operation on corresponding rows of the *unit* matrix **I** of order m. In each case, show also that **P** is nonsingular.

33. If **A** is an $m \times n$ matrix, show that each of the elementary operations on *columns* of **A** can be accomplished by *post*multiplying **A** by a matrix **Q**, where **Q** is formed by performing that operation on corresponding columns of the unit matrix **I** of order n. In each case, show also that **Q** is nonsingular.

34. Show that any nonsingular matrix can be reduced to the unit matrix of the same order by use of only elementary *row* operations and also by use of only elementary *column* operations. [Consider the process used in the Gauss-Jordan reduction.] Thus deduce that *if* $\mathbf{B} = \mathbf{P}\,\mathbf{A}\,\mathbf{Q}$, *where* **P** *and* **Q** *are nonsingular, then* **B** *can be obtained from* **A** *by use of elementary row and column operations.*

Section 1.9.

35. (a) By investigating ranks of relevant matrices, show that the following set of equations possesses a one-parameter family of solutions:

$$\begin{aligned} 2x_1 - x_2 - x_3 &= 2, \\ x_1 + 2x_2 + x_3 &= 2, \\ 4x_1 - 7x_2 - 5x_3 &= 2. \end{aligned}$$

(b) Determine the general solution.

36. (a) Show that the set

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

can possess a nontrivial solution only if $\lambda = 1$ or $\lambda = -3$.

(b) Obtain the general solution in each case.

37. The matrix

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

is the augmented matrix of a system of linear algebraic equations. Determine for what fixed values of a and b (if any) the system possesses the following:

(a) A unique solution.
(b) A one-parameter solution.
(c) A two-parameter solution.
(d) No solution.

Section 1.10.

38. Determine the dimension of the vector space generated by each of the following sets of vectors:

(a) $\{1, 1, 0\}, \{1, 0, 1\}, \{0, 1, 1\}$.

(b) $\{1, 0, 0\}, \{0, 1, 0\}, \{0, 0, 1\}, \{1, 1, 1\}$.

(c) $\{1, 1, 1\}, \{1, 0, 1\}, \{1, 2, 1\}$.

39. Determine whether the vector $\{6, 1, -6, 2\}$ is in the vector space generated by the vectors $\{1, 1, -1, 1\}$, $\{-1, 0, 1, 1\}$, and $\{1, -1, -1, 0\}$.

40. (a) Determine the angle θ between the vectors

$$\mathbf{u} = \{1, 1, 1, 1\}, \qquad \mathbf{v} = \{1, 0, 0, 1\}.$$

(b) Determine the Hermitian angle θ_H between the complex vectors

$$\mathbf{u} = \{0, i, 1\}, \qquad \mathbf{v} = \{i, 1 + i, 1\}.$$

41. (a) If $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{Q}$ is an $n \times s$ matrix, and if the elements of the jth column of $\mathbf{Q}$ are considered to comprise the elements of a vector $\mathbf{v}_j$, show that the jth column of $\mathbf{A}\,\mathbf{Q}$ is the vector $\mathbf{A}\,\mathbf{v}_j$.

(b) If $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{P}$ is an $r \times m$ matrix, and if the elements of the ith row of $\mathbf{P}$ are considered to comprise the elements of a transposed vector $\mathbf{u}_i^T$, show that the ith row of $\mathbf{P}\,\mathbf{A}$ is the transposed vector $\mathbf{u}_i^T\,\mathbf{A}$.

(c) With the notation of parts (a) and (b), show that the typical element b_{ij} of the product $\mathbf{P}\,\mathbf{A}\,\mathbf{Q}$ is the scalar $\mathbf{u}_i^T\,\mathbf{A}\,\mathbf{v}_j$.

42. (a) Prove that if the Gramian of two real vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ vanishes, then $\mathbf{u}_1$ and $\mathbf{u}_2$ are linearly dependent. [Notice that, if $G = 0$, then the equations

$$c_1\mathbf{u}_1^T\,\mathbf{u}_1 + c_2\mathbf{u}_1^T\,\mathbf{u}_2 = 0 \qquad \text{and} \qquad c_1\mathbf{u}_2^T\,\mathbf{u}_1 + c_2\mathbf{u}_2^T\,\mathbf{u}_2 = 0$$

possess a nontrivial solution. Multiply the first equation by c_1, the second by c_2, add, and interpret the result.]

(b) Generalize the result of part (a) to the case of n vectors.

43. If $\mathbf{u}$ and $\mathbf{v}$ are real vectors, use the fact that the quantity

$$(\mathbf{u} + \lambda\,\mathbf{v})^T(\mathbf{u} + \lambda\,\mathbf{v}) = (\mathbf{u}, \mathbf{u}) + 2\lambda(\mathbf{u}, \mathbf{v}) + \lambda^2(\mathbf{v}, \mathbf{v})$$

is *nonnegative* for all real values of λ to deduce the *Schwarz inequality*:

$$|(\mathbf{u}, \mathbf{v})| \leqq (\mathbf{u}, \mathbf{u})^{1/2}(\mathbf{v}, \mathbf{v})^{1/2}.$$

44. Deal as in Problem 43 with the nonnegative real quantity

$$(\bar{\mathbf{u}} + \lambda\,\bar{\mathbf{v}})^T(\mathbf{u} + \lambda\,\mathbf{v}),$$

where $\mathbf{u}$ and $\mathbf{v}$ are complex vectors and λ is real, to deduce the generalized form of the *Schwarz inequality*:

$$\tfrac{1}{2}\,|(\bar{\mathbf{u}}, \mathbf{v}) + (\mathbf{u}, \bar{\mathbf{v}})| \leqq (\bar{\mathbf{u}}, \mathbf{u})^{1/2}\,(\bar{\mathbf{v}}, \mathbf{v})^{1/2}.$$

45. Establish the *parallelogram law*,

$$l^2(\mathbf{u} + \mathbf{v}) + l^2(\mathbf{u} - \mathbf{v}) = 2l^2(\mathbf{u}) + 2l^2(\mathbf{v}),$$

where $\mathbf{u}$ and $\mathbf{v}$ are real vectors, and interpret the result geometrically when $\mathbf{u}$ and $\mathbf{v}$ are three-dimensional vectors. Also show that this result remains true when $\mathbf{u}$ and $\mathbf{v}$ are complex if l is replaced by l_H.

Section 1.11.

46. Show that the set of equations

$$\begin{aligned} x_1 + x_2 + x_3 &= 3, \\ x_1 + x_2 - x_3 &= 1, \\ 3x_1 + 3x_2 - 5x_3 &= 1 \end{aligned}$$

possesses a one-parameter family of solutions, and verify directly that the vector $\mathbf{c}$ whose elements comprise the right-hand members is orthogonal to all vector solutions of the transposed homogeneous set of equations.

47. (a) Prove that if the set $\mathbf{A}\,\mathbf{x} = \mathbf{0}$ possesses an r-parameter set of nontrivial solutions, then the same is true of the transposed set $\mathbf{A}^T\,\mathbf{x}' = \mathbf{0}$, and conversely, when $\mathbf{A}$ is square.

(b) Interpret the statement at the end of Section 1.10 in the case when the transposed set $\mathbf{A}^T\,\mathbf{x}' = \mathbf{0}$ possesses no nontrivial solution.

48. Prove that no nonzero vector $\mathbf{v}$ can be in both the solution space of the system $\mathbf{A}\,\mathbf{x} = \mathbf{0}$ and in the row space of $\mathbf{A}$. [Assume that

$$\mathbf{v} = k_1\boldsymbol{\alpha}_1 + \cdots + k_r\boldsymbol{\alpha}_r = C_1\mathbf{u}_1 + \cdots + C_{n-r}\mathbf{u}_{n-r},$$

with the notation of Section 1.11, and form the scalar product of each $\mathbf{u}_i$ into the two equal right-hand members of this relation.]

Section 1.12.

49. Show that the problem

$$\begin{aligned} x_1 - 2x_2 &= \lambda x_1, \\ x_1 - x_2 &= \lambda x_2 \end{aligned}$$

does not possess real nontrivial solutions for any values of λ.

50. (a) Determine the characteristic numbers (λ_1, λ_2) and corresponding unit characteristic vectors $(\mathbf{e}_1, \mathbf{e}_2)$ of the matrix

$$\mathbf{A} = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}.$$

(b) Verify that $\mathbf{e}_1$ and $\mathbf{e}_2$ are orthogonal.

(c) If $\mathbf{v} = \{1, 1\}$, determine α_1 and α_2 so that

$$\mathbf{v} = \alpha_1\mathbf{e}_1 + \alpha_2\mathbf{e}_2.$$

(d) Use the results of part (a), together with equation (105), to obtain the solution of the following set of equations:

$$\begin{aligned} 5x_1 + 2x_2 &= \lambda x_1 + 2, \\ 2x_1 + 2x_2 &= \lambda x_2 + 1. \end{aligned}$$

Consider the exceptional cases separately.

51. (a) Suppose that the n characteristic vectors of the real symmetric matrix $\mathbf{A}$ are *not* normalized (reduced to unit length). If they are denoted by $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$, show that (105) must be replaced by the equation

$$\mathbf{x} = \sum_{k=1}^{n} \frac{(\mathbf{u}_k, \mathbf{c})}{\lambda_k - \lambda} \frac{\mathbf{u}_k}{(\mathbf{u}_k, \mathbf{u}_k)}.$$

(b) Verify this result in the case of Problem 50(d).

52. Suppose that a sequence of approximations $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(r)}, \cdots$ to the vector solution $\mathbf{x}$ of the equation

$$\mathbf{x} = \mathbf{M}\,\mathbf{x} + \mathbf{c}$$

is generated by the recurrence formula

$$\mathbf{x}^{(r+1)} = \mathbf{M}\,\mathbf{x}^{(r)} + \mathbf{c} \qquad (r = 0, 1, 2, \cdots),$$

where $\mathbf{x}^{(0)}$ is an initial approximation, and assume that $\mathbf{M}$ is an $n \times n$ real, symmetric matrix. Let the characteristic numbers of $\mathbf{M}$ be denoted by $\lambda_1, \cdots, \lambda_n$, with corresponding mutually orthogonal characteristic vectors $\mathbf{u}_1, \cdots, \mathbf{u}_n$.

(a) If $\boldsymbol{\epsilon}^{(r)} \quad \mathbf{x} - \mathbf{x}^{(r)}$, show that

$$\boldsymbol{\epsilon}^{(r+1)} = \mathbf{M}\,\boldsymbol{\epsilon}^{(r)} \qquad (r = 0, 1, 2, \cdots).$$

(b) Noticing that the initial error vector $\boldsymbol{\epsilon}^{(0)}$ can be expressed in the form

$$\boldsymbol{\epsilon}^{(0)} = \sum_{k=1}^{n} \alpha_k \mathbf{u}_k,$$

for *some* values of $\alpha_1, \cdots, \alpha_n$, show that there follows

$$\boldsymbol{\epsilon}^{(r)} = \sum_{k=1}^{n} \alpha_k \lambda_k{}^r \mathbf{u}_k.$$

(c) Deduce that $\mathbf{x}^{(r)}$ converges to $\mathbf{x}$ as $r \to \infty$, regardless of the form of the initial approximation $\mathbf{x}^{(0)}$, if and only if $|\lambda_k| < 1$ $(k = 1, 2, \cdots, n)$. [This result can also be established, by a less simple argument, when $\mathbf{M}$ is not necessarily symmetric.]

53. To illustrate the result of Problem 52, suppose that the system

$$\begin{aligned} x_1 - \alpha x_2 &= c_1, \\ -\alpha x_1 + x_2 - \alpha x_3 &= c_2, \\ -\alpha x_2 + x_3 &= c_3 \end{aligned}$$

is solved iteratively by use of the relations

$$\begin{aligned} x_1{}^{(r+1)} &= \alpha x_2{}^{(r)} + c_1, \\ x_2{}^{(r+1)} &= \alpha(x_1{}^{(r)} + x_3{}^{(r)}) + c_2, \\ x_3{}^{(r+1)} &= \alpha x_2{}^{(r)} + c_3. \end{aligned}$$

Show that convergence is guaranteed if and only if $|\alpha| < \sqrt{2}/2$.

Section 1.13.

54. Construct a set of three mutually orthogonal unit vectors which are linear combinations of the vectors $\{1, 0, 2, 2\}$, $\{1, 1, 0, 1\}$, and $\{1, 1, 0, 0\}$.

55. Prove that the vector $\mathbf{v} = \{2, 1, 2, 0\}$ is in the space generated by the three vectors defined in Problem 54 and express $\mathbf{v}$ as a linear combination of the selected vectors $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$.

Sections 1.14, 1.15.

56. If A is a (homogeneous) quadratic form in $x_1, x_2, \cdots, x_n$, prove that

$$A = \frac{1}{2} \sum_{k=1}^{n} x_k \frac{\partial A}{\partial x_k}.$$

57. Construct an orthonormal modal matrix $\mathbf{Q}$ corresponding to the matrix

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

and verify that $\mathbf{Q}^T \mathbf{A} \mathbf{Q} = [\lambda_i \delta_{ij}]$. (Notice the footnote on page 31.)

58. Reduce the quadratic form $A = x_1^2 + 3x_2^2 + 3x_3^2 - 2x_2x_3$ to a canonical form by making an appropriate change in variables, $\mathbf{x} = \mathbf{Q}\,\mathbf{x}'$, where $\mathbf{Q}$ is an orthogonal matrix.

59. Let $\mathbf{M}$ represent a modal matrix of a real symmetric matrix $\mathbf{A}$, the modal columns of which are orthogonal, but not necessarily reduced to unit length. If the characteristic vectors whose elements comprise successive columns of $\mathbf{M}$ are denoted by $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$, show that

$$\mathbf{M}^T \mathbf{M} = \begin{bmatrix} l_1^2 & 0 & \cdots & 0 \\ 0 & l_2^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & l_n^2 \end{bmatrix}$$

and

$$\mathbf{M}^T \mathbf{A} \mathbf{M} = \begin{bmatrix} \lambda_1 l_1^2 & 0 & \cdots & 0 \\ 0 & \lambda_2 l_2^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_n l_n^2 \end{bmatrix},$$

where $l_i^2 = \mathbf{u}_i^T \mathbf{u}_i$. Hence deduce also that the form $A = \mathbf{x}^T \mathbf{A}\,\mathbf{x}$ is reduced to the form

$$\mathbf{A} = \lambda_1 l_1^2 \mathbf{x}_1'^2 + \lambda_2 l_2^2 \mathbf{x}_2'^2 + \cdots + \lambda_n l_n^2 \mathbf{x}_n'^2$$

by the change in variables $\mathbf{x} = \mathbf{M}\,\mathbf{x}'$. [Notice that this form reduces to the canonical form (127) if the vectors $\mathbf{u}_i$ are normalized, so that $\mathbf{M}$ is an orthogonal matrix.]

60. (a) Prove that the product of two orthogonal matrices is also an orthogonal matrix.

(b) Prove that the inverse of an orthogonal matrix is also an orthogonal matrix.

Section 1.16.

61. Let $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$. Determine nonsingular matrices $\mathbf{P}$ and $\mathbf{Q}$ such that $\mathbf{P}\,\mathbf{A}\,\mathbf{Q} = \mathbf{B}$, where $\mathbf{B}$ is obtained by interchanging the two rows of $\mathbf{A}$ and then adding twice the first column to the third column. (See also Problems 32 and 33.)

Section 1.17.

62. (a) Determine the characteristic numbers (λ_1, λ_2) and corresponding Hermitian unit characteristic vectors $(\mathbf{e}_1, \mathbf{e}_2)$ of the problem

$$9x_1 + (2 + 2i)x_2 = \lambda x_1,$$
$$(2 - 2i)x_1 + 2x_2 = \lambda x_2,$$

where $i^2 = -1$, and verify that $\mathbf{e}_1$ and $\mathbf{e}_2$ are orthogonal in the Hermitian sense.

(b) If $\mathbf{H}$ is the coefficient matrix of the system of part (a) and $\mathbf{U}$ is the orthonormal modal matrix made up of $\mathbf{e}_1$ and $\mathbf{e}_2$, verify that

$$\bar{\mathbf{U}}^T \mathbf{H}\, \mathbf{U} = [\lambda_i \delta_{ij}].$$

(c) If $\mathbf{v} = \{1 + i, 1\}$, determine α_1 and α_2 such that

$$\mathbf{v} = \alpha_1 \mathbf{e}_1 + \alpha_2 \mathbf{e}_2,$$

where $\mathbf{e}_1$ and $\mathbf{e}_2$ are the vectors determined in part (a).

63. Describe the modification of the Gram-Schmidt orthogonalization procedure of Section 1.13 which applies when orthogonality and unit length are defined in the Hermitian sense.

64. Prove that an orthonormal modal matrix $\mathbf{U}$ of a Hermitian matrix $\mathbf{H}$ is a unitary matrix [i.e., that equation (149) is satisfied].

65. Prove that (148) holds when $\mathbf{H}$ is Hermitian, $\mathbf{U}$ is a corresponding orthonormal modal matrix, and λ_k is the kth characteristic number of $\mathbf{H}$.

66. (a) Show that, if a matrix is both unitary and Hermitian, it must satisfy the equation $\mathbf{U}^2 = \mathbf{I}$.

(b) Prove that any matrix of order two, of this type, is either the positive or negative unit matrix, or else is of the form

$$\mathbf{U} = \begin{bmatrix} a & r\, e^{i\alpha} \\ r\, e^{-i\alpha} & -a \end{bmatrix},$$

where a, r, and α are real and $a^2 + r^2 = 1$.

67. A matrix $\mathbf{N}$ is called a *normal matrix* if it commutes with its conjugate transpose, so that

$$\mathbf{N}\, \bar{\mathbf{N}}^T = \bar{\mathbf{N}}^T \mathbf{N}.$$

(Notice that hence, in particular, any *real* matrix $\mathbf{A}$ for which $\mathbf{A}^T = \pm\mathbf{A}$ or any *complex* matrix $\mathbf{B}$ for which $\bar{\mathbf{B}}^T = \pm\mathbf{B}$ is normal.)

(a) Prove that if $\mathbf{N}$ is normal, then $l_{II}(\mathbf{N}\,\mathbf{v}) = l_{II}(\bar{\mathbf{N}}^T \mathbf{v})$ and hence deduce that *if* $\mathbf{N}$ *is normal, then the relations*

$$\mathbf{N}\,\mathbf{v} = \mathbf{0}, \qquad \bar{\mathbf{N}}^T \mathbf{v} = \mathbf{0}$$

imply each other.

(b) Prove that $\mathbf{N} - \lambda\,\mathbf{I}$ is normal if $\mathbf{N}$ is normal.

(c) By replacing $\mathbf{N}$ by $\mathbf{N} - \lambda_k \mathbf{I}$ and $\mathbf{v}$ by $\mathbf{u}_k$ in the result of (a), prove that $\mathbf{N}\,\mathbf{u}_k = \lambda_k \mathbf{u}_k$ implies $\bar{\mathbf{N}}^T \mathbf{u}_k = \bar{\lambda}_k \mathbf{u}_k$, so that *if* $\mathbf{u}_k$ *is a characteristic vector of a normal matrix* $\mathbf{N}$, *corresponding to* λ_k, *then* $\mathbf{u}_k$ *also is a characteristic vector of* $\bar{\mathbf{N}}^T$, *corresponding to* $\bar{\lambda}_k$.

(d) Prove that *two characteristic vectors of a normal matrix* $\mathbf{N}$, *corresponding to distinct characteristic numbers, are orthogonal in the Hermitian sense.* [Use the relations $\mathbf{N}\,\mathbf{u}_1 = \lambda_1 \mathbf{u}_1$ and $\bar{\mathbf{N}}^T \mathbf{u}_2 = \bar{\lambda}_2 \mathbf{u}_2$.]

68. Assuming that a normal matrix possesses s linearly independent characteristic vectors in correspondence with a characteristic number of multiplicity s (see Problem 71), prove that *if* $\mathbf{N}$ *is normal, then an associated orthonormal modal matrix* $\mathbf{U}$ *is unitary and has the property that*

$$\bar{\mathbf{U}}^T \mathbf{N}\,\mathbf{U} = [\lambda_i \delta_{ij}],$$

where λ_i *is the ith characteristic number of* $\mathbf{N}$.

Section 1.18.

69. Show that if the first two columns of an orthogonal matrix $\mathbf{Q}$ comprise the elements of two unit characteristic vectors of a real symmetric matrix $\mathbf{A}$, then $\mathbf{Q}^{-1}\mathbf{A}\,\mathbf{Q}$ is of the form

$$\begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \alpha_{33} & \cdots & \alpha_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \alpha_{3n} & \cdots & \alpha_{nn} \end{bmatrix},$$

where λ_1 and λ_2 are the characteristic numbers corresponding respectively to the two characteristic vectors.

70. Modify the argument of Section 1.18 to deal with a *Hermitian* matrix $\mathbf{H}$. (Notice that $\bar{\mathbf{U}}^T \mathbf{H}\,\mathbf{U}$ is also Hermitian.)

71. Modify the argument of Section 1.18 to deal with a general *normal* matrix $\mathbf{N}$ (see Problems 67 and 68). [Use the fact that $\bar{\mathbf{N}}^T \mathbf{e}_1 = \bar{\lambda}_1 \mathbf{e}_1$ to show that, when $\mathbf{e}_1$ is the first column of $\mathbf{U}$, both $\bar{\mathbf{U}}^T \mathbf{N}\,\mathbf{U}$ and its conjugate transpose will have zeros following the first element throughout the first column.]

Section 1.19.

72. Determine whether the real form

$$A = x_1^2 + 2x_2^2 + x_3^2 - 2x_1x_2 + 2x_2x_3$$

is positive definite, by examining the characteristic numbers of the associated matrix.

73. Determine a real change in variables which reduces the forms

$$A = 3x_1^2 - 2x_1x_2 + 3x_2^2, \qquad B = 2x_1^2 + 2x_2^2$$

simultaneously to the canonical forms

$$A = \lambda_1\alpha_1^2 + \lambda_2\alpha_2^2, \qquad B = \alpha_1^2 + \alpha_2^2,$$

by using the methods of Section 1.19.

Section 1.20.

74. Find the *sum* and *product* of all characteristic numbers of the matrix

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

75. Determine whether the matrix **A** of Problem 74 is positive definite.

76. A real symmetric matrix **A** is said to be *negative definite* if its associated quadratic form $\mathbf{x}^T\mathbf{A}\,\mathbf{x}$ is nonpositive for all real **x**, and is zero only when $\mathbf{x} = \mathbf{0}$. State conditions under which this situation exists, (a) in terms of the characteristic numbers of **A**, and (b) in terms of the discriminants of **A**. (Notice that **A** is negative definite if and only if $-\mathbf{A}$ is positive definite.)

77. Determine for what values of c, if any, it is true that

$$x^2 + y^2 + z^2 \geqq 2c(xy + yz + zx)$$

for all real values of x, y, and z, with equality holding only when $x = y = z = 0$.

78. Prove that if **P** is a nonsingular real matrix then $\mathbf{A} = \mathbf{P}^T\mathbf{P}$ is positive definite.

79. Prove that if **A** is real, symmetric, and positive definite then there exists a nonsingular real matrix **P** such that $\mathbf{A} = \mathbf{P}^T\mathbf{P}$. [First show that **A** can be reduced to a form $\mathbf{R}^T\mathbf{D}\,\mathbf{R}$, where $\mathbf{D} = \operatorname{diag}(\lambda_1, \cdots, \lambda_n)$ and $|\mathbf{R}| \neq 0$, and then show that **D** can be rewritten appropriately as $\mathbf{G}^T\mathbf{G}$.]

Section 1.21.

80. A geometrical vector $\mathscr{x}$ is represented by the numerical vector $\mathbf{x} = \{1, 1, 1\}$ in terms of components along unit vectors $\mathbf{i}_1$, $\mathbf{i}_2$, and $\mathbf{i}_3$ coinciding with the axes of a rectangular $x_1x_2x_3$ coordinate system. If new axes are chosen in such a way that the new unit vectors are related to the original ones by the equations

$$\mathbf{i}_1' = \frac{\mathbf{i}_1 + \mathbf{i}_2}{\sqrt{2}}, \qquad \mathbf{i}_2' = \frac{\mathbf{i}_1 - \mathbf{i}_2}{\sqrt{2}}, \qquad \mathbf{i}_3' = \mathbf{i}_3,$$

determine the representation $\mathbf{x}'$ of $\mathscr{x}$ in terms of components of $\mathscr{x}$ along the new axes. Show also that the new coordinate system is also rectangular.

81. A numerical vector $\mathbf{y}$, representing $\mathscr{y}$, is related to the numerical vector $\mathbf{x}$ of Problem 80 by the equation $\mathbf{y} = \mathbf{A}\,\mathbf{x}$, where

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

Determine the components of the representation $\mathbf{y}'$ in the new system, *first*, by determining the components of $\mathbf{y}$ and transforming them directly, and *second*, by using equation (199) in connection with the result of Problem 80.

82. Prove that if the new unit vectors of (190) are mutually orthogonal, then the matrix (196) is an orthogonal matrix.

83. (a) Show that an orthogonal matrix of order two is necessarily of one of the following two types:

$$\mathbf{Q}^{(+)} = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}, \qquad \mathbf{Q}^{(-)} = \begin{bmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{bmatrix}.$$

[Notice that $|\mathbf{Q}^{(+)}| = +1$, and $|\mathbf{Q}^{(-)}| = -1$.]

(b) If $\mathbf{x}$ and $\mathbf{x}'$ are considered as two distinct vectors referred to the same axes, and are related by the equation $\mathbf{x} = \mathbf{Q}\,\mathbf{x}'$, verify that $\mathbf{x}$ is rotated into $\mathbf{x}'$ through the angle α by a positive (counterclockwise) rotation if $\mathbf{Q} = \mathbf{Q}^{(+)}$.

(c) If $\mathbf{x}$ and $\mathbf{x}'$ are considered as comprising the components of representations of the same geometrical vector, referred to original and rotated axes, respectively, verify that the coordinate transformation $\mathbf{x} = \mathbf{Q}^{(+)}\mathbf{x}'$ corresponds to a negative rotation of the original axes, through the angle α.

(d) If $\mathbf{Q} = \mathbf{Q}^{(-)}$ in parts (b) and (c), verify that the transformations then each involve a *reversed* rotation combined with a suitable *reflection.*

Section 1.22.

84. If $\mathbf{A}$ is a symmetric $n \times n$ matrix with n distinct characteristic numbers λ_i, show that any polynomial $P(\mathbf{A})$ can be expressed in the form

$$P(\mathbf{A}) = c_1\mathbf{A}^{n-1} + c_2\mathbf{A}^{n-2} + \cdots + c_{n-1}\mathbf{A} + c_n\mathbf{I}$$

where the c's are determined by the n simultaneous linear equations

$$P(\lambda_i) = \sum_{k=1}^{n} c_k\lambda_i^{n-k} \qquad (i = 1, 2, \cdots, n).$$

[In some cases this direct determination is more convenient than the use of Sylvester's formula (226).]

85. Let $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ and $\mathbf{B} = \mathbf{A}^5 - 3\mathbf{A}^4 + 2\mathbf{A} - \mathbf{I}$.

(a) Determine the characteristic numbers and corresponding characteristic vectors of **B**.

(b) Determine whether **B** is positive definite.

(c) Determine the elements of $\mathbf{A}^{100}$.

86. Suppose that **A** is real and symmetric of order two, with a repeated characteristic number $\lambda_1 = \lambda_2$.

(a) Obtain from (230) the evaluation

$$e^{\mathbf{A}} = e^{\lambda_1}\,\mathbf{A} - (\lambda_1 - 1)e^{\lambda_1}\,\mathbf{I}.$$

(b) Prove that **A** must in this case be a scalar matrix $\mathbf{A} = k\,\mathbf{I}$, and show that the evaluation of part (a) reduces to

$$e^{k\mathbf{I}} = e^{k}\,\mathbf{I}.$$

87. Suppose that the elements of a matrix $\mathbf{A}(t) = [a_{ij}(t)]$ are differentiable functions of a variable t.

(a) From the definition

$$\frac{d\mathbf{A}(t)}{dt} = \lim_{\Delta t \to 0} \frac{\mathbf{A}(t + \Delta t) - \mathbf{A}(t)}{\Delta t} \equiv \lim_{\Delta t \to 0} \frac{\Delta \mathbf{A}}{\Delta t},$$

prove that $d\mathbf{A}(t)/dt = [da_{ij}/dt]$.

(b) Prove that

$$\frac{d}{dt}(\mathbf{A}\,\mathbf{B}) = \frac{d\mathbf{A}}{dt}\,\mathbf{B} + \mathbf{A}\,\frac{d\mathbf{B}}{dt}.$$

(c) Specialize the result of part (b) in the case when $\mathbf{B} = \mathbf{A}$, and give an example to show that $d\mathbf{A}^2/dt \neq 2\mathbf{A}\,d\mathbf{A}/dt$ in general.

88. (a) If **A** is a real symmetric matrix, verify that the differential equation

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\,\mathbf{x}$$

is satisfied by $\mathbf{x} = e^{t\mathbf{A}}\,\mathbf{c}$, where **c** is a constant vector.

(b) Use this result and an appropriate modification of equation (230) to solve the system

$$\frac{dx_1}{dt} = x_1 + 2x_2,$$

$$\frac{dx_2}{dt} = 2x_1 + x_2,$$

subject to the initial conditions $\{x_1(0), x_2(0)\} = \{c_1, c_2\}$.

89. (a) Show that if **A** is a symmetric matrix of order n, with distinct characteristic numbers, then

$$\mathbf{A}^N = \sum_{k=1}^{n} \lambda_k{}^N\, Z_k(\mathbf{A}),$$

where Z_k is defined by (227) and N is a positive integer.

(b) Let λ_n be the dominant characteristic number (i.e., the characteristic number with largest absolute value). Noticing that for sufficiently large N the nth term in the preceding sum will then predominate, show that if $\mathbf{x}$ is an arbitrary vector there follows

$$\mathbf{A}^N \mathbf{x} \approx {\lambda_n}^N \mathbf{v}, \qquad \mathbf{A}^{N+1} \mathbf{x} \approx {\lambda_n}^{N+1} \mathbf{v},$$

where $\mathbf{v} = [Z_n(\mathbf{A})]\mathbf{x}$, when N is large, unless it happens that $\mathbf{v} = \mathbf{0}$.

(c) Hence deduce that, in general, if an arbitrary vector $\mathbf{x}$ is premultiplied repeatedly by a symmetric matrix $\mathbf{A}$, the vector obtained after $N + 1$ such multiplications is approximately λ_n times that obtained after N multiplications, where λ_n is the dominant characteristic number of $\mathbf{A}$, and hence also that the vectors obtained after successive multiplications tend, in general, to become multiples of the characteristic vector associated with λ_n.

(d) Show that the exceptional case, in which the vector $\mathbf{x}$ is such that $[Z_n(\mathbf{A})]\mathbf{x} = \mathbf{0}$, will occur if $\mathbf{x}$ happens to be a characteristic vector of $\mathbf{A}$, corresponding to a characteristic number $\lambda_k \neq \lambda_n$, or if $\mathbf{x}$ is a linear combination of such vectors.

[A more complete treatment of this procedure, from a somewhat different point of view, is given in Section 1.23.]

Section 1.23.

90. Determine the dominant characteristic number and the corresponding characteristic vector for the system

$$\begin{aligned} x_1 + \ x_2 + \ x_3 &= \lambda x_1, \\ x_1 + 3x_2 + 3x_3 &= \lambda x_2, \\ x_1 + 3x_2 + 6x_3 &= \lambda x_4. \end{aligned}$$

(Retain slide-rule accuracy.)

91. Show that the iterative method does not converge to a characteristic vector if $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$, regardless of the initial approximation. Explain.

92. Investigate the application of the iterative method in the case of the matrix $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ -2 & -1 \end{bmatrix}$. Explain.

Section 1.24.

93. Determine the two largest characteristic numbers, and corresponding characteristic vectors, of the system

$$\begin{aligned} x_1 + \ x_2 + \ x_3 + \ x_4 &= \lambda x_1, \\ x_1 + 2x_2 + 2x_3 + 2x_4 &= \lambda x_2, \\ x_1 + 2x_2 + 3x_3 + 3x_4 &= \lambda x_3, \\ x_1 + 2x_2 + 3x_3 + 4x_4 &= \lambda x_4. \end{aligned}$$

(Retain slide-rule accuracy.)

94. Determine all characteristic numbers, and the corresponding characteristic vectors, of the system

$$\begin{aligned} x_1 - x_2 &= \lambda x_1, \\ -x_1 + 2x_2 - x_3 &= \lambda x_2, \\ -x_2 + 2x_3 - x_4 &= \lambda x_3, \\ -x_3 + x_4 &= \lambda x_4. \end{aligned}$$

(Retain slide-rule accuracy.)

95. Determine the smallest characteristic number for the system (241), without using the inversion (251), by making use of the procedure described in connection with (239) and (240). (Retain slide-rule accuracy.)

96. Suppose that the iterative method of Section 1.23 fails to converge for a real symmetric matrix **A**, so that λ_n and $-\lambda_n$ are both dominant characteristic numbers. Take $\lambda_n > 0$, and write $\lambda_{n-1} = -\lambda_n$.

Show that, if r is sufficiently large, the input in the rth cycle is given approximately by

$$\mathbf{x}^{(r)} \approx \mathbf{v}_n + \mathbf{v}_{n-1},$$

where $\mathbf{v}_n$ and $\mathbf{v}_{n-1}$ are constant multiples of the unit characteristic vectors relevant to λ_n and $\lambda_{n-1} \equiv -\lambda_n$, respectively, whereas the output is then given approximately by

$$\mathbf{y}^{(r)} \approx \lambda_n(\mathbf{v}_n - \mathbf{v}_{n-1}).$$

Show further that *if this output is taken as the input for the next cycle*, so that

$$\mathbf{x}^{(r+1)} = \mathbf{y}^{(r)},$$

there follows also

$$\mathbf{y}^{(r+1)} \approx \lambda_n^2(\mathbf{v}_n + \mathbf{v}_{n-1}),$$

so that λ_n can then be determined approximately by the relation

$$\mathbf{y}^{(r+1)} \approx \lambda_n^2 \mathbf{x}^{(r)},$$

after which approximations to $\mathbf{v}_n$ and $\mathbf{v}_{n-1}$ are given by

$$\mathbf{v}_n \approx \frac{1}{2}\left[\mathbf{x}^{(r)} + \frac{1}{\lambda_n}\mathbf{y}^{(r)}\right], \qquad \mathbf{v}_{n-1} \approx \frac{1}{2}\left[\mathbf{x}^{(r)} - \frac{1}{\lambda_n}\mathbf{y}^{(r)}\right],$$

when r is sufficiently large.

97. Illustrate the technique developed in Problem 96 in the case of the symmetric matrix $\mathbf{A} = \begin{bmatrix} -4 & 3 \\ 3 & 4 \end{bmatrix}$.

Section 1.25.

98. Prove that the characteristic numbers of the problem $\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{B}\,\mathbf{x}$ are real when **A** and **B** are real and symmetric, and either **A** or **B** is positive definite.

99. Determine the characteristic numbers and vectors of the problem $\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{B}\,\mathbf{x}$, where

$$\mathbf{A} = \begin{bmatrix} 5 & 2 \\ 2 & 3 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix},$$

and verify that the characteristic vectors are orthogonal relative to both **A** and **B**.

100. Construct an orthonormal modal matrix associated with Problem 99, where the normalization is relative to **B**.

101. Use the results of Problems 99 and 100 to determine a change in variables which reduces the quadratic forms

$$A = 5x_1^2 + 4x_1x_2 + 3x_2^2, \qquad B = x_1^2 + 2x_2^2$$

simultaneously to the canonical forms

$$A = \lambda_1\alpha_1^2 + \lambda_2\alpha_2^2, \qquad B = \alpha_1^2 + \alpha_2^2.$$

102. Show that the condition (282a) is equivalent to the condition

$$(\mathbf{x}^{(r)}, \mathbf{x}^{(r)})_{\mathbf{A}} \approx \lambda_n(\mathbf{x}^{(r)}, \mathbf{x}^{(r)})_{\mathbf{B}}$$

and that (282b) is equivalent to the condition

$$(\mathbf{y}^{(r)}, \mathbf{y}^{(r)})_{\mathbf{B}} \approx \lambda_n(\mathbf{x}^{(r)}, \mathbf{x}^{(r)})_{\mathbf{A}}.$$

103. Prove that any set of nonzero real vectors which are mutually orthogonal relative to a real positive definite matrix **B** is a linearly independent set. [Assume the contrary and deduce a contradiction.]

104. Consider the equation $\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{B}\,\mathbf{x}$, where

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 10 & 1 & 0 \\ 1 & 10 & 1 \\ 0 & 1 & 10 \end{bmatrix}.$$

(a) Determine the largest characteristic number λ_3 by an iterative process, after rewriting the equation in the form $\mathbf{B}^{-1}\,\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{x}$.

(b) Determine λ_3 by an iterative process without rewriting the equation in the form suggested in part (a).

(Retain slide-rule accuracy.)

105. (a) Determine the smallest characteristic number λ_1 for the equation of Problem 104 by an iterative process, after rewriting the equation in the form $\mathbf{A}^{-1}\,\mathbf{B}\,\mathbf{x} = \lambda^{-1}\,\mathbf{x}$.

(b) Determine λ_1 by an iterative process without rewriting the equation in the form suggested in part (a).

(Retain slide-rule accuracy.)

Section 1.26.

106. If $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ -2 & -1 \end{bmatrix}$, determine the characteristic vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ of the

problem $\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{x}$, and the characteristic vectors $\mathbf{u}_1'$ and $\mathbf{u}_2'$ of the associated problem $\mathbf{A}^T\,\mathbf{x}' = \lambda\,\mathbf{x}'$, and verify the validity of equation (289).

107. (a) Suppose that $\mathbf{A}$ posesses n distinct characteristic numbers $\lambda_1, \cdots, \lambda_n$, with corresponding characteristic vectors $\mathbf{u}_1, \cdots, \mathbf{u}_n$, and denote corresponding characteristic vectors of $\mathbf{A}^T$ by $\mathbf{u}_1', \cdots, \mathbf{u}_n'$. Obtain the solution of the problem $\mathbf{A}\,\mathbf{x} - \lambda\,\mathbf{x} = \mathbf{c}$ in the form

$$\mathbf{x} = \sum_{k=1}^{n} \frac{(\mathbf{u}_k', \mathbf{c})}{\lambda_k - \lambda} \frac{\mathbf{u}_k}{(\mathbf{u}_k', \mathbf{u}_k)},$$

when $\lambda \neq \lambda_1, \cdots, \lambda_n$. [Compare Problem 51.]

(b) Discuss the situation when λ assumes a characteristic value λ_p. (Notice also that this case is described by the result of replacing $\mathbf{A}$ by $\mathbf{A} - \lambda_p\mathbf{I}$ in the statement at the end of Section 1.11.)

108. With the terminology of Problem 107, use the result at the end of Section 1.9 to show that the elements of $\mathbf{u}_r$ are proportional to the cofactors of respective elements in any *row* of the matrix

$$\mathbf{A} - \lambda_r\mathbf{I} \equiv \begin{bmatrix} a_{11} - \lambda_r & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} - \lambda_r & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} - \lambda_r \end{bmatrix},$$

whereas the elements of $\mathbf{u}_r'$ are proportional to the cofactors of respective elements in any *column* of that matrix, if not all the relevant cofactors vanish. Verify this conclusion in the example of Problem 106.

109. Let $\mathbf{M}$ represent a modal matrix of *any* square matrix $\mathbf{A}$ of order n with n linearly independent characteristic vectors $\mathbf{u}_1, \cdots, \mathbf{u}_n$ corresponding to the respective characteristic numbers $\lambda_1, \cdots, \lambda_n$, where the $\mathbf{u}$'s which are used to construct $\mathbf{M}$ need be neither orthogonal nor of unit length and the λ's need be neither real nor distinct.

(a) By appropriately modifying the argument of equations (119)-(122) of Section 1.14, prove that

$$\mathbf{M}^{-1}\,\mathbf{A}\,\mathbf{M} = \mathbf{D},$$

where $\mathbf{D}$ is the diagonal matrix

$$\mathbf{D} = [\lambda_i\delta_{ij}],$$

so that $\mathbf{A}$ is thus diagonalized by a *similarity* transformation.

(b) Deduce that any such matrix $\mathbf{A}$ can be determined from its characteristic numbers and vectors by use of the relation

$$\mathbf{A} = \mathbf{M}\,\mathbf{D}\,\mathbf{M}^{-1}$$

and show also that

$$\mathbf{A}^r = \mathbf{M}\,\mathbf{D}^r\,\mathbf{M}^{-1}$$

when r is any nonnegative integer.

110. Consider the nonsymmetric matrix

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

(a) Show that $\mathbf{A}$ has a double characteristic number $\lambda_1 = \lambda_2 = 1$, corresponding to a single characteristic vector $\mathbf{u}_1$.

(b) Determine a matrix $\mathbf{P}$ such that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ is in the Jordan canonical form,

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},$$

by taking the vector $\mathbf{u}_1$ as the first column of $\mathbf{P}$ and determining the unknown elements of the second column of $\mathbf{P}$ in such a way that the upper right element of $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ is unity. Notice that infinitely many such $\mathbf{P}$'s are obtained.

111. Prove the following theorem: *If* $\mathbf{A}$ *is a real matrix with nonnegative elements and if the sum of the elements in each column* (or *in each row*) *of* $\mathbf{A}$ *is less than one, then the characteristic numbers of* $\mathbf{A}$ *are smaller than one in absolute value.* [Let $\mathbf{u} = \{u_1, \cdots, u_n\}$ be a characteristic vector corresponding to λ. Deduce from the relation $\lambda u_i = \sum_k a_{ik}u_k$ that $|\lambda|\,|u_i| \leqq \sum_k a_{ik}\,|u_k|$, and sum over i. Note that λ is also a characteristic number of $\mathbf{A}^T$.]

Section 1.27.

Determine the natural frequencies and natural modes of vibration of the mechanical system of Figure 1.1 in the following cases:

112. Assume $k_1 = 2k$, $k_2 = k_3 = k$; $m_1 = m_2 = m_3 = m$.

113. Assume $k_1 = 2k$, $k_2 = k_3 = k$; $m_1 = m_2 = m$, $m_3 = 2m$.

114. Assume $k_1 = 0$, $k_2 = k_3 = k$; $m_1 = m_2 = m_3 = m$.

115. Assume $k_1 = 0$, $k_2 = k_3 = k$; $m_1 = m_2 = m$, $m_3 = 2m$.

[In most physical problems of this type the fundamental mode (corresponding to the *smallest* natural frequency) is such that the initial approximation $\{1, 1, 1, \cdots, 1\}$ is a convenient one. In the highest natural mode, the successive masses generally tend to oscillate with opposite phases, so that the initial approximation $\{1, -1, 1, \cdots, \pm 1\}$ usually leads to more rapid convergence. In Problems 114 and 115, the system of masses and springs is unattached to a support, and the characteristic number associated with $\omega = 0$ corresponds to motion of the system as a rigid body.]

Minimal Properties of Characteristic Numbers.

116. Let $\mathbf{A}$ denote a real symmetric matrix of order n, with characteristic numbers $\lambda_1, \cdots, \lambda_n$, arranged in increasing *algebraic* order ($\lambda_1 \leqq \lambda_2 \leqq \cdots \leqq \lambda_n$), and corresponding normalized and orthogonalized characteristic vectors $\mathbf{e}_1, \cdots, \mathbf{e}_n$.

(a) If $\mathbf{x}$ is an arbitrary real vector with n components, and hence expressible in the form

$$\mathbf{x} = c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + \cdots + c_n\mathbf{e}_n = \sum_{k=1}^{n} c_k\mathbf{e}_k,$$

establish the relations

$$\mathbf{x}^T\mathbf{x} = c_1^2 + c_2^2 + \cdots + c_n^2 = \sum_{k=1}^{n} c_k^2,$$

$$\mathbf{A}\mathbf{x} = \lambda_1 c_1\mathbf{e}_1 + \lambda_2 c_2\mathbf{e}_2 + \cdots + \lambda_n c_n\mathbf{e}_n = \sum_{k=1}^{n} \lambda_k c_k\mathbf{e}_k,$$

and

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \lambda_1 c_1^2 + \lambda_2 c_2^2 + \cdots + \lambda_n c_n^2 = \sum_{k=1}^{n} \lambda_k c_k^2.$$

(b) Deduce that

$$\frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \frac{\lambda_1 c_1^2 + \lambda_3 c_2^2 + \cdots + \lambda_n c_n^2}{c_1^2 + c_2^2 + \cdots + c_n^2},$$

and hence also that

$$\left| \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} \right| \leqq |\lambda_i|_{\max}.$$

(c) Prove that

$$\lambda_n - \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \frac{\sum_{k=1}^{n} (\lambda_n - \lambda_k)c_k^2}{\sum_{k=1}^{n} c_k^2} \geqq 0,$$

for *any* real vector $\mathbf{x}$,

(d) If $\mathbf{x}$ is orthogonal to the characteristic vectors $\mathbf{e}_{r+1}, \mathbf{e}_{r+2}, \cdots, \mathbf{e}_n$, show that

$$\lambda_r - \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \frac{\sum_{k=1}^{r} (\lambda_r - \lambda_k)c_k^2}{\sum_{k=1}^{r} c_k^2} \geqq 0.$$

(e) Show that, if $\mathbf{x} = \mathbf{e}_i$, there follows

$$\frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} \equiv \frac{\mathbf{e}_i^T\mathbf{A}\mathbf{e}_i}{\mathbf{e}_i^T\mathbf{e}_i} = \lambda_i \qquad (i = 1, 2, \cdots, n).$$

117. Let $\mathbf{A}$ be a real symmetric matrix, with characteristic numbers $\lambda_1 \leqq \lambda_2 \leqq \cdots \leqq \lambda_n$ and corresponding normalized and orthogonalized characteristic vectors $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$. Deduce the following results from the results of Problem 116.

(a) The number λ_n is the maximum value of $(\mathbf{x}^T\mathbf{A}\mathbf{x})/(\mathbf{x}^T\mathbf{x})$ for all real vectors $\mathbf{x}$, and this maximum value is taken on when $\mathbf{x}$ is identified with a characteristic vector associated with λ_n.

(b) The number λ_r is the maximum value of $(\mathbf{x}^T\mathbf{A}\mathbf{x})/(\mathbf{x}^T\mathbf{x})$ for all real vectors $\mathbf{x}$ which are simultaneously orthogonal to the characteristic vectors associated with $\lambda_{r+1}, \lambda_{r+2}, \cdots, \lambda_n$, and this maximum value is taken on when $\mathbf{x}$ is identified with a characteristic vector associated with λ_r.

(c) The number λ_n is the maximum value of $\mathbf{e}^T \mathbf{A}\,\mathbf{e}$ for all real *unit* vectors $\mathbf{e}$, the number λ_r is the maximum value for all unit vectors simultaneously orthogonal to $\mathbf{e}_{r+1}, \mathbf{e}_{r+2}, \cdots, \mathbf{e}_n$, and these successive maxima are taken on when $\mathbf{e}$ is identified with the relevant unit characteristic vector.

118. Suppose that Problems 116 and 117 are modified in such a way that $\lambda_1 \leqq \lambda_2 \leqq \cdots \leqq \lambda_n$ are the characteristic numbers of the problem $\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{B}\,\mathbf{x}$, where $\mathbf{A}$ and $\mathbf{B}$ are real and symmetric, and also $\mathbf{B}$ is positive definite, and $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_n$ comprise an orthonormal set of corresponding characteristic vectors, the orthogonality and normality being relative to the matrix $\mathbf{B}$. Show that the results of those Problems again apply if $\mathbf{x}^T\mathbf{x}$ is replaced by $\mathbf{x}^T\mathbf{B}\,\mathbf{x}$ throughout, and if "unit vectors" are of unit length relative to $\mathbf{B}$.

119. Suppose that the characteristic numbers of a real symmetric matrix $\mathbf{A}$ are arranged in order of increasing *absolute value*.

(a) Deduce from the result of Problem 116(b) that the use of equation (250a), in connection with iterative approximation to characteristic quantities, leads to approximations to λ_n which are not greater than λ_n in absolute value.

(b) Show that the use of equation (250b) amounts to approximating λ_n by a ratio of the form

$$\frac{\lambda_1^2 c_1^2 + \lambda_2^2 c_2^2 + \cdots + \lambda_n^2 c_n^2}{\lambda_1 c_1^2 + \lambda_2 c_2^2 + \cdots + \lambda_n c_n^2},$$

and deduce that such an approximation is conservative if all characteristic numbers λ_i are *positive*.

120. (a) If $\mathbf{A}$ is a real symmetric matrix, all of whose elements are *nonnegative* ($a_{ij} \geqq 0$), deduce from preceding results that the characteristic number of largest magnitude is *positive* (although its negative *may* then *also* be a characteristic number), and that all components of the corresponding characteristic vector $\mathbf{e}_n$ are of the same sign, and hence may be taken to be all nonnegative. [Consider the nature of $\mathbf{e}^T\mathbf{A}\,\mathbf{e}$.]

(b) Show that the result of part (a) is also true of the dominant characteristic quantities for the problem $\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{B}\,\mathbf{x}$ if $\mathbf{B}$ is real, symmetric, and positive definite, $\mathbf{A}$ is real and symmetric, and $a_{ij} \geqq 0$.

[The result of part (a) is also true for *any* real square matrix $\mathbf{A}$ with nonnegative elements.]

121. If $\mathbf{A}$ is a real symmetric matrix with characteristic numbers λ_i and corresponding characteristic vectors $\mathbf{e}_i$, show that there follows

$$\mathbf{e}_i^T(\mathbf{A}\,\mathbf{x} - \lambda_i\mathbf{x}) = 0 \qquad (i = 1, 2, \cdots, n),$$

for any real vector $\mathbf{x}$. Hence, with the notation $\mathbf{y} = \mathbf{A}\,\mathbf{x}$ for the "transform" of $\mathbf{x}$, deduce that

$$\mathbf{e}_i^T(\mathbf{y} - \lambda_i\mathbf{x}) = 0 \qquad (i = 1, 2. \cdots, n),$$

for any real vector $\mathbf{x}$.

122. Suppose that $\mathbf{A}$ is a real symmetric matrix with *no negative elements.*

(a) Deduce from the results of Problems 120(a) and 121 that the components of the vector $\mathbf{y} - \lambda_n \mathbf{x}$, where $\mathbf{y} \quad \mathbf{A}\,\mathbf{x}$, then cannot all be of the same sign (unless they all vanish, so that $\mathbf{x}$ is a multiple of $\mathbf{e}_n$).

(b) Deduce that, *in this case*, if the input $\mathbf{x}$ of the iterative method of Section 1.23 posesses only nonegative elements, then the dominant characteristic number λ_n is not larger than the largest ratio y_i/x_i of corresponding elements of the output and input vectors, and not less than the smallest such ratio:

$$\min_i \frac{y_i}{x_i} \leqq \lambda_n \leqq \max_i \frac{y_i}{x_i}.$$

[This result also holds for *any* real square matrix *with no negative elements*.]

123. Prove that the statement of Problem 122(b) is true also for the application of the iterative method to the determination of the dominant characteristic number of the problem $\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{B}\,\mathbf{x}$ if $\mathbf{B}$ is real, symmetric, and positive definite whereas $\mathbf{A}$ is real and symmetric and composed only of nonnegative elements, and if also $\mathbf{A}$ and $\mathbf{B}$ are *commutative*. [Use Problem 120(b) and a generalization of Problem 121, with $\mathbf{y} = \mathbf{B}^{-1}\,\mathbf{A}\,\mathbf{x} = \mathbf{A}\,\mathbf{B}^{-1}\,\mathbf{x}$.]

124. Suppose that $\mathbf{A}$ is a real, symmetric, positive definite matrix such that, whereas all diagonal elements are positive, all elements off the diagonal are either negative or zero. [See, for example, the matrix of coefficients in (295).]

(a) Show that, if α is any positive constant *larger than the largest diagonal element of* $\mathbf{A}$, then the matrix $\mathbf{M} \quad \alpha\,\mathbf{I} - \mathbf{A}$ is a symmetric matrix, all of whose elements are nonnegative.

(b) Show that the characteristic numbers μ_i of $\mathbf{M}$ are given by $\mu_i = \alpha - \lambda_i$, where λ_i are the characteristic numbers of $\mathbf{A}$, and that the characteristic vector of $\mathbf{M}$ associated with μ_i is that of $\mathbf{A}$ associated with λ_i.

(c) Use the result of Problem 120(a) to show that the largest μ_i is positive. Deduce that the dominant μ_i is $\mu_1 = \alpha - \lambda_1$, where λ_1 is the smallest characteristic number of $\mathbf{A}$, and that all components of the corresponding characteristic vector have the same sign. Hence show that the *smallest* characteristic number of a matrix $\mathbf{A}$ of the type under consideration is not larger than the largest diagonal element of $\mathbf{A}$, and that all components of the associated characteristic vector may be taken as nonnegative. [Notice that the matrix $\mathbf{M}$ can be obtained more easily than the matrix $\mathbf{A}^{-1}$, for the purpose of determining λ_1 and the corresponding characteristic vector by matrix iteration.]

125. Generalize the results and procedures of Problem 124 to the case of the problem $\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{B}\,\mathbf{x}$ where $\mathbf{A}$ is a matrix of the type described in that problem, and $\mathbf{B}$ is a positive definite matrix with no negative elements. [Here $\mathbf{M} = \alpha\,\mathbf{B} - \mathbf{A}$, where $\alpha > (b_{ii}/a_{ii})_{\max}$.]

Section 1.28.

126. Prove that the relation

$$\|f(x) + g(x)\|^2 = \|f(x)\|^2 + \|g(x)\|^2$$

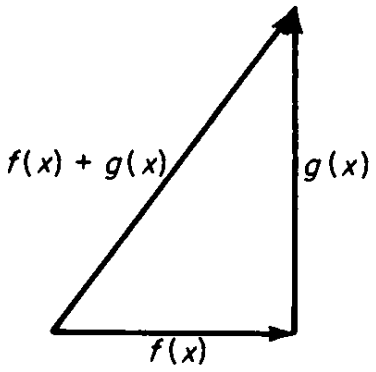

FIGURE 1.2

is true, over a prescribed interval (a, b), if and only if f and g are orthogonal over (a, b). Notice that if we think of $\|f(x)\|^2 = \int_a^b f^2\,dx$ as the "square of the length of $f(x)$" in the function space relevant to (a, b), then this result is analogous to the Pythagorean theorem (see Figure 1.2).

127. By noticing that, if all integrals are evaluated over an interval (a, b), the quantity

$$\int [f(x) + \lambda g(x)]^2\,dx \equiv \int f^2\,dx + 2\lambda \int fg\,dx + \lambda^2 \int g^2\,dx$$

is necessarily nonnegative for any real value of λ, deduce that

$$\left(\int fg\,dx\right)^2 \leqq \left(\int f^2\,dx\right)\left(\int g^2\,dx\right)$$

and hence

$$\left|\int fg\,dx\right| \leqq \left(\sqrt{\int f^2\,dx}\right)\left(\sqrt{\int g^2\,dx}\right).$$

This relation is known as the *Schwarz inequality*. Show also that equality holds if and only if $g(x)$ differs from a constant multiple of $f(x)$ by a trivial function.

Deduce that if we define the "angle between $f(x)$ and $g(x)$" in the function space relevant to (a, b) by the equation

$$\cos\theta[f,g] = \frac{\int_a^b fg\,dx}{\sqrt{\int_a^b f^2\,dx}\sqrt{\int_a^b g^2\,dx}} = \frac{(f,g)}{\|f\| \cdot \|g\|}$$

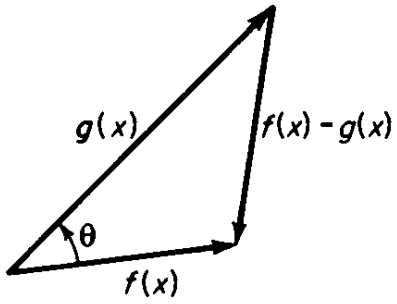

FIGURE 1.3

then θ is a *real* angle. Notice that this definition is completely analogous to the geometrical definition of the angle between two vectors (see Figure 1.3).

128. With the terminology of Problem 127, establish the "law of cosines,"

$$\|f - g\|^2 = \|f\|^2 + \|g\|^2 - 2\,\|f\| \cdot \|g\| \cos\theta[f,g],$$

in function space.

129. Verify the truth of the identity

$$\int (f-g)^2\,dx \quad \begin{cases} \left(\sqrt{\int f^2\,dx} - \sqrt{\int g^2\,dx}\right)^2 + 2\left[\sqrt{\int f^2\,dx}\sqrt{\int g^2\,dx} - \int fg\,dx\right] \\ \left(\sqrt{\int f^2\,dx} + \sqrt{\int g^2\,dx}\right)^2 - 2\left[\sqrt{\int f^2\,dx}\sqrt{\int g^2\,dx} + \int fg\,dx\right] \end{cases}$$

where each integral is evaluated over the interval (a, b). Use the Schwarz inequality (Problem 127) to show that each quantity in square brackets is nonnegative, and hence deduce the relation

$$\|f\| - \|g\| \leq \|f - g\| \leq \|f\| + \|g\|,$$

where neither equality can hold unless $g(x)$ differs from a constant multiple of $f(x)$ by a trivial function.

To what geometrical relation is this function-theoretical result analogous?

130. As a special case of the Schwarz inequality (Problem 127), deduce that

$$\frac{1}{b-a}\int_a^b f\,dx \leqq \sqrt{\frac{1}{b-a}\int_a^b f^2\,dx}.$$

[The left-hand member is the mean value of $f(x)$ over (a, b), the right-hand member the so-called *root mean square* (rms) value.]

131. Establish the validity of the following statement: "The rms value of the sum of two functions over a given interval is not greater than the sum of the separate rms values, and not less than their difference."

132. Let $f_1(x), f_2(x). \cdots, f_n(x), \cdots$ comprise an infinite set of functions defined over (a, b). Describe a procedure for forming from this set a set of linear combinations $\phi_1(x), \phi_2(x), \cdots, \phi_n(x), \cdots$ which is orthonormal over (a, b). [See Section 1.13.]

133. Show that the functions $f_0(x) = x$ and $f_k(x) = \sin \mu_k x$ $(k = 1, 2, \cdots)$ comprise an orthogonal set over $(0, 1)$ if the constants μ_k are the positive roots of the transcendental equation $\tan \mu_k = \mu_k$.

134. Show that the complex functions $f_k(x) = e^{ikx}$, where k takes on *all* integral values, comprise a set which is orthogonal in the Hermitian sense over any real interval $(a, a + 2\pi)$. Determine the normalizing factors.

Section 1.29.

135. Show that the functions defined in Problem 133 are the characteristic functions of the following Sturm-Liouville problem:

$$\frac{d^2y}{dx^2} + \mu^2 y = 0; \qquad y(0) = 0, \quad y(1) = y'(1).$$

136. Determine the coefficients in the expansion

$$1 = A_0 x + \sum_{k=1}^{\infty} A_k \sin \mu_k x \quad (0 < x < 1),$$

where $\tan \mu_k = \mu_k$.

137. (a) If a function $F(x)$ possesses the expansion

$$F(x) = A_0 x + \sum_{k=1}^{\infty} A_k \sin \mu_k x \qquad (0 < x < 1),$$

where $\tan \mu_k = \mu_k$, obtain the solution of the problem

$$\frac{d^2y}{dx^2} + \lambda y = F(x); \qquad y(0) = 0, \quad y(1) = y'(1),$$

in the form

$$y(x) = \frac{A_0}{\lambda} x + \sum_{k=1}^{\infty} \frac{A_k}{\lambda - \mu_k{}^2} \sin \mu_k x \qquad (0 < x < 1),$$

when $\lambda \neq 0, \mu_1{}^2, \mu_2{}^2, \cdots$.

(b) Use the result of Problem 136 to obtain this solution in the special case when $F(x) = 1$.

138. Suppose that y is a nontrivial solution of the equation

$$\mathscr{L}y + \lambda r y = 0 \qquad (a < x < b)$$

where $\mathscr{L}$ is defined by equation (321), with p, q, and r real.

(a) By multiplying the equal members of this equation by the complex conjugate function $\bar{y}$, integrating the resultant relation over (a, b), and transforming the result by integration by parts, show that

$$\lambda \int_a^b r y \bar{y}\, dx = \int_a^b p y' \bar{y}'\, dx - \int_a^b q y \bar{y}\, dx - [p \bar{y} y']_a^b.$$

(b) If $p(x)$ and $r(x)$ are *positive* and $q(x)$ *nonpositive* over (a, b) and if y satisfies the end conditions

$$\alpha_2 y(a) + \beta_1 y'(a) = 0, \qquad \alpha_1 y(b) + \beta_2 y'(b) = 0,$$

show that λ is real and positive when also α_1 and β_1 are of the *same* sign (or $\alpha_1\beta_1 = 0$) and α_2 and β_2 are of *opposite* sign (or $\alpha_2\beta_2 = 0$). [Thus, under these conditions, the characteristic values of λ are real and positive and the corresponding characteristic functions can be taken to be real (by rejecting a permissible imaginary multiplicative constant).]