FINITE MATRICES

BY

W. L. FERRAR, M.A., D.Sc.
FELLOW OF HERTFORD COLLEGE
OXFORD

OXFORD
AT THE CLARENDON PRESS
1951

Oxford University Press, Amen House, London E.C.4
GLASGOW NEW YORK TORONTO MELBOURNE WELLINGTON
BOMBAY CALCUTTA MADRAS CAPE TOWN
Geoffrey Cumberlege, Publisher to the University

PRINTED IN GREAT BRITAIN

PREFACE

THIS book is written for graduate students and for undergraduates whose degree courses include more matrix theory than a text-book of elementary properties will provide. In it I have tried to give an account of the theory of finite matrices, including their invariant factors and elementary divisors, which can be read with reasonable ease by mathematicians who are not specialists in this particular field. I have worked with the ordinary numbers of analysis and have not considered, save for an odd reference or two, the demands of an abstract algebra. My aim throughout has been to make the argument simple and straightforward.

When I began the book I expected that the whole of it would be concerned with the presentation of results long since known in some form or other. On reaching the chapter on functions of matrices I found that, starting from a few 'well-known' facts, the theory unfolded itself naturally and easily, but that only patches of it here and there appeared to have been published before. Accordingly, Chapter V is largely a first essay at a connected account of this part of the theory.

My indebtedness to other books and to research papers is very great. The reader who wishes to acquire a knowledge of the wider field within which my own limited treatment lies should consult, among others:

H. W. Turnbull and A. C. Aitken, *An Introduction to the Theory of Canonical Matrices* (Blackie, 1932);

W. V. D. Hodge and D. Pedoe, *Methods of Algebraic Geometry* (Cambridge, 1947);

G. Julia, *Introduction mathématique aux théories quantiques* (Gauthier-Villars, 1949), Part I on Finite Matrices and Part II on Hilbert Space and Infinite Matrices;

A. A. Albert, *Modern Higher Algebra* (Chicago, 1937);

J. H. M. Wedderburn, *Lectures on Matrices* (Amer. Math. Soc. Colloquium Publications, vol. xvii, 1934);

C. C. MacDuffee, *The Theory of Matrices* (Chelsea Publishing Co., New York, 1946: reprint of first edition).

For anything concerning matrices that was known prior to 1932 MacDuffee's book is invaluable. A similar account of what has been done since 1932 would be a great asset; is it too much to hope that a scholar might one day write it or edit a series of B.Sc. and Ph.D. theses written to that end? The present book makes no pretence to be complete, even in the central topics of finite matrices: it attempts a clear and readable account of the principal theorems and no more.

I end with an acknowledgement of my debt to the staff of the Clarendon Press. I have no immediate plans for another book with which to tax their skill and forbearance and so, on this occasion, I wish particularly to thank all of them for the way in which, over a period of some fifteen years, a series of not too tidy manuscripts has been transformed into well-printed books.

HERTFORD COLLEGE, OXFORD W. L. F.
April 1951

CONTENTS

I.	INTRODUCTION	1
II.	EQUIVALENT MATRICES	15
III.	EQUIVALENT λ-MATRICES	26
IV.	COLLINEATION	48
V.	INFINITE SERIES AND FUNCTIONS OF MATRICES	87
VI.	CONGRUENCE	135
VII.	MATRIX EQUATIONS	153
VIII.	MISCELLANEOUS NOTES	168
	INDEX	181

CHAPTER I

INTRODUCTION

1. Scope of the chapter

THE aim of this introductory chapter is to provide a résumé of the more elementary properties of matrices. I have thought it useful, for both writer and reader, to note explicitly, even though it be briefly and sketchily, the accepted facts about matrices on which the later chapters will be based. Some proofs, but not all, will be given. I have tried to hold the balance between brevity and clarity and, accordingly, I have omitted many details that would find a place in a full account of what is given here in outline.

2. Notation

(*a*) A set of mn numbers arranged in m columns and n rows is called a matrix.† Thus

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ . & . & . & . \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}$$

is a matrix. The square bracket is a conventional symbol which is read as 'the matrix'. The individual numbers are referred to as the ELEMENTS OF THE MATRIX.

We shall normally use the square bracket whenever we wish to indicate that an array of numbers is to be considered as a matrix: thus

$$[x_1, x_2,..., x_n]$$

indicates that the n letters x_1, x_2,..., x_n are to be considered as a ONE-ROW MATRIX. To indicate that n letters x_1, x_2,..., x_n are to be considered as a ONE-COLUMN MATRIX we use the special notation

$$\{x_1, x_2,..., x_n\}.$$

When a matrix has n rows and n columns we refer to it as a SQUARE MATRIX OF ORDER n.

† Some writers insist that the laws of addition and multiplication must be laid down before the use of the word matrix can be justified.

(*b*) Capital italic letters A, B,..., X will be used to denote matrices, in general square matrices of order n. To indicate the actual elements of a matrix we shall write down the element in the ith row and kth column; thus

$$A = [a_{ik}], \qquad B = [\xi_{ki}]$$

means that A has the element a_{ik} in the ith row and kth column, while B has the element ξ_{ki} in the ith row and kth column.

The DETERMINANT whose elements are precisely those of a square matrix A is denoted by $|A|$. When $|A| = 0$ the matrix is said to be SINGULAR and when $|A| \neq 0$ the matrix is said to be NON-SINGULAR.

We sometimes use a special notation for matrices having a single row or a single column. Clarendon letters **a**, **b**,..., **x** denote single-column matrices and (anticipating the later definition of a transpose) **a**′, **b**′,..., **x**′ denote single-row matrices; thus

$$\mathbf{x} = \{x_1,..., x_n\}, \qquad \mathbf{x}' = [x_1,..., x_n]$$

means that **x** is the single-column matrix and **x**′ the single-row matrix having the elements shown.

(*c*) It being understood that, unless the contrary is stated, all literal suffixes run from 1 to n, we shall use the SUMMATION CONVENTION for repeated literal suffixes. With this convention

$$a_{rs}x_s \quad \text{denotes} \quad \sum_{s=1}^{n} a_{rs}x_s,$$

and

$$a_{rs}x_{rs} \quad \text{denotes} \quad \sum_{r=1}^{n}\sum_{s=1}^{n} a_{rs}x_{rs}.$$

On the other hand, a repeated numerical suffix, such as the suffix 1 in $a_{r1}x_1$, will not imply a summation.

When the occasion arises we shall enclose a literal suffix in brackets to denote that there is to be no summation with respect to that particular suffix; thus

$$a_{rs}x_{r(s)} \quad \text{denotes} \quad \sum_{r=1}^{n} a_{rs}x_{rs},$$

there being no summation with respect to s.

(*d*) The square matrix of order n having unity in each place on its leading diagonal and zero elsewhere is called the UNIT

MATRIX of order n; it is denoted by I. Sometimes we use I_r, I_s,... to denote unit matrix of order $r, s, \dots$.

A matrix having zero in every place is called a NULL MATRIX and is denoted by 0.

A square matrix of order n whose only non-zero elements occur in its leading diagonal is called a DIAGONAL MATRIX.

3. Addition and multiplication

(a) *Addition.* $$[a_{ik}]+[b_{ik}] = [a_{ik}+b_{ik}]. \quad (1)$$

The definition applies to any two matrices A and B, not necessarily square, provided that each has the same number of rows and each has the same number of columns. Moreover, from the definition, $$A+B = B+A.$$

(b) *Multiplication.*

$$[a_{ik}]\times[b_{ik}] = [a_{i\lambda}b_{\lambda k}]. \quad (2)$$

The definition applies to any two matrices A and B, not necessarily square, provided that the number of columns in A is equal to the number of rows in B. The product AB has as many rows as A and as many columns as B.

From the definition,

$$A(B+C) = AB+AC$$

and $$(B+C)A = BA+CA.$$

Further, $$[a_{ik}]\times[b_{ik}]\times[c_{ik}],$$

whether the triple product be formed by

$$(AB)C \quad \text{or} \quad A(BC),$$

is equal to $$[a_{i\lambda}b_{\lambda\mu}c_{\mu k}] \quad (3)$$

and is commonly denoted by ABC. Products of four or more matrices are formed on the same pattern: thus

$$ABC\dots Z = [a_{i\lambda}b_{\lambda\mu}c_{\mu\nu}\dots z_{\rho k}],$$

the i, k being the only suffixes that do not imply summation.

By their definitions, AB and BA are matrices whose elements are formed by different processes and, in general, AB is not equal to BA. On the other hand

$$AI = IA = A$$

for every square matrix A of order n.

Finally, the equation $AB = 0$ may be true when neither A nor B is a null matrix; but if either A or B is known to be non-singular when $AB = 0$, then the other is necessarily a null matrix.

4. Related matrices

(a) *The reciprocal of a matrix.*

When $A = [a_{ik}]$, the determinant obtained from $|A|$ by deleting the rth row and sth column and multiplying by the sign-factor $(-1)^{r+s}$ is denoted by A_{rs}; it is called the cofactor of a_{rs} in $|A|$. The matrix†

$$[A_{ki}]$$

is called the ADJUGATE or ADJOINT of A.

It is a well-known result in the theory of determinants that

$$a_{ij}A_{kj} = 0 \quad (i \neq k) \tag{4}$$

and

$$a_{ij}A_{(i)j} = |A| \quad (i = 1,2,\ldots,n). \tag{5}$$

Hence, when A is a non-singular matrix,

$$[a_{ik}]\times[A_{ki}/|A|] = [a_{ij}A_{kj}/|A|] = I$$

and similarly, on working with columns instead of rows at lines (4) and (5),

$$[A_{ki}/|A|]\times[a_{ik}] = I.$$

Accordingly, we call the matrix

$$[A_{ki}/|A|]$$

the RECIPROCAL of A and denote it by A^{-1}.

When A is a singular matrix, $|A| = 0$ and the division by $|A|$ is no longer valid; the reciprocal is not then definable.

Moreover A^{-1} is the only matrix with the property that its product by A is equal to I. For, if $RA = I$, then

$$(R-A^{-1})A = RA-I = 0;$$

on multiplying by A^{-1}, we get

$$(R-A^{-1})AA^{-1} = 0,$$

or, since $AA^{-1} = I$, $\quad R-A^{-1} = 0.$

Thus a matrix R for which $RA = I$ must be equal to A^{-1}. Similarly, $AR = I$ implies $R = A^{-1}$.

† Notice that A_{ki} comes in the ith row and kth column.

(b) *The powers of a matrix.*

The notations A^2, A^3,..., stand for AA, AA^2,...; A^{-2}, A^{-3},... stand for $A^{-1}A^{-1}$, $A^{-1}A^{-2}$,...; and with this notation

$$A^r \times A^s = A^s \times A^r = A^{r+s}$$

for all integer values of r and s (positive, negative, or zero) provided that A^0 is interpreted to be I.

(c) *The transpose of a matrix.*

The matrix whose rth column ($r = 1, 2,...$) is the rth row of A is called the TRANSPOSE of A and is commonly denoted by A'. Thus $A = [a_{ik}]$ gives $A' = [a_{ki}]$.

When $A = A'$, that is when $a_{ik} = a_{ki}$, the matrix A is said to be SYMMETRICAL.

(d) *Functions of a matrix.*

When $a_0, a_1,..., a_p$ are numbers and

$$f(x) \equiv a_0 + a_1 x + ... + a_p x^p$$

is a polynomial in a single variable x, the matrix-sum†

$$a_0 I + a_1 A + ... + a_p A^p$$

is a single matrix that is conveniently denoted by $f(A)$. If this matrix is non-singular, it has a reciprocal and this is conveniently denoted by $1/f(A)$.

The product of this by $g(A)$, where $g(x)$ is another polynomial in x, yields a matrix that is conveniently denoted by $g(A)/f(A)$. The resulting matrix is said to be a rational function of A.

5. The law of reversal for transposes and reciprocals

Let $A = [a_{ik}]$, $B = [b_{ik}]$.

Then $AB = [a_{ij} b_{jk}]$, $(AB)' = [a_{kj} b_{ji}]$.

But

$$B'A' = [b_{ki}] \times [a_{ki}]$$
$$= [b_{ji} a_{kj}]$$
$$= [a_{kj} b_{ji}] = (AB)'.$$

† The matrix $a_1 A$, where a_1 is an ordinary number, is defined to be the matrix whose elements are a_1 times the elements of A; e.g.

$$2\begin{bmatrix} 1 & 4 \\ 3 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 8 \\ 6 & 10 \end{bmatrix}.$$

Hence, *the transpose of a product is the product of the transposes taken in the reverse order.*

By extension, $(ABC)' = C'B'A'$

and $(AB...K)' = K'...B'A'.$

Again

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

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

Accordingly, *the reciprocal of a product is the product of the reciprocals taken in the reverse order.*

By extension, $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$

and $(AB\ldots K)^{-1} = K^{-1}\ldots B^{-1}A^{-1}.$

6. Simple matrix equations

(*a*) When A, B are given square matrices of order n and B is non-singular, the matrix equations

$$A = BX, \qquad A = YB$$

in the 'unknowns' X and Y have the unique solutions

$$X = B^{-1}A, \qquad Y = AB^{-1}.$$

That these are solutions follows at once from the fact that

$$BB^{-1} = B^{-1}B = I.$$

Moreover, each solution is unique: if, for example, BR is also equal to A, then $B(R-X) = 0$ and, since B is non-singular,

$$R-X = 0.$$

(*b*) The unique solution of the matrix equation

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

where A is a non-singular square matrix of order n and $\mathbf{x}$, $\mathbf{b}$ are single-column matrices of n rows, is $\mathbf{x} = A^{-1}\mathbf{b}$. It provides the solution of the n linear equations

$$a_{ik}x_k = b_i \quad (i = 1, 2,\ldots, n).$$

Similarly, the matrix equation

$$\mathbf{y}'A = \mathbf{b}'$$

has the solution $\mathbf{y}' = \mathbf{b}'A^{-1}$ and provides the solution of the n equations
$$y_k a_{ki} = b_i \quad (i = 1, 2, ..., n).$$

7. Submatrices

A matrix
$$P \equiv \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$$

can be denoted, on introducing symbols P_1, P_2, P_3, P_4, where
$$P_1 \equiv \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}, \qquad P_2 \equiv \begin{bmatrix} p_{13} \\ p_{23} \end{bmatrix},$$
$$P_3 \equiv [p_{31} \quad p_{32}], \qquad P_4 \equiv [p_{33}],$$

by
$$\begin{bmatrix} P_1 & P_2 \\ P_3 & P_4 \end{bmatrix}.$$

The matrices P_1, P_2, P_3, P_4 are called submatrices of P.

When a second matrix Q is divided into submatrices on the same pattern as P, a little calculation shows that
$$P+Q = \begin{bmatrix} P_1+Q_1 & P_2+Q_2 \\ P_3+Q_3 & P_4+Q_4 \end{bmatrix}$$

and
$$PQ = \begin{bmatrix} P_1 Q_1+P_2 Q_3 & P_1 Q_2+P_2 Q_4 \\ P_3 Q_1+P_4 Q_3 & P_3 Q_2+P_4 Q_4 \end{bmatrix}.$$

In this sum and product the 'elements' are themselves matrices; for example, P_1+Q_1 is a matrix of two rows and two columns, while $P_3 Q_1+P_4 Q_3$ is a matrix of one row and two columns.

The symmetry of the previous example is not an essential feature of the process. In multiplication, for example, what is essential is that, in each $P_r Q_s$ that occurs, P_r shall have as many columns as Q_s has rows. As an illustration of a non-symmetrical arrangement (it has no other interest and is devised purely as an illustration), let
$$P_{11} = [p_{11} \quad p_{12}], \qquad P_{12} = [p_{13}],$$
$$Q_{11} = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \end{bmatrix}, \qquad Q_{12} = \begin{bmatrix} q_{14} \\ q_{24} \end{bmatrix},$$
$$Q_{21} = [q_{31} \quad q_{32} \quad q_{33}], \qquad Q_{22} = [q_{34}].$$

Then

$$[p_{11} \quad p_{12} \mid p_{13}] \times \begin{bmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ \hline q_{31} & q_{32} & q_{33} & q_{34} \end{bmatrix}$$

$$= [P_{11} \quad P_{12}] \times \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}$$

$$= [P_{11}Q_{11}+P_{12}Q_{21} \quad P_{11}Q_{12}+P_{12}Q_{22}]$$

$$= [p_{1k}q_{k1} \quad p_{1k}q_{k2} \quad p_{1k}q_{k3} \quad p_{1k}q_{k4}],$$

the summation with regard to k being for $k = 1, 2, 3$.

8. The rank of a matrix

(a) *Minors.*

Let A be a matrix, not necessarily square. From it delete all rows save a certain r rows and all columns save a certain r columns. When $r > 1$ the elements that remain form a square matrix of order r and the determinant of this matrix is called a minor of A of order r. A single element of A may be considered to be a minor of order 1.

(b) *Definition of rank.*

A matrix has rank r ($\geqslant 1$) when r is the largest integer for which we can state that 'not ALL *minors of order r are zero'.*

To understand the definition we note that a minor of order $k+1$ can be expanded by its first row as a sum of multiples of minors of order k, so that if all minors of order k are zero, then all minors of order $k+1$ are zero. The converse is not true; for example, in

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

the only minor of order 4 is the determinant of the matrix and its value is zero, but the minor of order 3

$$\begin{vmatrix} 1 & 2 & 4 \\ 0 & 5 & 1 \\ 1 & 0 & 2 \end{vmatrix}$$

is not equal to zero.

It is sometimes convenient to speak of a null matrix, in which every element is zero, as being of rank zero.

(c) *Linear dependence.*

Consider an array of three rows

$$\begin{matrix} a_1, & b_1, & \ldots, & z_1, \\ a_2, & b_2, & \ldots, & z_2, \\ a_3, & b_3, & \ldots, & z_3. \end{matrix}$$

If the three rows are related in such a way that there are numbers λ_1 and λ_2 for which

$$\rho_3 = \lambda_1 \rho_1 + \lambda_2 \rho_2 \quad (\rho = a, b, \ldots, z) \tag{6}$$

we say that the third row is the SUM OF MULTIPLES (λ_1 and λ_2) of the first and second rows. When the three rows, or two of them, are related in such a way that there are numbers λ_1, λ_2, λ_3, of which two at least are not zero, and for which

$$\lambda_1 \rho_1 + \lambda_2 \rho_2 + \lambda_3 \rho_3 = 0 \quad (\rho = a, b, \ldots, z), \tag{7}$$

we say that the three rows are LINEARLY DEPENDENT. We say that the rows are LINEARLY INDEPENDENT if (7) is satisfied only when $\lambda_1 = \lambda_2 = \lambda_3 = 0$.

The definitions extend to any number of rows or of columns.

(d) *Rank and linear dependence.*

The rank of a matrix is equal to the number of linearly independent rows in the matrix, as the following theorem shows.

Let A be a matrix of rank r and let a non-zero minor of A of order r have elements from the αth, βth,..., κth rows of A (r rows in all). Let A have a further row, say the θth. Then† *there are numbers λ_α, λ_β,..., λ_κ for which*

$$\rho_\theta = \lambda_\alpha \rho_\alpha + \lambda_\beta \rho_\beta + \ldots + \lambda_\kappa \rho_\kappa,$$

so that the θth row is the sum of multiples of the αth, βth,..., κth rows.

Thus we can, when the number of rows of A exceeds its rank r, select r rows of A and express every other row as a sum of multiples of the r selected rows. Moreover, *it is not possible to*

† This and other properties noted in this section are proved in Ferrar, *Algebra* (Oxford, 1941), at chapter viii. Further references to this book will be indicated by F. and the appropriate page number.

select q rows of A, where $q < r$, and then express every other row as a sum of multiples of the q selected rows.

There are similar results for columns.

9. Linear equations

Consider the m linear equations

$$\sum_{k=1}^{n} a_{ik} x_k = b_i \quad (i = 1,..., m) \tag{8}$$

in the n unknowns $x_1,..., x_n$. Let the matrices A, B be given by

$$A = \begin{bmatrix} a_{11} & . & . & a_{1n} \\ . & . & . & . \\ a_{m1} & . & . & a_{mn} \end{bmatrix}, \qquad B = \begin{bmatrix} a_{11} & . & . & a_{1n} & b_1 \\ . & . & . & . & . \\ a_{m1} & . & . & a_{mn} & b_m \end{bmatrix};$$

let A be of rank r and B of rank r'. Then, from the nature of definition of rank, $r \leqslant r'$; moreover

When $r = r'$ the equations (8) *are consistent (that is, there is at least one set of values of the unknowns that satisfies all the equations) and when $r < r'$ the equations* (8) *are not consistent.*

When all the b_i in (8) are zero, we are concerned with what are known as HOMOGENEOUS LINEAR EQUATIONS, namely

$$\sum_{k=1}^{n} a_{ik} x_k = 0 \quad (i = 1,..., m). \tag{9}$$

The ranks r and r' of the above discussion are now necessarily equal and the equations are always consistent. On the other hand, the equations are always satisfied by

$$x_1 = x_2 = ... = x_n = 0$$

and this may be the only solution. The following theorem† summarizes the more important results about such a set of equations.

Let A, in (9), be of rank r. Then $r \leqslant n$, since A has n columns.

(i) *When $r = n$, the equations* (9) *have no solution other than*

$$x_1 = x_2 = ... = x_n = 0.$$

† For proofs and further details, see F. 98–105.

(ii) *When* $r = n-1$, *the equations have effectively only one non-zero solution*† *and if this is*

$$\xi_1, \quad \xi_2, \quad ..., \quad \xi_n,$$

all other non-zero solutions are of the form

$$\lambda\xi_1, \quad \lambda\xi_2, \quad ..., \quad \lambda\xi_n.$$

(iii) *When* $r < n-1$, *the equations* (9) *have* $n-r$ *linearly independent non-zero solutions and every non-zero solution can be expressed as a sum of multiples of these* $n-r$ *solutions.*

10. The rank of a product of two matrices

We note two important theorems of frequent application.

(*a*) *The rank of a product AB cannot exceed the rank of either factor.*

(*b*) *When B is a non-singular square matrix of the same order as the square matrix A, the matrices*

$$A, \quad AB, \quad BA$$

all have the same rank.

The proof of these theorems follows fairly directly from the following result,‡ one that is often used apart from its immediate connexion with the ranks of matrices.

(*c*) Let A have n_1 rows and n columns and let B have n rows and n_2 columns; then AB has n_1 rows and n_2 columns. *Every minor of AB of order greater than n, if there are such minors, is equal to zero; and every minor of AB of order* $t \leqslant n$ *is* either *the product of a t-rowed minor of A by a t-rowed minor of B* or *is the sum of a number of such products.*

This contains as a special case the more elementary result

When A and B are square matrices of the same order, the determinant $|AB| = |A| \times |B|$. The proof of this follows at once from (2) of § 3. Its extension to a product of three or more matrices,

$$|ABC...K| = |A|.|B|.|C|.....|K|,$$

is also a direct consequence of the definition of a product of matrices.

† 'Non-zero' because at least one ξ is different from zero.

‡ F. 109.

11. The characteristic equation of a matrix

When A is the square matrix $[a_{ik}]$, the determinant of the matrix $A-\lambda I$ is given by

$$|A-\lambda I| = \begin{vmatrix} a_{11}-\lambda & a_{12} & . & . & a_{1n} \\ a_{21} & a_{22}-\lambda & . & . & a_{2n} \\ . & . & . & . & . \\ a_{n1} & a_{n2} & . & . & a_{nn}-\lambda \end{vmatrix}.$$

When expanded, this determinant is of the form

$$f(\lambda) \equiv (-1)^n(\lambda^n+p_1\lambda^{n-1}+...+p_n), \tag{10}$$

where the p_r are polynomials in the n^2 elements a_{ik}. The roots of the equation

$$|A-\lambda I| = 0 \tag{11}$$

are called the LATENT ROOTS of the matrix A and the equation itself is called the CHARACTERISTIC EQUATION of A. Moreover†

(*a*) *Every square matrix satisfies its own characteristic equation; that is, if*

$$|A-\lambda I| = (-1)^n(\lambda^n+p_1\lambda^{n-1}+...+p_n),$$

then $$A^n+p_1A^{n-1}+...+p_{n-1}A+p_nI = 0.$$

(*b*) *The characteristic roots of a symmetrical matrix having real elements are all real numbers.*

(*c*) *The characteristic roots of a Hermitian matrix* (cf. § 13) *are all real numbers.*

12. Special notations

(*a*) *The diagonal matrix.*

The notation $$\text{diag}\{\alpha_1, \alpha_2,..., \alpha_r\}$$

indicates a matrix, of r rows and columns, which has the elements shown in its leading diagonal and zeros in all other positions. The extension of the notation to cover submatrices is a common one; thus, when

$$A_1 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \qquad A_2 = \begin{bmatrix} a_{33} & a_{34} \\ a_{43} & a_{44} \end{bmatrix},$$

the matrix $$\text{diag}\{A_1, A_2\}$$

† F. 111 for (*a*); F. 146 for (*b*); F. 158, Ex. 14, for (*c*).

is, in full,
$$\begin{bmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ 0 & 0 & a_{33} & a_{34} \\ 0 & 0 & a_{43} & a_{44} \end{bmatrix}.$$

(b) *Functions of diagonal matrices.*

Let $A_1, A_2,..., A_k$ be square matrices of orders $r_1, r_2,..., r_k$ and let $B_1, B_2,..., B_k$ be a second set of square matrices of orders $r_1, r_2,..., r_k$. Let

$$A = \text{diag}\{A_1, A_2,..., A_k\}, \qquad B = \text{diag}\{B_1, B_2,..., B_k\}.$$

Then $$AB = \text{diag}\{A_1 B_1, A_2 B_2,..., A_k B_k\}, \tag{12}$$

as can be seen by considering the matrix product set out in the form

$$\begin{bmatrix} A_1 & 0 & . & . & 0 \\ 0 & A_2 & . & . & 0 \\ . & . & . & . & . \\ 0 & 0 & . & . & A_k \end{bmatrix} \times \begin{bmatrix} B_1 & 0 & . & . & 0 \\ 0 & B_2 & . & . & 0 \\ . & . & . & . & . \\ 0 & 0 & . & . & B_k \end{bmatrix}.$$

The non-zero elements of the inner products of the first r_1 rows of A (when A is set out in full as a matrix of $r_1+r_2+...+r_k$ rows) by the first r_1 columns of B (when B is similarly set out in full) are precisely the inner products of the rows of A_1 by the columns of B_1. They yield the submatrix $A_1 B_1$; and so for the other elements of (12).

In particular, when $B_i = A_i$ ($i = 1,..., k$), the result (12) gives

$$A^2 = \text{diag}\{A_1^2, A_2^2,..., A_k^2\}$$

and we may build from this the result

$$f(A) = \text{diag}\{f(A_1), f(A_2),..., f(A_k)\}, \tag{13}$$

where $f(A)$ is any polynomial in A or is a rational function $g(A)/h(A)$, provided $h(A)$ is not a singular matrix.

(c) *Special types of matrix.*

We use A' to denote the transpose of A, and $\bar{A}$ to denote the matrix whose elements are the conjugate complexes† of the

† When $z = x+iy$, $\bar{z} = x-iy$.

elements of A. The following list defines the terms there introduced:

$A' = A$	A is *symmetric* $(a_{ij} = a_{ji})$,
$\bar{A}' = A$	A is *Hermitian* $(a_{ij} = \bar{a}_{ji})$,
$A'A = I$	A is *orthogonal*,
$\bar{A}'A = I$	A is *unitary*,
$A' = -A$	A is *skew-symmetric*,
$\bar{A}' = -A$	A is *skew-hermitian*.

CHAPTER II

EQUIVALENT MATRICES

Much of this chapter is taken up by details about the 'elementary transformations' of a matrix. On getting to the end of the chapter, the reader may well feel a sense of anti-climax, in that these details have led to a result (compare Theorem 2, p. 22) which is so general that it is of little use as it stands. The detail is, however, a basis for the work of later chapters and this, rather than the result immediately obtained, justifies its inclusion.

1. Preliminary

1.1. *Field of numbers*

DEFINITION 1. *A set of numbers, real or complex, is said to form a* FIELD OF NUMBERS *if it satisfies the conditions*

(i) *whenever r, s belong to the set,*

$$r+s, \qquad r-s, \qquad r\times s$$

also belong to the set;

(ii) *whenever r, s belong to the set and s is not zero,*

$$r\div s$$

also belongs to the set.

1.2. *Elementary transformations of matrices*

DEFINITION 2. *The* ELEMENTARY TRANSFORMATIONS *of a matrix are*

(i) *the interchange of two rows or of two columns;*

(ii) *the multiplication of the elements of a row (or column) by a number other than zero;*

(iii) *the addition to the elements in one row (or column) of a multiple of the elements in another row (or column).*

Simple examples of (ii) and (iii) are:

$$\text{(ii)} \quad \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \quad \text{transformed to} \quad \begin{bmatrix} 1 & 3x \\ 2 & 4x \end{bmatrix},$$

$$\text{(iii)} \quad \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \quad \text{transformed to} \quad \begin{bmatrix} 1 & 3+y \\ 2 & 4+2y \end{bmatrix}.$$

These examples are sufficient to illustrate the point that the 'number' in (ii) and the 'multiple' in (iii) may or may not correspond to numbers in a given field F (Definition 2 does not require the x and y of the examples to belong to any particular field) and it is as well, before going farther, to give a precise definition of an elementary transformation *within a given field.*

DEFINITION 3. *The* ELEMENTARY TRANSFORMATIONS *of a matrix* WITHIN A GIVEN FIELD F *are*

(i) *the interchange of two rows or of two columns;*

(ii) *the multiplication of the elements of a row (or column) by a number, other than zero, belonging to* F;

(iii) *the addition to the elements in one row (or column) of a multiple of the elements in another row (or column), the multiple in question being a number that belongs to* F.

I have thought it worth while to stress this point because attention to it helps one to realize the difference between what one can and cannot do when the transformations are restricted to those within a given field. For example, on denoting 'transforms into' by the symbol $\rightarrow$,

$$\begin{bmatrix} 3+\sqrt{5} & 6+2\sqrt{5} \\ 1 & 4 \end{bmatrix} \rightarrow \begin{bmatrix} 3+\sqrt{5} & 0 \\ 1 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 3+\sqrt{5} & 0 \\ 0 & 2 \end{bmatrix}$$

by transformations within the field of rational numbers—the steps are col. $2-2\times$(col. 1) and col. $1-\frac{1}{2}\times$(col. 2). On the other hand,

$$\begin{bmatrix} 3+\sqrt{5} & 7+2\sqrt{5} \\ 1 & 4 \end{bmatrix} \rightarrow \begin{bmatrix} 3+\sqrt{5} & 0 \\ 1 & x \end{bmatrix}$$

is not possible by transformations within this field; to obtain a zero in the position shown, we must use an irrational multiple.

1.3. *Elementary transformations as matrix multiplications*

Throughout this sub-section A denotes a square matrix whose ith row is

$$a_{i1} \quad a_{i2} \quad \ldots \quad a_{in}$$

and I denotes the unit matrix of order n.

(i) *The interchange of two rows.*

Let I_{ij}, A_{ij} denote the result of interchanging the ith and jth rows of I and of A respectively. Then

$$I_{ij} A = A_{ij}.$$

For example, with $n = 4$, $i = 2$, $j = 4$,

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} a_{11} & . & . & a_{14} \\ a_{21} & . & . & a_{24} \\ a_{31} & . & . & a_{34} \\ a_{41} & . & . & a_{44} \end{bmatrix} = \begin{bmatrix} a_{11} & . & . & a_{14} \\ a_{41} & . & . & a_{44} \\ a_{31} & . & . & a_{34} \\ a_{21} & . & . & a_{24} \end{bmatrix}.$$

Similarly, if I^{ij}, A^{ij} denote the matrices obtained by interchanging the ith and jth columns,

$$AI^{ij} = A^{ij}.$$

Before proceeding we note that

(a) $I_{ij} \equiv I^{ij}$, so that $I_{ij} \equiv (I_{ij})'$;
(b) $I_{ij}^2 \equiv I$, so that $I_{ij} \equiv (I_{ij})^{-1}$;
(c) the value of the determinant $|I_{ij}|$ is ± 1.

We note also that pre-multiplication by I_{ij} interchanges two rows of A and post-multiplication interchanges two columns, while the product

$$I_{ij}\, A I^{ij} \tag{1}$$

is given by interchanging two rows of A and then interchanging two columns in the result. Further, by (a) and (b) above, (1) may be written in either of the forms

$$I_{ij}\, A (I_{ij})^{-1}, \tag{2}$$

$$(I^{ij})' A (I^{ij}). \tag{3}$$

(ii) *The multiplication of the elements of a row by a number other than zero.*

Let K_i be the matrix derived from I by replacing the 1 in the ith diagonal position by k. Then

$$K_i A$$

is the result of multiplying the ith row of A by k; for example,

$$\begin{bmatrix} 1 & . & . \\ . & 1 & . \\ . & . & k \end{bmatrix} \times \begin{bmatrix} a & b & c \\ d & e & f \\ x & y & z \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ kx & ky & kz \end{bmatrix}.$$

Similarly, AK_i is the result of multiplying the ith column of A by k. Again, as in (i), pre-multiplication affects rows and post-multiplication affects columns.

We note the properties

(a) K_i is a symmetrical matrix and its reciprocal $(K_i)^{-1}$ is a matrix of the same type having k^{-1} instead of k in the ith diagonal place.

(b) When k is a number belonging to a given field F, the elements of K_i all belong to F, for every field must contain 1. [We note this obvious point because we shall wish to quote it later: it has, of course, no importance in itself.]

(c) The value of the determinant $|K_i|$ is k.

(iii) *The addition to the elements in one row of a multiple of the elements in another row.*

Let $i \neq j$; let H_{ij} be the matrix having all its elements zero save for the one non-zero element, h, where the ith row crosses the jth column. Then

$$H_{ij}A \quad (A \equiv [a_{rs}])$$

is a matrix whose ith row is h times the jth row of A and whose other rows consist of zeros: for example,

$$\begin{bmatrix} . & . & . \\ . & . & h \\ . & . & . \end{bmatrix} \times \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} . & . & . \\ ha_{31} & ha_{32} & ha_{33} \\ . & . & . \end{bmatrix}.$$

It follows that the result of adding h times the jth row of A to its ith row is given by

$$(I+H_{ij})A. \tag{4}$$

Similarly AH_{ij} is a matrix whose jth column is h times the ith column of A and whose other columns consist of zeros, while the result of adding h times the ith column of A to its jth column is

$$A(I+H_{ij}). \tag{5}$$

We note the properties

(a) $H_{ij}^2 = 0, \quad (I+H_{ij})(I-H_{ij}) = I.$

(b) The elements of $I+H_{ij}$ belong to a given field F if and only if the multiple h belongs to F.

(c) The value of the determinant $|I+H_{ij}|$ is unity whatever be the value of h. This is particularly important when we come to deal (Chap. III) with λ-matrices, where h will be a polynomial in a variable λ; the value of the determinant $|I+H_{ij}|$

will still be unity and hence will be independent of λ (the major point at issue).

1.4. *Details of notation*

We introduce the definition that follows with a view to convenience of diction at a later stage.

DEFINITION 4. *When XA or AX is the result of an elementary transformation of A, we say that X is the matrix of that transformation.*†

When X is an I_{ij} or I^{ij}, we say it is of Type I;
when X is a K_i, we say it is of Type II;
when X is an $I+H_{ij}$, we say it is of Type III.

Moreover

(i) when X is of Type I, it is its own reciprocal [as is obvious from the nature of things; to interchange the ith and jth rows twice in succession is to leave the rows unaltered];

(ii) when X is of Type II, its reciprocal is also of Type II [we merely replace k by k^{-1}];

(iii) when X is of Type III, its reciprocal is also of Type III [the reciprocal of $I+H_{ij}$ is, by (a), $I-H_{ij}$ and the latter is got from the former by writing $-h$ instead of $+h$].

To sum up, the matrix X of an elementary transformation is non-singular; its reciprocal X^{-1} is also the matrix of an elementary transformation; moreover, X and X^{-1} are of the same type.

1.5. *Sequences of elementary transformations*

Let $X_1,\ldots, X_p$ and $Y_1,\ldots, Y_q$ be the matrices of elementary transformations; let

$$U = X_1 \ldots X_p, \qquad V = Y_1 \ldots Y_q,$$

and

$$B = UAV.$$

Then B is a matrix derived from A by a sequence of elementary transformations. Now U and V are non-singular and have reciprocals given by

$$U^{-1} = X_p^{-1} \ldots X_1^{-1}, \qquad V^{-1} = Y_q^{-1} \ldots Y_1^{-1}$$

† We shall not normally need to specify whether the product in question is XA or AX.

and, since $X_p^{-1},..., Y_1^{-1}$ are the matrices of elementary transformations (§ 1.4) and

$$A = U^{-1}BV^{-1},$$

A is derived from B by a sequence of elementary transformations.

Hence, if B is derived from A by a sequence of elementary transformations, then A can be derived from B by a sequence of elementary transformations. This fact justifies the definition now to be given.

1.6. *Formal definition of equivalence*

DEFINITION 5. *Two matrices are equivalent if it is possible to pass from one to the other by a sequence of elementary transformations.*

As we saw in § 1.5, when A and B are equivalent there are non-singular matrices U and V for which

$$UAV = B, \qquad U^{-1}BV^{-1} = A. \tag{6}$$

Sometimes it is of importance to know whether the transformations envisaged in Definition 5 lie within a given field F.

DEFINITION 6. *Two matrices are equivalent in a given field* F *if it is possible to pass from one to the other by a sequence of elementary transformations lying within that field.*

Our previous work justifies such a definition. For if we can go from A to B by elementary transformations in F that use numbers $k_1, k_2,...$ for transformations of Type II and multiples $h_1, h_2,...,$ for transformations of Type III, then we can go from B to A by elementary transformations that involve the numbers $1/k_1, 1/k_2,...$ in Type II and multiples $-h_1, -h_2,...$ in Type III; moreover, $1/k$ and $-h$ belong to F provided that k and h belong to F.

Further, when A and B are equivalent in F there are non-singular matrices† U and V, with all their elements in F, that satisfy

$$UAV = B, \qquad U^{-1}BV^{-1} = A. \tag{7}$$

1.7. *Equivalent matrices have the same rank*

Since U and V in (6) are non-singular matrices, the ranks of A and B are equal.‡

† Write U, V as products, as in § 1.5: each X_r and Y_r has its elements in F.
‡ Chap. I, § 10.

As we shall see in § 2, the converse theorem is true; two square matrices each of order n and rank r are equivalent.

1.8. *Equivalence is transitive*

From the definition of the term 'equivalence', if A is equivalent to B and B is equivalent to C, then A is equivalent to C.

2. Properties of equivalent matrices

[Throughout this section, unless the contrary is expressly stated, the letters A, B denote square matrices of order n: and all numbers belong to a given field F.]

2.1. THEOREM 1. *Let the elements of A belong to the field* F *and let the rank of A be $r < n$. Then A is equivalent in* F *to the matrix*

$$J = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix},$$

where the submatrix I_r is the unit matrix of order r and the zeros denote null submatrices.

PROOF. By hypothesis, A contains a minor of order r whose determinant is not equal to zero. Hence appropriate changes of rows and columns (if such be necessary) make A equivalent to

$$A_1 = \begin{bmatrix} M & P \\ Q & R \end{bmatrix},$$

where M has r rows and columns and $|M| \neq 0$. The $(r+p)$th row of A_1 is linearly dependent in F on the first r rows, and so the $(r+p)$th row can be made a row of zeros by subtracting suitable multiples of the first r rows. Thus A_1 is equivalent in F to

$$A_2 = \begin{bmatrix} M & P \\ 0 & 0 \end{bmatrix},$$

where the zeros denote null submatrices having $n-r$ rows.

On working in a similar way with columns, A_2 is equivalent in F to

$$A_3 = \begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix},$$

which is therefore equivalent in F to A.

Suppose M is, in full,

$$\begin{bmatrix} a_1 & b_1 & . & . & k_1 \\ a_2 & b_2 & . & . & k_2 \\ . & . & . & . & . \\ a_r & b_r & . & . & k_r \end{bmatrix}.$$

At least one number in the first row is not zero, since $|M| \neq 0$, and by a preliminary change of columns in A_3, if such be necessary, we may suppose $a_1 \neq 0$. By subtracting suitable multiples of the first column of A_3 from the second, third,..., rth columns, we obtain a matrix A_4, equivalent to A_3, having

$$\begin{bmatrix} a_1 & 0 & . & . & 0 \\ a_2 & \beta_2 & . & . & \kappa_2 \\ . & . & . & . & . \\ a_r & \beta_r & . & . & \kappa_r \end{bmatrix}$$

in its top left-hand corner and having zeros elsewhere.

On working by rows, A_4 is equivalent to A_5, a matrix having in the top left-hand corner

$$\begin{bmatrix} a_1 & 0 & . & . & 0 \\ 0 & \beta_2 & . & . & \kappa_2 \\ . & . & . & . & . \\ 0 & \beta_r & . & . & \kappa_r \end{bmatrix};$$

moreover one at least of $\beta_2,..., \kappa_2$ is not zero (or the rank of A_5 would be less than r).

We can proceed step by step until we reach an A_n, equivalent in F to A, having

$$\text{diag}\{\alpha_1, \alpha_2,..., \alpha_r\}$$

in the top left-hand corner and having zeros everywhere else. A final set of elementary transformations, multiplying the rows by $\alpha_1^{-1}, \alpha_2^{-1},..., \alpha_r^{-1}$, shows that A is equivalent in F to J.

COROLLARY. *When A is of rank n, it is equivalent in F to I_n.*

2.2. THEOREM 2. *A and B are square matrices of order n with elements in F; each is of rank r. Then A and B are equivalent in F.*

PROOF. When $r < n$, both A and B are equivalent in F to the matrix J of § 2.1. When $r = n$, both A and B are equivalent in F to I_n. Hence, by § 1.8, A and B are equivalent in F.

2.3. THEOREM 3. *The matrix A is equivalent in* F *to any matrix*

$$RAS,$$

where R, S are non-singular matrices of order n with elements in F.

We have only to observe that RAS has the same rank as A and Theorem 3 follows at once from Theorem 2.

2.4. *Note on Theorems* 1–3

The very generality of the results shows that we shall gain but little knowledge of any particular matrix by studying the set of equivalent matrices. The crux of the matter is that rank is preserved, that and nothing else save the number of rows and columns.

2.5. *Bilinear forms*

Let
$$A(x,y) \equiv a_{ij}x_i y_j,$$
a bilinear form in the $2n$ variables

$$x_1, \quad x_2, \quad \ldots, \quad x_n, \qquad y_1, \quad y_2, \quad \ldots, \quad y_n,$$

be transformed by the substitutions

$$x = RX, \qquad y = SY,$$

in which x, y, X, Y denote single-column matrices with elements x_i, y_i, X_i, Y_i and R, S are non-singular square matrices.

The bilinear form is† $x'Ay$ and transforms into $X'BY$, where $B = R'AS$. The matrices A and B are equivalent (Theorem 3) and the bilinear forms are also said to be equivalent.

But again the generality is too great: it is not often that one is interested in transforming a bilinear form by means of a substitution R acting on the x and a completely unrelated substitution S acting on the y. The problem gains in interest and importance when the R and S are related, e.g. when they are transposes or reciprocals.

† F. 126; $x' = X'R'$ and so $x'Ay = X'R'ASY = X'BY$.

3. Rectangular matrices

When A has n_1 rows and n_2 columns,
X has n_1 rows and n_1 columns,
and Y has n_2 rows and n_2 columns,
the product XA has n_1 rows and n_2 columns, and XAY has n_1 rows and n_2 columns.

Any elementary transformation of A that affects its rows is obtained by forming a product XA, where X is a square matrix of order n_1 and is of one of the three types considered in § 1.3. Similarly, an elementary transformation of a matrix B that has n_1 rows and n_2 columns is obtained by forming a product BY, where Y is a square matrix of order n_2 and is of one of the three types considered in § 1.3.

As in § 1.5, a matrix B derived from A by a sequence of elementary transformations can be expressed as

$$X_1 \dots X_p A Y_1 \dots Y_q$$

or

$$UAV,$$

where U and V are non-singular square matrices, U is of order n_1 and V is of order n_2. Also, $A = U^{-1}BV^{-1}$. Thus Theorems 1 and 2 apply without change to a rectangular matrix A, while Theorem 3 becomes

Let R, S be non-singular square matrices of orders n_1, n_2 respectively, let A have n_1 rows and n_2 columns, and let the elements of all three matrices belong to some given field F. *Then the product RAS is equivalent to A in* F.

4. A note on later chapters

Theorem 3 shows that equivalence may also be defined thus:

Let A be a square matrix of order n. Let R, S be any two non-singular matrices of order n. Then the matrix RAS is equivalent to A.

The one thing that is preserved under this very general kind of equivalence is rank and order, and, as an instrument of investigation into the properties of matrices associated with particular algebraic forms, this kind of equivalence is too general to be of much use. In later chapters we investigate the problem of equivalence when R and S are related by some definite law. For example, to anticipate a result proved in Chapter IV, when S is any non-singular matrix, $R = S^{-1}$, and two matrices A and B are equivalent in the sense that

$$B = S^{-1}AS,$$

it will follow that the matrices A and B have the same characteristic equation; in these circumstances, not only rank and order, but also the eigenvalues† or characteristic roots are preserved.

Before proceeding to these special kinds of equivalence, we consider (Chap. III) the general kind of equivalence for matrices whose elements are functions of a variable λ.

† The names 'latent roots', 'characteristic roots', 'eigenvalues' are in general use for what we called (Chap. I, p. 12) the latent roots.

CHAPTER III

EQUIVALENT λ-MATRICES

1. Definitions

In this chapter we shall be concerned with matrices whose elements are polynomials in a variable λ with coefficients in a given field F. To avoid constant repetition in our explanations, we state at the outset that, throughout this chapter, ALL COEFFICIENTS OF POLYNOMIALS, ALL 'NUMBERS', AND ALL 'CONSTANT MULTIPLES' REFERRED TO IN THE TEXT BELONG TO A GIVEN FIELD F.

DEFINITION 7. *A matrix $A = [a_{ik}]$ in which some or all of the elements a_{ik} are polynomials in a variable λ is called a λ-matrix.*

The concept is a generalization from the particular matrix

$$A - \lambda C = [a_{ik} - \lambda c_{ik}],$$

in which each element is a linear function of λ.

The determinant $|A|$ of a λ-matrix is, in general, a polynomial in λ; it may be a constant independent of λ. In particular this constant may be zero, and then $|A| = 0$ for all values of λ: we usually indicate this by saying '$|A|$ is identically zero'.

DEFINITION 8. *The λ-matrix A is said to be singular when $|A|$ is identically zero and to be non-singular when $|A|$ is not identically zero.*

DEFINITION 9. *The λ-matrix A is said to be of rank r ($\geqslant 1$) when r is the largest integer for which we can state that 'not all minors of order r are identically zero'.*

2. Elementary λ-transformations

2.1. The elementary λ-transformations of a matrix, usually though not necessarily a λ-matrix, are

(i) the interchange of two rows or of two columns;

(ii) the multiplication of the elements in a row (or column) by a constant other than zero;

(iii) the addition to the elements in one row (or column) of the elements in another row (or column) multiplied by a polynomial in λ or by a constant.

The only difference between λ-transformations and those in Chapter II is that we now admit in (iii) as a 'multiple' not only constants in F but also polynomials in λ with coefficients in F. In (ii) the multiplier is still a constant, independent of λ.

2.2. *Transformation as matrix multiplication*

As in Chapter II, §§ 1.3–1.5, an elementary λ-transformation of a matrix A may be expressed as a matrix product, a product XA when the transformation affects rows and a product AX when it affects columns; further, the determinant $|X|$ is a non-zero constant. In particular (we quote from Chap. II, § 1.4),

(i) when X is of Type I, it is its own reciprocal and $|X| = 1$;

(ii) when X is of Type II and $|X| = k$, its reciprocal is also of Type II and $|X^{-1}| = k^{-1}$;

(iii) when X is of Type III, say $X = I+H_{ij}$, where H_{ij} has a polynomial in λ or a non-zero constant at the cross of the ith row and jth column and has zeros elsewhere, its reciprocal is $I-H_{ij}$, which is also of Type III, and $|X| = 1$.

The one point of difference from Chapter II is that an X of Type III is now, in general, a λ-matrix; but the value of the determinant $|X|$ is still unity.

2.3. *Sequences of elementary λ-transformations*

Let $X_1,..., X_p$ and $Y_1,..., Y_q$ be the matrices of elementary λ-transformations; let

$$U = X_1 ... X_p, \qquad V = Y_1 ... Y_q.$$

Let
$$UAV = B, \tag{1}$$

so that B is a matrix derived from A by a sequence of elementary λ-transformations. Then

$$|U| = |X_1| ... |X_p|, \qquad |V| = |Y_1| ... |Y_q|,$$

and the values of $|U|$ and $|V|$ are (by § 2.2) non-zero constants. The matrices U and V have reciprocals

$$U^{-1} = X_p^{-1} ... X_1^{-1}, \qquad V^{-1} = Y_q^{-1} ... Y_1^{-1};$$

also
$$U^{-1}BV^{-1} = A. \tag{2}$$
Since the reciprocals X^{-1} and Y^{-1} are the matrices of elementary λ-transformations (§ 2.2, where reciprocals are shown to correspond type for type), (2) shows that A can be derived from B by a sequence of elementary λ-transformations.

Hence, if B is derived from A by a sequence of elementary λ-transformations, then A can be derived from B by a sequence of λ-transformations.

2.4. *Definition of λ-equivalence*

DEFINITION 10. *Two matrices are λ-equivalent if it is possible to pass from one to the other by a sequence of elementary λ-transformations.*

As for ordinary matrices, the relation (1) above, in which U and V are non-singular, shows that λ-equivalent matrices have the same rank (Chap. I, § 10).

2.5. Before proceeding, we note one or two points of detail.

(*a*) We have seen, in § 2.3, that when A and B are λ-equivalent, there are matrices U and V for which
$$UAV = B \tag{1}$$
and the determinants $|U|$ and $|V|$ are non-zero constants. At a later stage we shall prove the converse, namely, if B is defined by (1) and the determinants $|U|$ and $|V|$ are non-zero constants, then A and B are λ-equivalent.

(*b*) *If A is λ-equivalent to B and B is λ-equivalent to C, then A is λ-equivalent to C.* This proposition is inherent in the definition of λ-equivalence. Obvious as it is, the proposition is worth noting explicitly, since in the sequel we shall encounter long chains of λ-equivalent matrices and it will be important to realize that the last member of the chain is a λ-equivalent of the first.

3. A fundamental lemma

3.1. *Preliminary.* The first main objective of this chapter is to show that a λ-matrix of rank r is λ-equivalent to a matrix
$$\left[\begin{array}{c:c} E & O \\ \hdashline O & O \end{array}\right],$$

wherein the O's denote null submatrices and E is a diagonal matrix, of order r, whose elements along the diagonal are either non-zero constants or polynomials in λ. We shall build up our proof of this from the lemma of § 3.2.

We first need a word or two of clarification about factors. Let $f(\lambda)$, $f_1(\lambda)$ be two polynomials in λ with coefficients in F; in particular cases either f or f_1, or both, may be constants in F, but f may not be identically zero.

When $$f_1 \equiv 0, \qquad f_1 \equiv 0 . f.$$

When f_1 is not identically zero,

EITHER there is a unique polynomial $q(\lambda)$, with coefficients in F, for which
$$f_1 \equiv qf, \tag{3}$$

OR there is a unique pair of polynomials $q(\lambda)$ and $r(\lambda)$, with their coefficients in F, for which†
$$f_1 \equiv qf + r$$
and the degree of r is less than the degree of f.

NOTE. q and r may be constants, but, as we have set out the various possibilities, r cannot be identically zero, though q may be.

When (3) holds, we say that f is a factor of f_1. It is perhaps worth stressing that the above requires the coefficients of all polynomials concerned to belong to a field F and does not require them to be integers. For example, to take two extreme cases,

(a) $$x + \tfrac{1}{2} \equiv \tfrac{1}{2}(2x+1)$$
so that, by (3), $2x+1$ is a factor of $x+\frac{1}{2}$;

(b) Let $k \neq 0$ be a constant in F and let
$$f \equiv a_0 \lambda^m + a_1 \lambda^{m-1} + \ldots + a_m$$
be any polynomial (other than simply $0+0+\ldots+0$) with coefficients in F. Then f/k is also a polynomial with coefficients in F. But
$$f \equiv k \times (f/k)$$
and so, under the present definition of 'factor', k is always a factor of f.

The point is important in the proof of Theorem 4 (p. 33).

† We can find q and r from the fact that q is the quotient and r the remainder when f divides f_1.

3.2. *Statement and proof of the lemma*

We first dispose of some details of notation, devised to clarify and shorten the work.

(a) By the FIRST ELEMENT OF A MATRIX we shall mean the element in the first row and first column.

(b) We shall use the symbol r, s to denote 'THE ELEMENT IN THE rth ROW AND sth COLUMN'; e.g. we write 'x is r, s' or 'let r, s be f'.

(c) By a NON-ZERO ELEMENT we shall mean one that is not identically zero; we are not here concerned with the fact that an element may be zero for some particular value of λ.

LEMMA. *Let A be a λ-matrix with a non-zero first element $f(\lambda)$. Then*

EITHER *$f(\lambda)$ is a factor of all other non-zero elements of A,*

OR *there is a λ-equivalent matrix B with a non-zero first element of lower degree than $f(\lambda)$.*

PROOF. Let A have at least one non-zero element of which f is not a factor.

(I) Suppose that the first row of A contains such an element, say

$$f_1(\lambda) \quad \text{is} \quad 1, j,$$

where f is not a factor of f_1. We may write

$$f_1 \equiv qf+r,$$

where r is non-zero and is of lower degree than f.

When $q \equiv 0$, the interchange of columns 1 and j gives a λ-equivalent matrix having r as its first element.

When q is non-zero, add to the jth column $-q$ times the first: in the λ-equivalent matrix so formed

$$r \quad \text{is} \quad 1, j$$

and an interchange of columns gives a λ-equivalent matrix with r as its first element.

Hence, *if the first row or* (*by a similar argument*) *the first column of A contains a non-zero element of which f is not a factor, A is λ-equivalent to a matrix B whose first element is non-zero and is of lower degree than f.*

(II) Now suppose f is a factor of every non-zero element in the first row or first column, but that A has a non-zero element ϕ of which f is not a factor. Suppose that

$$\phi \quad \text{is} \quad i, k \quad (i \neq 1,\ k \neq 1).$$

We consider separately the circumstances

(a) $1, k$ is non-zero, (b) $1, k$ is 0.

(a) Let $1, k$ be non-zero. By hypothesis, it has f as a factor: let it be qf. Add $-q$ times the first column to the kth; in the λ-equivalent matrix so formed, say C,

$$1, 1 \quad \text{is} \quad f; \qquad 1, k \quad \text{is} \quad 0;$$
$$i, 1 \quad \text{is} \quad q_i f \text{ (say)}; \qquad i, k \quad \text{is} \quad f_2,$$

where $$f_2 \equiv \phi - qq_i f.$$

In this, f_2 is non-zero and f is not a factor of f_2 [q_i may be zero, a constant in F, or a polynomial in λ].

Now form from C a λ-equivalent matrix by adding to the elements of the first column the elements of the kth column. In this matrix, D say,

$$1, 1 \quad \text{is} \quad f; \qquad i, 1 \quad \text{is} \quad q_i f + f_2$$

and f is not a factor of the non-zero element in $i, 1$. By (I), D is λ-equivalent to a matrix B whose first element is non-zero and of lower degree than f. Moreover, A is λ-equivalent to C, to D, and to B.

(b) Let $1, k$ be zero. Then in A

$$1, 1 \quad \text{is} \quad f; \qquad 1, k \quad \text{is} \quad 0;$$
$$i, 1 \quad \text{is} \quad q_i f; \qquad i, k \quad \text{is} \quad \phi.$$

We can proceed at once to the second step (from C to D) of the argument in (a), using ϕ instead of f_2. Again A is λ-equivalent to a matrix B of the type required.

We have thus proved that when A contains a non-zero element of which f is not a factor, it is λ-equivalent to a matrix with a non-zero first element of lower degree than f. This establishes the lemma.

4. The normal form of a λ-matrix

We are now in a position to prove a key theorem.

THEOREM 4. *Let A be a λ-matrix of order n and rank $r \geqslant 1$. When $r < n$, A is λ-equivalent to a matrix of the form*

$$N \equiv \begin{bmatrix} E & O \\ O & O \end{bmatrix},$$

where the O's denote null sub-matrices and

$$E = \operatorname{diag}\{E_1, E_2, \dots, E_r\},$$

each E_s being either unity or a polynomial in λ with unity as coefficient of the highest power of λ. Moreover†

E_1 is a factor of E_2, E_2 a factor of $E_3, \dots, E_{r-1}$ a factor of E_r.

When $r = n$, A is λ-equivalent to E above.

PROOF. *First step.* This finds a λ-equivalent matrix A_1, whose first element is non-zero.

The matrix A has at least one non-zero element (since $r \geqslant 1$). Accordingly,

EITHER the first element of A is non-zero, when we take

$$A_1 \equiv A,$$

OR the first element is zero, but some other element x is non-zero. We then take A_1 to be the matrix obtained from A by the interchange of rows or of columns, possibly both, necessary to make x the first element.

Let the first element of A_1 be $f_1(\lambda)$.

Second step. This finds a λ-equivalent matrix B whose first element is a factor of all non-zero elements of B.

If f_1 is a factor of all non-zero elements of A_1, we define B to be A_1. If f_1 is not a factor of all non-zero elements of A_1, then (by the lemma of § 3.2) A_1 is λ-equivalent to a matrix A_2 with a non-zero first element $f_2(\lambda)$ of lower degree than $f_1(\lambda)$.

If f_2 is a factor of all non-zero elements of A_2, we define B to be A_2; if not, A_2 is λ-equivalent to an A_3 with a non-zero first element $f_3(\lambda)$ of lower degree than $f_2(\lambda)$; and so on.

† See the note in § 5.4 (p. 38) on the fact that each E_k is a factor of the succeeding E_{k+1}.

Proceeding thus, we can define B provided that the sequence of matrices

$$A_1, \quad A_2, \quad ...,$$

with non-zero first elements

$$f_1, \quad f_2, \quad ...$$

reaches a point at which f_k is a factor of all non-zero elements of A_k, when the sequence ends.

Now the degrees of $f_1, f_2,...$ decrease at each step. Hence, if the sequence $A_1, A_2,...$ does not end while f is still a polynomial, we shall arrive after a finite number of steps at an f_k that is merely a constant and this constant f_k will be a factor of every non-zero element of A_k [cf. (b) of § 3.1]. Hence the sequence $A_1, A_2,...$ ends, after a finite number of steps, with some A_m whose first element is non-zero and is a factor of all non-zero elements of A_m. We take B to be A_m. Accordingly,

A is λ-equivalent to a matrix B whose first element is non-zero and is a factor of all non-zero elements of B.

NOTE. This first element may be, and often is, a constant.

Third step. This finds a λ-equivalent matrix C given by

$$C \equiv \left[\begin{array}{c|ccc} \beta & 0 & \cdot & \cdot \quad 0 \\ \hline 0 & & & \\ \cdot & & P & \\ \cdot & & & \\ 0 & & & \end{array}\right], \tag{4}$$

in which β is a non-zero element, all other elements in the first row or first column are zero, and β *is a factor of every non-zero element in the submatrix P.*

Let the first row of the matrix B above be

$$\beta \quad q_2\beta \quad q_3\beta \quad ... \quad q_n\beta,$$

in which some of the q's may be identically zero. By taking in succession

$$\text{col. } 2-q_2(\text{col. } 1),$$
$$\text{col. } 3-q_3(\text{col. } 1),$$

and so on, we obtain a λ-equivalent matrix B_1 whose first row is

$$\beta \quad 0 \quad 0 \quad ... \quad 0;$$

also β is still a factor of every non-zero element of this matrix B_1.

By working on the rows of B_1 we obtain a λ-equivalent matrix C of the form (4). Moreover, C is of rank r (being λ-equivalent to A) and β is non-zero; hence P is of rank $r-1$.

Fourth step. This applies to the 2nd, 3rd,..., nth rows and columns the manipulations which the first three steps have applied to the 1st row and column.

(i) By a first step, involving interchange among the 2nd, 3rd,..., nth rows or columns (or both) of C, but leaving the first row and column of C unchanged, we replace P by a sub-matrix P_1 whose first element is non-zero;

(ii) By a second step, involving interchanges as in (i) and adding to rows (or columns) of C multiples of other rows (or columns), but always excluding the first row and column of C from any part in such manipulations, we replace P_1 by a sub-matrix Q whose first element is non-zero and is a factor of all non-zero elements of Q;

(iii) By further adding to rows (or columns) of C multiples of other rows (or columns), again excluding the first row and column of C from such manipulations, we replace Q by a sub-matrix R of the type

$$\begin{bmatrix} \gamma & 0 & 0 & . & . & 0 \\ 0 & \gamma_{33} & \gamma_{34} & . & . & \gamma_{3n} \\ . & . & . & . & . & . \\ 0 & \gamma_{n3} & \gamma_{n4} & . & . & \gamma_{nn} \end{bmatrix},$$

in which γ is non-zero and is a factor of every non-zero γ_{ik}.

Since β, in (4), was a factor of every non-zero element of P and γ is obtained as a sum of multiples of elements of P, β is a factor of γ. Hence C is λ-equivalent to a matrix D given by

$$D \equiv \begin{bmatrix} \beta & 0 & 0 & . & . & 0 \\ 0 & \gamma & 0 & . & . & 0 \\ 0 & 0 & \gamma_{33} & . & . & \gamma_{3n} \\ . & . & . & . & . & . \\ 0 & 0 & \gamma_{n3} & . & . & \gamma_{nn} \end{bmatrix},$$

in which β and γ are non-zero, β is a factor of γ, and γ is a factor of every non-zero γ_{ik}; moreover, the sub-matrix $[\gamma_{ik}]$ is of rank $r-2$ (for D is of rank r).

After r repetitions of this process we arrive at a matrix K, which is λ-equivalent to A and is given by

$$K \equiv \begin{bmatrix} \beta & 0 & . & . & 0 & \\ 0 & \gamma & . & . & 0 & \\ . & . & . & . & . & O \\ 0 & 0 & . & . & \kappa & \\ & & O & & & X \end{bmatrix},$$

in which the r elements $\beta, \gamma,..., \kappa$ are non-zero and each is a factor of its successor. But since K is λ-equivalent to A, its rank is r; and so the sub-matrix X must be a null matrix.

Fifth step. Let $b, c,..., k$ be the coefficients of the highest powers of λ in $\beta, \gamma,..., \kappa$ respectively [if $\beta,..., \delta$ (say) are constants, let $b = \beta,..., d = \delta$]. Multiply the first row of K by $1/b$, the second by $1/c$, and so on. The matrix N so formed is λ-equivalent to K, and so also to A. This establishes the theorem, for

(i) each of the r elements $\beta/b,..., \kappa/k$ in the diagonal of N is either unity or a polynomial in λ having unity as coefficient of its highest power of λ;

(ii) in the sequence

$$\beta/b, \quad \gamma/c, \quad ..., \quad \kappa/k$$

each term is a factor of its successor.

DEFINITION 11. *The form of matrix given in Theorem 4 is called the* EQUIVALENT NORMAL FORM *(sometimes simply the normal form) of the λ-matrix A.*

5. The H.C.F. of minors of order t

5.1. *Preliminary.* In this section we shall prove that the $E_1, E_2,..., E_r$ of Theorem 4 can be determined from a knowledge of the H.C.F. of minors of orders 1, 2,..., r respectively of the original matrix A. We first give a word of explanation about the H.C.F., i.e. the highest common factor, of a number of polynomials with coefficients in a given field F.

Let $f_1(\lambda), f_2(\lambda),..., f_m(\lambda)$ be a number of polynomials with coefficients in F. Then f_1 and f_2 have an H.C.F. g_1 say, which is either unity or is a polynomial in λ with unity as the coefficient of its highest power of λ. Moreover, *this H.C.F. may be determined by operations within* *the field* F *and without factorizing* f_1 *and* f_2.† Also g_1 and f_3 have an H.C.F., g_2 say: and so on until we arrive at g_{m-1}, which is the H.C.F. of the m polynomials $f_1, f_2,..., f_m$.

This H.C.F. is either unity or is a polynomial in λ with unity as the coefficient of its highest power of λ.

5.2. *H.C.F. of t-rowed minors*

THEOREM 5. *When A and B are λ-equivalent matrices, the H.C.F. of the t-rowed minors of A is equal to the H.C.F. of the t-rowed minors of B.*

We shall give two proofs of this important theorem.

FIRST PROOF. Since B is λ-equivalent to A, there are non-singular λ-matrices U, V for which

$$B = UAV.$$

Let B_t, U_t, A_t, V_t denote typical t-rowed minors of B, U, A, V. Then (Chap. I, § 10) every non-zero t-rowed minor of UAV is of the form

$$\sum U_t\, A_t\, V_t,$$

the sum possibly reducing to a single term; that is,

$$B_t = \sum U_t\, A_t\, V_t.$$

Hence, for a given t, every B_t contains as a factor the H.C.F. of the A_t.

But also A is λ-equivalent to B and therefore every A_t contains as a factor the H.C.F. of the B_t. It follows from these two results taken together that the H.C.F. of the A_t is equal to the H.C.F. of the B_t.

SECOND PROOF. This considers the actual change in the values of minors that is effected by a single elementary λ-transformation.

Let B be derived from A by an elementary λ-transformation. Let $G(\lambda)$ be the H.C.F. of the t-rowed minors of A, and $G_1(\lambda)$ the H.C.F. of the t-rowed minors of B. If the elementary transformation is an interchange of rows (or columns) or is the multiplication of a row (or column) by a constant, the values of all minors are either unaltered or multiplied by a non-zero constant. This leaves the H.C.F. unaltered.

† This point is covered by many algebra text-books; e.g. W. L. Ferrar, *Higher Algebra* (Oxford, 1948), p. 222.

Let B be derived from A by an elementary λ-transformation of Type III; say, col. $p+h(\lambda)\times$col. q replaces col. p. The only minors whose values are altered are those that contain elements from col. p and do not contain the corresponding elements from col. q. The value of any such minor, A_t say, is altered to $A_t+h(\lambda)M_t$, where M_t is also a t-rowed minor of A. Each A_t and M_t contains $G(\lambda)$ as a factor: hence

$$G_1(\lambda) \textit{ contains } G(\lambda) \textit{ as a factor.}$$

But A can also be derived from B by an elementary λ-transformation and, therefore,

$$G(\lambda) \textit{ contains } G_1(\lambda) \textit{ as a factor.}$$

It follows that $G \equiv G_1$ and that any sequence of elementary λ-transformations leaves the H.C.F. of the t-rowed minors of a matrix unaltered.

5.3. *Application of Theorem 5 to the normal form*

THEOREM 6. *Let a matrix A, of rank r, have a normal λ-equivalent form*

$$N = \begin{bmatrix} E & O \\ O & O \end{bmatrix},$$

where the O's denote null submatrices and†

$$E = \text{diag}\{E_1, E_2,..., E_r\}.$$

Then the H.C.F.

of the elements of A is	E_1,
of the 2-rowed minors of A is	$E_1 E_2$,
of the 3-rowed minors of A is	$E_1 E_2 E_3$,
. . . .	. . .
of the r-rowed minors of A is	$E_1 E_2 \ldots E_r$.

PROOF. By Theorem 5, the H.C.F. of the elements of A is equal to the H.C.F. of the elements

$$E_1, \quad E_2, \quad ..., \quad E_r.$$

But (Theorem 4) each of the E's is a factor of its successor and therefore the H.C.F. of the E's is E_1.

Let $1 < t \leqslant r$. The H.C.F. of the t-rowed minors of N is the H.C.F. of the polynomials obtained by forming the product of

† When $r = n$, the O's are absent and $N \equiv E$.

any t of the E's. Since each E is a factor of its successor, this H.C.F. is $E_1 E_2 \dots E_t$. By Theorem 5, the H.C.F. of the t-rowed minors of A is also $E_1 E_2 \dots E_t$.

5.4. *Note on the factors* $E_1, E_2, \dots, E_n$

The fact that E_r is a factor of E_{r+1}, which we noted in Theorem 4, is important and is often used in later sections. As Theorem 6 shows, this fact embodies a property of determinants, namely, that the H.C.F.'s of minors of one, two, three,... rows and columns are necessarily of the pattern

$$E_1, \quad E_1 E_2, \quad E_1 E_2 E_3, \quad \dots,$$

where each E_r is a factor of the succeeding E_{r+1}. It may be helpful to establish this in an elementary way.

In the first place, when every element a_{ij} of a determinant A contains E_1 as a factor, every 2-rowed minor of A, such as

$$\begin{vmatrix} a_{is} & a_{it} \\ a_{js} & a_{jt} \end{vmatrix} = a_{is} a_{jt} - a_{it} a_{js},$$

is either zero or contains E_1^2 as a factor. The H.C.F. of 2-rowed minors is therefore $E_1 E_2$, where E_2 contains E_1 as a factor.

Next, with a determinant of order three, say

$$\Delta = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix},$$

let the H.C.F. of the elements be E_1 and the H.C.F. of the co-factors $A_1, \dots, C_3$ be $E_1 E_2$. Then, by Jacobi's theorem,

$$a_1 \Delta = B_2 C_3 - B_3 C_2, \tag{5}$$

which contains $(E_1 E_2)^2$ as a factor. If $a_1 = \alpha_1 E_1, \dots, c_3 = \gamma_3 E_1$, results like (5) show that each term

$$\alpha_1 \Delta, \quad \alpha_2 \Delta, \quad \dots, \quad \gamma_3 \Delta$$

contains $E_1 E_2^2$ as a factor. Moreover, the H.C.F. of $\alpha_1, \alpha_2, \dots, \gamma_3$ is unity (since the H.C.F. of $a_1, a_2, \dots, c_3$ is E_1), and so Δ contains the factor $E_1 E_2^2$. Hence, when the H.C.F. of the elements of a determinant (of order $m \geqslant 3$) is E_1 and of 2-rowed minors is $E_1 E_2$, the H.C.F. of 3-rowed minors contains a factor $E_1 E_2^2$ and is of the form $E_1 E_2 E_3$, where E_2 is a factor of E_3.

Finally, let A be a determinant of order n, say $A = |a_{ij}|$, and let A_{ij} be the cofactor of a_{ij}. Then, by an extension of Jacobi's theorem,†

$$\begin{vmatrix} A_{is} & A_{it} \\ A_{js} & A_{jt} \end{vmatrix} = A.(\text{minor of } A \text{ having } n-2 \text{ rows}). \tag{6}$$

Accordingly, if we suppose that the H.C.F. of $(n-2)$-rowed minors is $E_1 E_2 \dots E_{n-2}$ and of $(n-1)$-rowed minors is

$$E_1 E_2 \dots E_{n-1},$$

(6) gives

$$(E_1 E_2 \dots E_{n-1})^2 \lambda_{isjt} = A.(E_1 E_2 \dots E_{n-2}) \mu_{isjt},$$

where the H.C.F. of the various μ_{isjt} is unity. Hence A contains the factor $E_1 \dots E_{n-2} E_{n-1}^2$. Accordingly, if for any determinant of order $m \geqslant n$, the H.C.F. of $(n-2)$-rowed minors is $E_1 \dots E_{n-2}$ and of $(n-1)$-rowed minors is $E_1 \dots E_{n-1}$, the H.C.F. of n-rowed minors has a factor $E_1 \dots E_{n-2} E_{n-1}^2$ and so is of the form $E_1 E_2 \dots E_n$, where E_n contains E_{n-1} as a factor.

It follows, by induction, that the H.C.F.'s of the non-zero minors of a determinant follow the pattern

$$E_1, \quad E_1 E_2, \quad E_1 E_2 E_3, \quad \dots,$$

where each E_r is a factor of the succeeding E_{r+1}.

6. The invariant factors $E_1, E_2, \dots, E_r$

THEOREM 7. *When A is λ-equivalent to B, the two matrices have the same normal λ-equivalent form N, the same rank r and, for $t = 1, 2, \dots, r$, the same H.C.F. of t-rowed minors.*

Conversely, when A and B have the same rank r and, for $t = 1, 2, \dots, r$, the same H.C.F. of t-rowed minors, the two matrices are λ-equivalent and have the same normal λ-equivalent form N.

PROOF. Let A be λ-equivalent to B. Then, if A is λ-equivalent to N, B is also λ-equivalent to N. The two matrices therefore have the same rank and, when $1 \leqslant t \leqslant r$, the H.C.F. of t-rowed minors of either is $E_1 E_2 \dots E_t$. This proves the first part of the theorem.

Let A and B have the same rank and, for $t = 1, 2, \dots, r$, the

† F. 57; Theorem 18.

same H.C.F. of t-rowed minors, say $D_1, D_2,..., D_r$ respectively. Define $E_1, E_2,..., E_r$ by

$$E_1 = D_1, \quad E_2 = D_2/D_1, \quad ..., \quad E_r = D_r/D_{r-1}.$$

Then (by Theorem 6) both A and B have the same normal λ-equivalent form N given by

$$N = \left[\begin{array}{c|c} E & O \\ \hline O & O \end{array}\right],$$

where $E = \text{diag}(E_1, E_2,..., E_r)$. Further, they are then λ-equivalent to each other. This proves the second part of the theorem.

In view of the fact that the E's are unaltered by any sequence of elementary λ-transformations and are factors of the various H.C.F.'s of minors, the polynomials

$$E_1, \quad E_2, \quad ..., \quad E_r$$

are commonly called the INVARIANT FACTORS of any matrix from which they derive.

When A is a square λ-matrix of order n and is non-singular, $r = n$. There is then only one minor of order n, the determinant $|A|$ itself and the H.C.F. of n-rowed minors is $|A|$; thus, apart from a possible constant factor, $|A|$ is the product of $E_1, E_2,..., E_n$.

7. Alternative definition of λ-equivalence

Let R, S be any two λ-matrices whose determinants $|R|$, $|S|$ are non-zero constants, independent of λ. Let A be a given λ-matrix and let B be given by

$$B = RAS. \tag{7}$$

Then, on solving (7),

$$A = R^{-1}BS^{-1} \tag{8}$$

where R^{-1} and S^{-1} are also λ-matrices, since $|R|$ and $|S|$ are constants (cf. the definition of a reciprocal matrix, Chap. I, § 4; the elements of R^{-1} are polynomials in the elements of R divided by $|R|$).†

Since R, S are non-singular, the matrices A and B have the

† If $|R|$ is a polynomial in λ and is not a constant, the elements of R^{-1} are rational functions of λ with denominators $|R|$: they are not polynomials in λ.

same rank, r say. By a repetition of the argument used in the first proof of Theorem 5, A and B have the same H.C.F. of t-rowed minors for $1 \leqslant t \leqslant r$. Hence, by Theorem 7, the matrices A and B are λ-equivalent.

Conversely, a fact we have had occasion to remark several times already, if A and B are λ-equivalent, there are matrices R and S for which $|R|$ and $|S|$ are constants and the equations (7) and (8) are satisfied.

The theory of equivalent matrices can, in fact, be started from this end, λ-equivalence being defined thus:

Two λ-matrices A and B are said to be λ-equivalent if there are λ-matrices R and S, whose determinants $|R|$ and $|S|$ are equal to non-zero constants, for which

$$B = RAS, \qquad A = R^{-1}BS^{-1}.$$

In this definition all constants and all coefficients of polynomials in λ are understood to belong to a given field F; as we have seen in earlier sections, there are certain points at which the argument breaks down if this proviso is dropped.

8. Elementary divisors

8.1. Let A be a given λ-matrix of rank r and let its λ-equivalent normal form N be given by

$$N = \begin{bmatrix} E & O \\ O & O \end{bmatrix},$$

where $E = \text{diag}\{E_1, E_2,..., E_r\}$, each of $E_1,..., E_{r-1}$ is a factor of its successor, and each of the diagonal elements of E is either unity or a polynomial in λ having unity as the coefficient of its highest power of λ.

As we have had occasion to remark earlier (§ 5.1, p. 35), the E_t can be determined by rational processes within the field F by continued application of the algorithm for finding the H.C.F. of two polynomials; for example, if F is the field of real rational numbers, an invariant factor $E_3 \equiv \lambda^2+2$ appears as such and not as the product $(\lambda+i\sqrt{2})(\lambda-i\sqrt{2})$. On the other hand, it is often convenient to express each E_t as a product of linear factors†

† We must suppose that the field F is one in which a polynomial *can* be expressed as a product of linear factors.

and powers of linear factors, say

$$E_i = (\lambda-\alpha)^{a_i}(\lambda-\beta)^{b_i}(\lambda-\gamma)^{c_i}\dots.$$

When this is done and we have written

$$\begin{array}{ll} E_1 \text{ as the product of} & (\lambda-\alpha)^{a_1}(\lambda-\beta)^{b_1}\dots(\lambda-\kappa)^{k_1}, \\ E_2 \quad ,, \quad ,, \quad ,, & (\lambda-\alpha)^{a_2}(\lambda-\beta)^{b_2}\dots(\lambda-\kappa)^{k_2}, \\ \cdot\ \cdot\ \cdot\ \cdot & \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ E_r \text{ as the product of} & (\lambda-\alpha)^{a_r}(\lambda-\beta)^{b_r}\dots(\lambda-\kappa)^{k_r}, \end{array}$$

such of the factors as are not unity are called the ELEMENTARY DIVISORS of the matrix A.

Since each of $E_1,\dots, E_{r-1}$ is a factor of its successor,

$$a_1 \leqslant a_2 \leqslant \dots \leqslant a_r,$$
$$b_1 \leqslant b_2 \leqslant \dots \leqslant b_r,$$

and so on.

When we know the INVARIANT FACTORS $E_1, E_2,\dots, E_r$ we can, assuming the possibility of solving the appropriate equations in λ, determine all the elementary divisors. Equally, given all the elementary divisors, the appropriate multiplications determine all the invariant factors.

We shall return to the subject of elementary divisors and invariant factors later. Meanwhile, we state a theorem for λ-equivalence expressed in terms of these factors and divisors.

8.2. *A necessary and sufficient condition for λ-equivalence*

THEOREM 8. (i) *Let A, B be square λ-matrices of order n. Then a necessary and sufficient condition that A and B be λ-equivalent is that the two matrices have the same invariant factors.*

(ii) *Equally, a necessary and sufficient condition that A and B be λ-equivalent is that the two matrices have the same set of elementary divisors.*

PROOF. (i) When the invariant factors of a matrix are $E_1, E_2,\dots, E_r$, its rank is r and the matrix is λ-equivalent to the matrix N of Theorem 6. Hence, when two matrices have the same invariant factors, they are both λ-equivalent to the same form N and so are λ-equivalent to each other.

Again, if A and B are λ-equivalent and A is λ-equivalent to a normal form N, then also B is λ-equivalent to N and therefore has the same invariant factors as A.

(ii) This follows at once from (i) in view of the way in which the invariant factors $E_1, E_2,..., E_r$ are expressed (§ 8.1) in terms of the elementary divisors.

9. Matrices that are linear in λ

In this section all matrices are assumed to be square matrices of order n.

9.1. *Preliminary*

Let A be a λ-matrix, each element being a polynomial (possibly a constant) in λ with coefficients in a field F. It is said to be of degree k when λ^k is the highest power of λ that occurs among the elements. Such a matrix may be written in the form

$$A \equiv a_k \lambda^k + a_{k-1}\lambda^{k-1} + ... + a_0 \quad (a_k \neq 0),$$

in which $a_k,..., a_0$ are matrices with elements in F and a_k is not the null matrix; it is a polynomial in λ with matrix coefficients. Conversely, such a polynomial in λ can be expressed as a λ-matrix of degree k.

We recall from Chapter I, § 10, that when b is non-singular, the matrix products

$$ab, \qquad ba$$

have the same rank as a. In particular, when b is non-singular and a is not the null matrix, the rank of the products ab and ba is necessarily positive and, accordingly, neither ab nor ba can then be the null matrix. Hence

LEMMA 1. *When*

$$B = b_l \lambda^l + b_{l-1}\lambda^{l-1} + ... + b_0$$

and $|b_l| \neq 0$, *the products* AB *and* BA *are* λ*-matrices of degree* $k+l$.

The next lemma is an analogue, for polynomials with matrix coefficients, of a well-known result for polynomials with coefficients in a field F (compare Chap. III, § 3.1).

LEMMA 2. *When* b_l *is non-singular, there is a unique pair of matrices* Q_1 *and* R_1 *for which*

$$A \equiv Q_1 B + R_1$$

and either $R_1 \equiv 0$ *or* R_1 *is a* λ*-matrix of degree less than* l *(possibly a constant); and there is a unique pair of matrices* Q_2 *and* R_2 *for which*

$$A \equiv BQ_2 + R_2$$

and either $R_2 \equiv 0$ *or* R_2 *is a* λ*-matrix of degree less than* l *(possibly a constant).*

The reader will see the need for the condition 'b_l is not singular' if, when $k > l$, he makes the first step in the division sum

$$b_l\,\lambda^l + \dots)a_k\,\lambda^k + \dots(a_k\,b_l^{-1}\lambda^{k-l} + \dots;$$

all the terms in the quotient involve the reciprocal of b_l.

9.2. *The equivalence of* $a_1 - \lambda b_1$ *and* $a_2 - \lambda b_2$

We set out to prove the following theorem, one that has many applications in the literature of canonical matrices.†

THEOREM 9. *Let* a_1, a_2, b_1, b_2 *be matrices with elements in* F *and let* b_1 *and* b_2 *be non-singular. Let the matrices*

$$A_1 \equiv a_1 - \lambda b_1, \qquad A_2 \equiv a_2 - \lambda b_2$$

be λ*-equivalent.*

Then there are non-singular matrices p, q *with elements in* F *for which*

$$A_2 \equiv pA_1\,q;$$

moreover,

$$a_2 = pa_1\,q, \qquad b_2 = pb_1\,q.$$

PROOF. We reserve small letters, e.g. p, q, for matrices whose elements are constants in F: a matrix indicated by a capital letter will be a λ-matrix.

Since A_1 and A_2 are λ-equivalent, there are λ-matrices P and Q for which

$$A_2 = PA_1\,Q. \tag{1}$$

The determinants $|P|$ and $|Q|$ are known to be independent of λ, but the matrices P and Q are, in general, λ-matrices. Moreover (cf. § 7, p. 40), Q^{-1} is also a λ-matrix. Let

$$P \equiv A_2\,P_1 + p, \qquad Q^{-1} \equiv S_1\,A_1 + s,$$

† W. V. D. Hodge and D. Pedoe, *Methods of Algebraic Geometry* (Cambridge, 1947), p. 93, give the theorem when one only of the two b's is required to be non-singular.

where, by Lemma 2 and the fact that A_1, A_2 are linear in λ, p and s are constants, possibly zero.

Then, from (1), $A_2 Q^{-1} \equiv PA_1$ and so

$$A_2 S_1 A_1 + A_2 s \equiv A_2 P_1 A_1 + pA_1.$$

On rearrangement this gives

$$A_2(S_1 - P_1)A_1 \equiv pA_1 - A_2 s. \tag{2}$$

Now the right-hand side is, at most, linear in λ, while the left-hand side is

$$(a_2 - \lambda b_2)(S_1 - P_1)(a_1 - \lambda b_1). \tag{3}$$

Suppose that $S_1 - P_1$ is not the null matrix, but is of degree t, where $t \geqslant 0$, in λ. Then, since b_1 and b_2 are non-singular, (3) is of degree $t+2$ in λ and cannot be identically equal to $pA_1 - A_2 s$, which is at most linear in λ. Hence $S_1 - P_1 \equiv 0$ and, from (2),

$$A_2 s \equiv pA_1. \tag{4}$$

We complete the proof by showing that s is non-singular. Let

$$Q \equiv Q_1 A_2 + q.$$

Then, since $I \equiv QQ^{-1}$,

$$I \equiv (Q_1 A_2 + q)(S_1 A_1 + s),$$

$$I - qs \equiv (Q_1 A_2 + q)S_1 A_1 + Q_1 A_2 s,$$

or, on using (4),

$$\begin{aligned} I - qs &\equiv (Q_1 A_2 + q)S_1 A_1 + Q_1 pA_1 \\ &\equiv (Q_1 A_2 S_1 + qS_1 + Q_1 p)A_1. \end{aligned} \tag{5}$$

Since $A_1 \equiv a_1 - \lambda b_1$ and b_1 is non-singular, the right-hand side of (5) is of degree 1 at least in λ unless it is identically zero. But the left-hand side of (5) is a constant and the hypothesis that the right-hand side is not identically zero is therefore untenable. Accordingly

$$I = qs$$

or, what is the same thing, s is non-singular and its reciprocal $s^{-1} = q$.

It follows from (4) that

$$A_2 \equiv pA_1 s^{-1} \equiv pA_1 q$$

and in this p, q are independent of λ.

Moreover, for all λ, this gives

$$a_2-\lambda b_2 \equiv p(a_1-\lambda b_1)q,$$

whence $$a_2 = pa_1 q, \qquad b_2 = pb_1 q.$$

9.3. *The equivalence of $a_1-\lambda I$, $a_2-\lambda I$*

When, in Theorem 9, $b_1 = b_2 = I$, the unit matrix of order n, the condition that b_1 and b_2 are non-singular is automatically satisfied. The resulting theorem is

THEOREM 10. *Let a_1 and a_2 be matrices with elements in* F *and let the matrices*

$$A_1 \equiv a_1-\lambda I, \qquad A_2 \equiv a_2-\lambda I$$

be λ-equivalent. Then there is a non-singular matrix t, with elements in F, *for which*

$$A_2 \equiv tA_1 t^{-1}, \tag{6}$$

and $$a_2 = ta_1 t^{-1}. \tag{7}$$

Conversely, if (7) *holds, then* (6) *also holds and the matrices A_1, A_2 are λ-equivalent.*

PROOF. As in the proof of Theorem 9, there are constant matrices p, q, for which

$$A_2 \equiv pA_1 q, \qquad a_2 = pa_1 q, \qquad I = pIq;$$

and the last of these shows that $q = p^{-1}$. Hence the first part of Theorem 10 follows on putting $t = p$.

Conversely, when (7) holds,

$$a_2-\lambda I \equiv t(a_1-\lambda I)t^{-1},$$

so that (6) follows. Since the determinant $|t|$ is independent of λ, it follows from (6) that A_2 is λ-equivalent to A_1 (cf. § 7, p. 40).

9.4. *Application of invariant factors*

On combining Theorem 8 and Theorem 10 we obtain a result of wide application.

THEOREM 11. *Let a_1 and a_2 be matrices with elements in* F. *Then a necessary and sufficient condition for the existence of a matrix t, with elements in* F, *for which*

$$a_2 = ta_1 t^{-1}$$

is that the two matrices $a_1 - \lambda I$ and $a_2 - \lambda I$ should have the same invariant factors or, what is the same thing, the same set of elementary divisors.

9.5. THEOREM 12. *Let A be a square matrix whose elements are constants. The determinant of the matrix $\lambda I - A$ is equal to the product of the invariant factors of the matrix.*

PROOF. Let the invariant factors of $\lambda I - A$ be

$$E_1(\lambda), \quad E_2(\lambda), \quad \ldots, \quad E_n(\lambda).$$

Then (Theorem 4, with $r = n$, since the determinant $|\lambda I - A|$ is not identically zero), the matrix $\lambda I - A$ is λ-equivalent to the matrix

$$E \equiv \text{diag}\{E_1(\lambda), E_2(\lambda), \ldots, E_n(\lambda)\},$$

and so (§ 7) there are matrices R and S, whose determinants $|R|$ and $|S|$ are constants, for which

$$\lambda I - A = RES.$$

Hence
$$|\lambda I - A| = kE_1(\lambda)E_2(\lambda)\ldots E_n(\lambda), \tag{8}$$
where k is a constant, and, since the coefficient of the highest power of λ in each $E_i(\lambda)$ is unity, $k = 1$.

COROLLARY. *Let the latent roots of A be $\alpha, \beta, \gamma, \ldots$. Then the elementary divisors of $\lambda I - A$ are powers of $\lambda - \alpha$, $\lambda - \beta$, $\lambda - \gamma, \ldots$.*

For α, β, $\gamma, \ldots$ are the roots of the equation $|\lambda I - A| = 0$ and so
$$|\lambda I - A| \equiv (\lambda - \alpha)^u (\lambda - \beta)^v \ldots. \tag{9}$$
By (8) above and the fact that the elementary divisors are factors of the $E_1(\lambda), \ldots, E_n(\lambda)$, any elementary divisor must be of the form $(\lambda - \rho)^x$ where ρ is one of α, β, $\gamma, \ldots$.

CHAPTER IV

COLLINEATION

1. Introductory

1.1. *The field* F

The elements of all matrices, all 'constants', 'multiples', and 'coefficients' that occur in this chapter belong to a given field F. In the course of the text we refer to this field only when we wish to draw special attention to it; lack of reference to F when a constant or multiple is mentioned does not imply that the constant or multiple may lie outside F.

1.2. *Equivalence restricted by a particular condition*

In Chapter II we considered a very general form of 'equivalence' of two matrices. The outcome of our investigation was that, with the definition of equivalence there adopted, any two matrices of order n having the same rank were equivalent. We proved that

provided R and S are non-singular matrices, all matrices RAS are equivalent to A.

In the present chapter we impose a further condition on R and S; in short, we require R to be S^{-1}. There are at least three points of view from which one can see the force of such a restriction: we note them briefly in § 1.3.

1.3. *The condition* $R = S^{-1}$

(*a*) We know the importance, in many connexions, of the characteristic equation

$$|A-\lambda I| = 0$$

of a matrix A.

When we impose the condition

$$RIS = I$$

or, what is the same thing, $R = S^{-1}$, and then consider two equivalent matrices A and B related by the equation

$$RAS = B,$$

we see that for all values of λ

$$R(A-\lambda I)S = B-\lambda I.$$

This means that when A is equivalent to B, each matrix $A-\lambda I$ is equivalent to the corresponding matrix $B-\lambda I$.

If any permissible values of λ lie outside the given field F, we shall be obliged to adjoin these values to F so as to make an extended field F_1 and then work in the field F_1.

(*b*) Let $\boldsymbol{\xi}$ be a single-column matrix with elements $\xi_1,\ldots, \xi_n$; and so for other letters. Let these elements be the current coordinates of a point in a system of n homogeneous coordinates and let A be a square matrix of order n: then the matrix relation

$$\mathbf{y} = A\mathbf{x} \tag{1}$$

expresses a relation between the variable point $\mathbf{x}$ and the variable point $\mathbf{y}$.

Now let the coordinate system be changed from $\boldsymbol{\xi}$ to $\boldsymbol{\eta}$ by means of a transformation given by

$$\boldsymbol{\eta} = T\boldsymbol{\xi}, \tag{2}$$

where T is a non-singular square matrix. The new coordinates, $\mathbf{X}$ and $\mathbf{Y}$, of the points $\mathbf{x}$ and $\mathbf{y}$ are then given by

$$\mathbf{X} = T\mathbf{x}, \quad \mathbf{Y} = T\mathbf{y}. \tag{3}$$

On substituting from (3) in the relation (1),

$$T^{-1}\mathbf{Y} = AT^{-1}\mathbf{X};$$

that is,

$$\mathbf{Y} = TAT^{-1}\mathbf{X}.$$

Thus the effect of replacing A in (1) by a matrix of the type TAT^{-1} amounts to considering the same geometrical relation expressed in a different coordinate system.

Alternatively, we may regard

$$\mathbf{X} = T\mathbf{x}, \quad \mathbf{Y} = T\mathbf{y}$$

as equations whereby, using the same system of $\boldsymbol{\xi}$ *coordinates throughout, we relate a new pair of points* $\mathbf{X}, \mathbf{Y}$ *to the old pair of points* $\mathbf{x}, \mathbf{y}$. *The equation*

$$\mathbf{Y} = TAT^{-1}\mathbf{X}$$

then gives the relation between the new pair of points.

(*c*) Let $\mathbf{l}'$ be a single-row matrix with elements $l_1,\ldots, l_n$ and $\mathbf{x}$ a single-column matrix with elements $x_1,\ldots, x_n$, the l's being tangential (or prime) coordinates and the x's point coordinates in any given system of homogeneous coordinates. Express the equation

$$l_1 x_1 + \ldots + l_n x_n = 0,$$

i.e.
$$\mathbf{l}'\mathbf{x} = 0 \tag{4}$$

in a new coordinate system $\mathbf{L}$, $\mathbf{X}$ defined by the point-coordinate transformation

$$\mathbf{X} = T\mathbf{x}$$

or, what is the same thing, $\mathbf{x} = T^{-1}\mathbf{X}$. The equation (4) is now

$$\mathbf{l}'T^{-1}\mathbf{X} = 0$$

and, on writing
$$\mathbf{L}' = \mathbf{l}'T^{-1},$$

this becomes
$$\mathbf{L}'\mathbf{X} = 0. \tag{5}$$

Thus the transformation of coordinates for both points and primes is contained in the two equations

$$\mathbf{X} = T\mathbf{x}, \qquad \mathbf{L}' = \mathbf{l}'T^{-1}. \tag{6}$$

Now the bilinear form

$$a_{ij} l_i x_j$$

is expressed in matrix notation by the single-element matrix $\mathbf{l}'A\mathbf{x}$. In the new system of coordinates defined by (6) this becomes

$$\mathbf{L}'TAT^{-1}\mathbf{X}.$$

Accordingly a bilinear form

$$a_{ij} l_i x_j \tag{7}$$

with matrix A, when expressed in terms of a new coordinate system defined by $\mathbf{X} = T\mathbf{x}$, becomes

$$b_{ij} L_i X_j$$

with a matrix B given by

$$B = TAT^{-1}.$$

That is to say, given a bilinear form $\mathbf{l}'A\mathbf{x}$, which involves both point and tangential coordinates of a system, the effect of replacing A by TAT^{-1} amounts to considering the same bilinear form expressed in a different coordinate system.

2. Definitions

2.1. An equation
$$\mathbf{X} = T\mathbf{x}, \tag{1}$$

wherein T is a non-singular square matrix, is commonly called a COLLINEATION. The name derives from the geometry associated with (1); when $n = 3$ in § 1.3 (c) and the point $\mathbf{x}$ lies on the line $\mathbf{l}'\mathbf{x} = 0$, the point $\mathbf{X}$ lies on the line $(\mathbf{l}'T^{-1})\mathbf{X} = 0$.

As we have seen in (b) and (c) of § 1.3, the effect of a collineation on a matrix A that occurs in a relation $\mathbf{y} = A\mathbf{x}$ or in a bilinear form $\mathbf{l}'A\mathbf{x}$ is to replace A by a matrix TAT^{-1}.

Definition 12 a. *When T is non-singular and*

$$B = TAT^{-1},$$

we say that B is a transform *of A and write*

$$B \sim A.$$

An alternative definition, useful in stressing the particular kind of equivalence involved and the association with collineation, is

Definition 12 b. *When T is non-singular and $B = TAT^{-1}$, we say that B is c-*equivalent *to A.*

Thus, to say that 'B is a transform of A' and to say that 'B is c-equivalent to A' are two ways of saying the same thing. Sometimes we use one form of words, sometimes the other.

We notice a small point in passing. The relation of c-equivalence is reflexive; for if $B = TAT^{-1}$, then

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

Note. The use of the term c-equivalent is not general. Some books use the term 'similar' or refer to equivalence (or transformation) 'within the sub-group of collineations'. Other books disregard all equivalence that is not collineatory, and refer to what we have called c-equivalence as 'equivalence' *tout simple*.

3. Elementary properties of c-equivalence

3.1. *Powers and rational functions of A*

Let $B = TAT^{-1}$. Then

$$B^2 = TAT^{-1}.TAT^{-1} = TAIAT^{-1} = TA^2T^{-1}$$

and, generally, for any positive integer k,

$$B^k = TA^kT^{-1}.$$

Also $$B^{-1} = (TAT^{-1})^{-1} = TA^{-1}T^{-1},$$

on using the law of reversal for reciprocals.

Again, let $f(A)$ denote a polynomial in A. Then, by the above,

$$f(B) = Tf(A)T^{-1}$$

and, on using the law of reversal for reciprocals,

$$\{f(B)\}^{-1} = T\{f(A)\}^{-1}T^{-1}$$

provided the matrix $f(A)$ is non-singular.

It follows that, when $g(A)$ is also a polynomial and the matrix $g(A)$ is non-singular,

$$\frac{f(B)}{g(B)} = T\frac{f(A)}{g(A)}T^{-1}. \tag{1}$$

That is to say, *any rational function of A undergoes the same multiplications, by T and T^{-1}, as A itself.*

3.2. *The characteristic equation*

As we saw in (*a*) of § 1.3, when

$$B = TAT^{-1}$$

and λ is any number,

$$T(A-\lambda I)T^{-1} = B-\lambda I. \tag{2}$$

Now the characteristic equation of a matrix A is

$$|A-\lambda I|.$$

From (2),

$$|B-\lambda I| = |T|.|A-\lambda I|.|T^{-1}|$$
$$= |A-\lambda I|,$$

since $|T|.|T^{-1}| = 1$. Hence

When A and B are c-equivalent they have the same characteristic equation.

3.3. *A sequence of transforms*

It is important in the sequel to remember that, if

$$A \sim B \quad \text{and} \quad B \sim C,$$

then $A \sim C$. For, if

$$A = TBT^{-1} \quad \text{and} \quad B = SCS^{-1},$$

then

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

By extension, when

$$A \sim B, \quad B \sim C, \quad \ldots, \quad K \sim L,$$

then also

$$A \sim L.$$

In the sequel we encounter long chains of transforms and the theorem sought is, usually, that the last matrix of the chain is a transform of the first.

4. Jacobi's canonical form

In Chapter II we saw that a square matrix A of rank r was equivalent, with the wide definition of equivalence we were then using, to a matrix having r non-zero elements in the diagonal and zero elements everywhere else. This property remained true for λ-equivalence and the type of matrix considered in Chapter III. We now ask 'Can we still make A equivalent to a matrix of diagonal type when we restrict ourselves to c-equivalence, that is when we are no longer considering matrices RAS but are restricted to matrices RAR^{-1}?' The final answer is 'Yes, when all the roots of $|A-\lambda I| = 0$ are different; sometimes, but not always, when two or more roots are equal'.

There are many ways of proving our main result (Theorem 14), a result that has been known to mathematicians for more than seventy years. Nearly every book one opens on the subject gives a different presentation of the matter and very few of these presentations make easy reading. We shall observe two principles in our own first treatment of the matter, and these are

(*a*) to work wholly in matrices and to avoid appeals to outside theories,

(*b*) to choose the more elementary and obvious in preference to the more advanced and subtle argument.

We reach the final 'classical canonical form' in three stages. Starting from a given square matrix A, we first find a transform of A with all zeros below the diagonal [§ 4.1, (4)]; we next obtain Jacobi's form [§ 4.4, (8)]; and the final form is obtained in § 5. Later [§ 9] we indicate a much shorter treatment that is less elementary in character.

Once the reader has mastered the facts and become familiar with the structure and properties of the matrices involved, he will probably prefer the less elementary to the more elementary at all stages of the work. The main point of the (somewhat lengthy) elementary treatment is to show how simple manipulations of the rows and columns lead inevitably to the classical canonical form of Theorem 14. It is, I know, a confession of my own shortcomings, but I was never able really to understand any of the published proofs of Theorem 14 until I had evolved for myself the simple calculations of §§ 4, 5, and, for the elementary divisor arguments, § 8.

4.1. *Zeros below the diagonal*

Let A be a given square matrix of order n and let its characteristic roots be $\lambda_1,\dots,\lambda_n$, which are not necessarily all different. Adjoin to the field F, in which lie the elements of A, such of the numbers $\lambda_1,\dots,\lambda_n$ as lie outside F and carry out the subsequent work in this enlarged field, say F_1 (cf. § 1.3 a).

Since $|A-\lambda_1 I| = 0$, the equation

$$Ax = \lambda_1 x \tag{1}$$

has a non-zero solution x, a single-column vector with elements

$$x_{11},\quad x_{21},\quad \dots,\quad x_{n1}. \tag{2}$$

Let $X = [x_{ij}]$ be a non-singular† matrix having (2) as its first column. The first column of AX is $\lambda_1 x_{j1}$. Also X^{-1}, the reciprocal of X, is the matrix

$$[X_{ji}/|X|],$$

where X_{ij} is the cofactor of x_{ij} in $|X|$.

Further,

$$X_{j1}x_{j1} = |X|,$$

$$X_{ji}\,x_{j1} = 0 \quad (i \neq 1).$$

Hence the first column of

$$X^{-1}AX$$

is λ_1 followed by $n-1$ zeros. Accordingly, A is c-equivalent to

$$X^{-1}AX = A_1 = \begin{bmatrix} \lambda_1 & b' \\ o & B \end{bmatrix}, \tag{3}$$

† Such can always be constructed. If $x_{11} \neq 0$ we put unity in the leading diagonal places after the first and we put zero everywhere else save in the first column; and an interchange of rows from this pattern will give what we want when $x_{11} = 0$ and, say, $x_{k1} \neq 0$.

wherein b' is a single-row submatrix, o is a column of zeros, and B a matrix of order $n-1$. Moreover, the characteristic roots of B are $\lambda_2,\ldots,\lambda_n$; for, by § 3.2, A_1 has the same characteristic roots as A, namely $\lambda_1, \lambda_2,\ldots,\lambda_n$.

Since λ_2 is a characteristic root of B, there is a non-zero column vector y with elements

$$y_{22}, \quad y_{32}, \quad \ldots, \quad y_{n2}$$

for which $By = \lambda_2 y$. Let Y be a non-singular square matrix having y as its first column; then the first column of $Y^{-1}BY$ (cf. the details of $X^{-1}AX$) is λ_2 followed by $n-2$ zeros. Further, when $Z = \text{diag}\{1, Y\}$,

$$Z^{-1}A_1 Z$$

is a matrix of the form†

$$A_2 = \left[\begin{array}{cc|c} \lambda_1 & \alpha_{12} & R \\ 0 & \lambda_2 & \\ \hline \multicolumn{2}{c|}{O} & C \end{array}\right],$$

wherein O is a two-column matrix of zeros and C a square matrix of order $n-2$ whose characteristic roots are $\lambda_3,\ldots,\lambda_n$.

By its method of derivation

$$A_2 = Z^{-1}A_1 Z = Z^{-1}X^{-1}AXZ$$

is c-equivalent to A. We continue the sequence A, A_1, $A_2,\ldots$, A_n of c-equivalent matrices until we have dealt with all the roots $\lambda_1,\ldots,\lambda_n$. The final matrix is of the form

$$\alpha = \begin{bmatrix} \lambda_1 & \alpha_{12} & \alpha_{13} & . & . & \alpha_{1n} \\ & \lambda_2 & \alpha_{23} & . & . & \alpha_{2n} \\ & . & . & . & . & . \\ & & & & & \lambda_n \end{bmatrix}, \tag{4}$$

in which all elements below the diagonal are zero. As we have seen in § 3.3, (4) is c-equivalent to A.

The form (4) is not unique, for we can arrange $\lambda_1,\ldots,\lambda_n$ in any one of $n!$ orders. We shall suppose in the sequel that, if

† This detail requires a little calculation. In submatrix notation, with o denoting a column of $n-1$ zeros and o' a row of $n-1$ zeros, $Z^{-1}A_1 Z$ is, by definition,

$$\begin{bmatrix} 1 & o' \\ o & Y^{-1} \end{bmatrix} \times \begin{bmatrix} \lambda_1 & b' \\ o & B \end{bmatrix} \times \begin{bmatrix} 1 & o' \\ o & Y \end{bmatrix} = \begin{bmatrix} 1 & o' \\ o & Y^{-1} \end{bmatrix} \times \begin{bmatrix} \lambda_1 & b'Y \\ o & BY \end{bmatrix} = \begin{bmatrix} \lambda_1 & b'Y \\ o & Y^{-1}BY \end{bmatrix}.$$

there are equal roots, these equal roots are not separated as we come along the diagonal of (4).

We go on to modify further the form (4). The next section, § 4.2, is concerned with preliminary details of notation.

4.2. *H changes*

We recall from Chapter II, § 1.3 that

$$P(I+H_{ij})$$

is the matrix obtained from P on replacing c_j (the jth column) by c_j+hc_i; that

$$(I-H_{ij})Q$$

is the matrix obtained from Q on replacing ρ_i (the ith row) by $\rho_i-h\rho_j$; and that

$$I-H_{ij} = (I+H_{ij})^{-1}.$$

This last equation shows that

$$R \equiv (I-H_{ij})P(I+H_{ij})$$

is a transform of P. We refer to R as an H_{ij} change of P and when we perform in succession a number of such changes, e.g. when we consider

$$(I-H_{rs})(I-H_{ij})P(I+H_{ij})(I+H_{rs}) = (I-H_{rs})R(I+H_{rs}),$$

we refer to the result as an H change of P. As in § 3.3, an H change of P is necessarily a transform of P.

4.3. *H changes of* (4)

In the matrix (4) let $\alpha_{ij} \neq 0$: as a glance at (4) shows, $i<j$. When we form the product

$$(I-H_{ij})\alpha(I+H_{ij}), \tag{5}$$

the only elements of α which are altered are those on the lines (1) and (2) of the diagram: those on (1) are changed when we replace c_j by c_j+hc_i, those on (2) when we replace ρ_i by $\rho_i-h\rho_j$. The other elements in the ith row or jth column are unaffected because all elements below the diagonal are zero.

The element a_{ij} is replaced in (5) by

$$\alpha_{ij}+h(\lambda_i-\lambda_j)$$

and when $\lambda_i \neq \lambda_j$ we can choose h so that this becomes zero.

Suppose that, in (4),

$$\lambda_1 = \lambda_2 = \dots = \lambda_i$$

and that $\lambda_j \neq \lambda_i$ when $j > i$. On making a succession of H_{ij} changes with $j = i+1, \dots, n$ and with the appropriate choice of

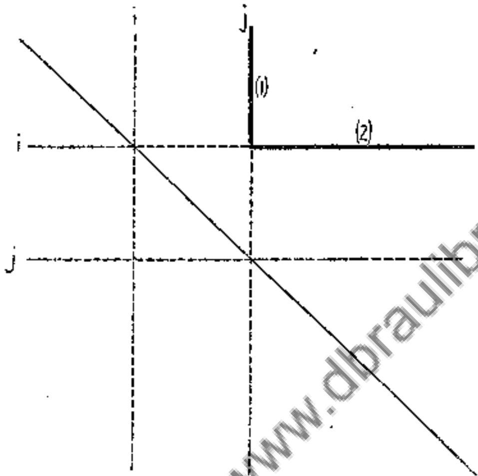


DIAGRAM 1.

h for each, we can replace α by a c-equivalent matrix of the same form as (4) but having zeros at

$$\alpha_{i,i+1}, \quad \alpha_{i,i+2}, \quad \dots, \quad \alpha_{i,n}.$$

Working with the $(i-1)$th row, we can replace

$$\alpha_{i-1,i+1}, \quad \alpha_{i-1,i+2}, \quad \dots, \quad \alpha_{i-1,n}$$

by zeros; and so on until the whole rectangle above the $(i+1)$th row and to the right of the ith column is zeros. Let the resulting matrix be

$$\beta = \left[\begin{array}{ccccc|ccc} \lambda_1 & \beta_{12} & . & . & \beta_{1i} & 0 & . & . & 0 \\ 0 & \lambda_1 & . & . & \beta_{2i} & 0 & . & . & 0 \\ . & . & . & . & . & . & . & . & . \\ 0 & 0 & . & . & \lambda_1 & 0 & . & . & 0 \\ \hline 0 & 0 & . & . & 0 & & & & \\ . & . & . & . & . & & & \gamma & \\ 0 & 0 & . & . & 0 & & & & \end{array}\right], \tag{6}$$

or, in submatrix notation,

$$\beta = \begin{bmatrix} B_1 & O \\ O & \gamma \end{bmatrix},$$

where the latent roots of γ are $\lambda_{i+1},..., \lambda_n$. Then β is an H change of α, and so

$$\beta \sim \alpha \sim A.$$

Before proceeding we note that, if λ_1 is not a repeated root, B_1 is the single-element matrix $[\lambda_1]$ and

$$\beta = \begin{bmatrix} B_1 & o' \\ o & \gamma \end{bmatrix},$$

where o' denotes a row and o a column of zeros.

In the same way we can replace γ by a matrix

$$\begin{bmatrix} B_2 & O \\ O & \delta \end{bmatrix},$$

the transform being obtained by means of H_{rs} changes of β in which $s > r > i$. As a glance at (6) shows, such a change, which involves the operations

$$c_s + hc_r, \qquad \rho_r - h\rho_s,$$

will not affect the first i rows and columns of β. We can continue this process until we obtain a transform of α, and so of A, which has non-zero matrices B_1, B_2,... diagonally placed and has zeros everywhere else.

4.4. *Summary*

What we have proved so far is this:

THEOREM 13. *Let A have k distinct characteristic roots $\lambda_1,..., \lambda_k$ and let λ denote a typical root. When λ is an r-ple root, let B be a matrix of the type*

$$B = \begin{bmatrix} \lambda & b_{12} & . & . & b_{1r} \\ 0 & \lambda & . & . & b_{2r} \\ . & . & . & . & . \\ 0 & 0 & . & . & \lambda \end{bmatrix}, \tag{7}$$

and when λ is a simple root, let B be the single-element matrix $[\lambda]$. Then A is c-equivalent to a matrix

$$J \equiv \text{diag}\{B_1, B_2,..., B_k\} \tag{8}$$

in which $B_1,\ldots, B_k$ are matrices of type (7) corresponding to the k distinct values of λ.

The form (8) above is called, sometimes the INTERMEDIATE FORM, sometimes JACOBI'S CANONICAL FORM. We go on to deduce from it what is commonly called the classical canonical form.

5. The classical canonical form

We showed in § 4 that a given square matrix A was c-equivalent to a matrix of the form

$$J \equiv \begin{bmatrix} B_1 & O & . & . & O \\ O & B_2 & . & . & O \\ . & . & . & . & . \\ O & O & . & . & B_k \end{bmatrix}. \tag{1}$$

In this, the O's represent null submatrices, $B_1,\ldots, B_k$ correspond to the k distinct roots of $|A-\lambda I| = 0$, and each B (corresponding to an r-ple root λ) is of the form

$$B \equiv \begin{bmatrix} \lambda & b_{12} & . & . & b_{1r} \\ 0 & \lambda & . & . & b_{2r} \\ . & . & . & . & . \\ 0 & 0 & . & . & \lambda \end{bmatrix}. \tag{2}$$

We go on to show that each B can be replaced by a form

$$\text{diag}\{C_s(\lambda), C_t(\lambda),\ldots\},$$

where each C is of the particular pattern given by

$$C_1(\lambda) = [\lambda], \quad C_2(\lambda) = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}, \quad C_3(\lambda) = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}.$$

In general

$$C_s(\lambda) = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & . & . & . & . \\ & & & \lambda & 1 \\ & & & & \lambda \end{bmatrix}; \tag{3}$$

in this, each diagonal element is λ, each element immediately above the diagonal is unity, and all other elements are zero.

NOTATION. We indicate a diagonal matrix in which $C_1(\lambda)$ enters p times, $C_2(\lambda)$ enters q times, and so on by the notation

$$\operatorname{diag}\{C_1^p(\lambda), C_2^q(\lambda),...\}.$$

DEFINITION 13. *When $C_1(\lambda), C_2(\lambda),...$ occur in a diagonal matrix, such as that given above, they are sometimes called* ELEMENTARY CLASSICAL SUBMATRICES.

5.1. *G changes*

In addition to the H changes of § 4.2 we shall obtain transforms by interchanges of rows and columns. We recall from Chapter II, § 1.3 that, when I_{ij} is the result of interchanging the ith and jth rows of I,

(i) $$I_{ij} = (I_{ij})^{-1},$$

so that $I_{ij}AI_{ij}$ is a transform of A;

(ii) AI_{ij} is the result of interchanging the ith and jth columns of A;

(iii) $I_{ij}B$ is the result of interchanging the ith and jth rows of B.

We refer to $I_{ij}AI_{ij}$ as a G change of A.

5.2. *The local effect of H and G changes*

Any H or G change of the matrix (1) that affects only the rows and columns which cross in the submatrix B_r leaves all the remaining B's and all the O's unaltered. Accordingly, when we are studying the effect of such changes on J, we may treat them as though they were operations by matrices

$$I+H_{ij}, \quad I-H_{ij}, \quad I_{ij}$$

of the same order as the submatrix B_r and acting solely on this submatrix. This we do in all subsequent work.

5.3. *X changes*

Consider a matrix, of order r,

$$P = \begin{bmatrix} \lambda & p_1 & 0 & . & . & 0 & 0 \\ 0 & \lambda & p_2 & . & . & 0 & 0 \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & \lambda & p_{r-1} \\ 0 & 0 & 0 & . & . & 0 & \lambda \end{bmatrix} \tag{4}$$

and the matrix

$$X = \text{diag}\{1, p_1, p_1 p_2, ..., p_1 p_2 ... p_{r-1}\},$$

where no one of p_1, p_2,... is zero. The matrix XPX^{-1}, when calculated, is found to be $C_r(\lambda)$. We shall refer to it as an X change of P.

Now suppose that P is embedded as a diagonal submatrix in, say,

$$K = \text{diag}\{P_s, P, P_{n-s-r}\}, \tag{5}$$

where P_s, P_{n-s-r} are square matrices of orders s, $n-s-r$. We can replace P by $C_r(\lambda)$ and leave P_s, P_{n-s-r} (and all zero submatrices of K above and below the diagonal submatrices) unaltered by operating on K with a matrix

$$(X) = \text{diag}\{I_s, X, I_{n-s-r}\},$$

where I_s, I_{n-s-r} are unit matrices of orders s, $n-s-r$. As a brief calculation (with submatrices) will show

$$(X)K(X)^{-1} = \text{diag}\{P_s, XPX^{-1}, P_{n-s-r}\}.$$

Thus, the matrix

$$(X)K(X)^{-1} \tag{6}$$

is a transform of K, has $C_r(\lambda)$ where K had P, and has its other elements identical with those of K. We call (6) an X change of K.

We use X changes in our later work.

5.4. *Note on procedure*

We turn our attention to any one B_s of (1). For convenience, we drop the suffix and consider (2) as a typical submatrix of (1).

We make a further simplification, chiefly so that the reader may follow the subsequent argument with a minimum of supplementary calculation. Let B^* be the result of putting $\lambda = 0$ in B. Then

$$B = \lambda I + B^*$$

and, for any non-singular T,

$$T^{-1}BT = \lambda I + T^{-1}B^*T; \tag{7}$$

that is to say, if we obtain a transform $T^{-1}B^*T$ of B^* and insert λ's in the diagonal of the resulting matrix, we obtain the corresponding transform of B. We carry out all our intermediate calculations with B^*.

5.5. *The simple case*

We first consider

$$B^* = \begin{bmatrix} 0 & b_{12} & . & . & b_{1r} \\ 0 & 0 & . & . & b_{2r} \\ . & . & . & . & . \\ 0 & 0 & . & . & 0 \end{bmatrix} \tag{8}$$

when
$$b_{12}\,b_{23}\ldots b_{r-1,r} \neq 0. \tag{9}$$

The transform

$$(I-H_{2j})B^*(I+H_{2j}) \quad (j > 2)$$

(i) leaves unchanged all the zeros in and below the diagonal of B^*;

(ii) leaves the elements b_{12}, b_{23},... unaltered, so that (9) is also true of the transform;

(iii) has $b_{1j}+hb_{12}$ as the jth element in the first row.

The choice $h = -b_{1j}/b_{12}$ makes the jth element in the first row a zero. Thus a succession of H_{2j} changes, involving

$$H_{23}, \quad H_{24}, \quad \ldots, \quad H_{2r},$$

gives a transform of B^* whose first row is

$$0 \quad b_{12} \quad 0 \quad 0 \quad \ldots \quad 0$$

and whose elements b_{23}, b_{34},... are those of the original B^*.

Performing on this transform a succession of H_{3j} changes, involving

$$H_{34}, \quad H_{35}, \quad \ldots, \quad H_{3r}$$

we obtain another transform whose first two rows are

$$\begin{matrix} 0 & b_{12} & 0 & 0 & \ldots & 0 \\ 0 & 0 & b_{23} & 0 & \ldots & 0 \end{matrix}$$

and, proceeding row by row, we obtain

$$B^* \sim P \equiv \begin{bmatrix} 0 & b_{12} & 0 & . & . & 0 & 0 \\ 0 & 0 & b_{23} & . & . & 0 & 0 \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & 0 & b_{r-1,r} \\ 0 & 0 & 0 & . & . & 0 & 0 \end{bmatrix}.$$

A final X change (cf. § 5.3) gives

$$B^* \sim C_r(0),$$

and so, by § 5.4, $$B \sim C_r(\lambda).$$

5.6. *The more difficult case: first step*

We now suppose that

$$b_{12}b_{23}\dots b_{r-1,r} = 0. \qquad (10)$$

If some $b_{1j} \neq 0$, we write $B^* \equiv B''$ (see below). If all the elements of the first row of B^* are zero, we 'isolate' the leading element and write B^*, in submatrix notation, as

$$\begin{bmatrix} C_1(0) & o' \\ o & B' \end{bmatrix},$$

where o' is a row and o a column of zeros. If the first row of B' is all zeros, we 'isolate' its leading element, and, unless B^* is merely $\text{diag}\{C_1^r(0)\}$, we can proceed thus until we reach a B'' with at least one non-zero in its first row. Thus

EITHER $\qquad B^* = \text{diag}\{C_1^r(0)\},$

OR $\qquad B^* = \text{diag}\{C_1^p(0), B''\} \quad (p \geqslant 0),$

$$B'' = \begin{bmatrix} 0 & b''_{12} & b''_{13} & . & . & b''_{1t} \\ 0 & 0 & b''_{23} & . & . & b''_{2t} \\ . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & 0 \end{bmatrix} \quad (p+t = r),$$

and at least one element in the first row of B'' is not zero.

In the latter event let b''_{1j} $(j \geqslant 2)$ be the first non-zero in $b''_{12}, b''_{13},\dots$. Then suitable H_{jk} changes, involving $k = j+1,\dots, r$, give, as in § 5.5, a transform of B'', D say, in which d_{1j} is the only non-zero element in the first row and all zeros in and below the diagonal remain unaltered.

Suppose the second row of D is

$$0 \quad 0 \quad d_{23} \quad \dots \quad d_{2t}.$$

If all are zero, we pass at once to the work of § 5.7, which deals with the third row. If the first non-zero is d_{2l} $(l > 2)$ and $l \neq j$, we proceed thus:

Suitable H_{lk} changes, with $k = l+1,\ldots, t$, give a transform of D in which

(i) the first row is the same as that of D,

(ii) the second row has just one non-zero element d_{2l}.

When $l > j$, that is, when the non-zero in the second row lies to the right of the non-zero in the first, we have attained our immediate aim. When $l < j$, we make the G change 'interchange the first two columns and then the first two rows' (noting that j is now necessarily greater than 3); this transforms

$$\begin{matrix} 0 & 0 & \ldots & . & d_{1j} & \ldots \\ 0 & 0 & \ldots & d_{2l} & . & \ldots \\ . & . & . & . & . & . \end{matrix}$$

into a matrix whose first two rows are

$$\begin{matrix} 0 & 0 & \ldots & d_{2l} & . & \ldots \\ 0 & 0 & \ldots & . & d_{1j} & \ldots \end{matrix}$$

Thus, when $l \neq j$, we secure a transform of B'', say E, whose first two rows are

$$\begin{matrix} 0 & 0 & \ldots & e_{1j} & . & \ldots \\ 0 & 0 & \ldots & . & e_{2l} & \ldots, \end{matrix} \tag{11}$$

the one non-zero element in the second row lying to the right of that in the first row.

It remains to consider what steps are to be taken when $l = j$. This can happen only when $j > 2$. Let the first two rows of D be

$$\begin{matrix} 0 & 0 & \ldots & d_{1j} & 0 & 0 \\ 0 & 0 & \ldots & d_{2j} & d_{2,j+1} & \ldots. \end{matrix}$$

On taking
$$(I-H_{21})D(I+H_{21}) \tag{12}$$
with $h = d_{2j}/d_{1j}$, we obtain a zero in place of d_{2j}: either all the elements of the second row are now zero (when we pass at once to the work of § 5.7) or we can proceed, as before, to the pattern (11).

To sum up, we now have a transform of B'', say E, in which

(i) there is just one non-zero e_{1j} ($j \geqslant 2$) in the first row, and

(ii) either the second row is all zeros or it has just one non-zero element, which lies to the right of e_{1j}.

5.7. *The second step*

We can proceed in this way, row by row, ensuring at each step that a row has at most one non-zero and that each non-zero lies to the right of any non-zero in a previous row.

For example, consider the third row, not all zeros, when e_{1j} and e_{2l} are the non-zeros of the first and second rows. If necessary, make a preliminary transformation (or transformations) of the type (12) to ensure that the first non-zero, e_{3m} say, of the third row is not in the jth or lth column. This done, H changes (as in §§ 5.5 and 5.6) make e_{3m} the only non-zero in the third row. Finally, either the non-zeros move to the right as we descend, or suitable G changes among the first three rows and columns will make it so.†

When we have thus worked through all the rows, we have transformed B'' into

$$P = \begin{bmatrix} . & . & . & p_{1j} & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & p_{sk} & . & . \\ . & . & . & . & . & . & . & . & . & . \end{bmatrix}, \qquad (13)$$

in which

(i) *there is just one non-zero p_{1j}, with $j \geqslant 2$, in the first row;*

(ii) *each remaining row has at most one non-zero and the non-zeros, if there are such, move to the right as we come down the rows.*

It may be noted, in passing, that there are at least $j-1$ rows of zeros when $j > 2$ and that, when $j = 2$, the condition (10) ensures at least two rows of zeros in B''.

5.8. *The third step*

We now transform P by moving its non-zero elements to their appropriate positions, one place to the right of the diagonal. Suppose, for example, that the non-zero in the first row of P is p_{1j}, where $j > 2$. Then

$$Q = I_{j2} P I_{j2}$$

has q_{12} as the only non-zero in its first row. If the jth row of

† Compare the G change that precedes (11).

P is all zeros, then the second row of Q is all zeros and we can consider at once the third row of Q. If P has a non-zero p_{jr}, this appears as a non-zero q_{2r}, possibly lying to the right of the non-zeros in the later rows of Q; we can, by suitable G changes as in §§ 5.6 and 5.7, replace Q by an R in which the non-zeros move to the right as we come down the rows. All elements in and below the diagonal remain zero.

At this stage we have replaced P by a transform, P_1 say, in which EITHER the first two rows are

$$\begin{matrix} \cdot & p_{12} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & p_{2l} & \cdot \end{matrix}$$

and non-zeros (if any) of later rows move to the right as we move down the rows, OR the first two rows are

$$\begin{matrix} \cdot & p_{12} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{matrix}$$

and the non-zeros (if any) of later rows move to the right as we descend.

We can move p_{2l} to the position p_{23} and make the necessary G changes to ensure that succeeding non-zeros move to the right as we come down the rows. The I_{j3} changes necessary to put p_{2l} into the position p_{23} may bring a line of zeros into the third row.

Thus we can, step by step, bring the non-zeros into the positions one place to the right of the diagonal. Let the resulting transform when all non-zeros are in position be denoted by V.

Now V has t rows, of which the last is necessarily a row of zeros: it may have other rows of zeros. Let the kth row be the first row of zeros. When $k = t$, an X change† gives

$$V \sim C_t(0).$$

When $k < t$, we isolate

$$\begin{bmatrix} 0 & v_{12} & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & v_{k-1,k} \\ 0 & 0 & \cdot & \cdot & 0 \end{bmatrix}$$

† Compare § 5.3. In fact, $k = t$ only when (9) holds. When (10) holds V must have at least two rows of zeros (cf. end of § 5.7).

as a leading submatrix, transform it by an X change into $C_k(0)$, and obtain

$$V \sim \begin{bmatrix} C_k(0) & O \\ O & V_1 \end{bmatrix}.$$

In this V_1 is of the same type as V, save that its first row may be all zeros.

There are three possibilities:

(i) V_1 is all zeros, when

$$V_1 = \text{diag}\{C_1^{t-k}(0)\};$$

(ii) V_1 has a non-zero in its first row and so is of exactly the same type as V;

(iii) V_1 has zero rows at first and then at least one non-zero before we reach the last row, when, as for B^* and B'' in § 5.6,

$$V_1 = \text{diag}\{C_1^s(0), V_2\} \quad (s > 0)$$

and V_2 has a non-zero in its first row.

Thus V transforms into one or other of

$$C_t(0), \qquad \text{diag}\{C_k(0),\ C_1^{t-k}(0)\},$$

$$\text{diag}\{C_k(0), V_2\}, \qquad \text{diag}\{C_k(0), C_1^s(0), V_2\},$$

the matrix V_2 being of the same type as V and having a non-zero in its first row. We can apply to V_2 the method used for V and continue until we have exhausted all the rows of V.

Hence, in all circumstances,

$$B'' \sim V \sim \text{diag}\{C_i(0), C_j(0),...\}$$

the number of submatrices C and their suffixes depending on the elements of B''. But, by § 5.6, either

$$B^* = B'',$$

$$B^* = \text{diag}\{C_1^r(0)\},$$

or

$$B^* = \text{diag}\{C_1^p(0), B''\} \quad (p > 0).$$

Hence, in all circumstances,

$$B^* \sim \text{diag}\{C_l(0), C_m(0),...\}. \tag{14}$$

The suffixes l, m,... are not necessarily different and do not necessarily increase as the sequence proceeds. The matrix (14) is, however, fully diagonal when reckoned by submatrices and appropriate G changes readily† give

$$B^* \sim \operatorname{diag}\{C_1^p(0), C_2^q(0),...\}, \tag{15}$$

where $p = 0$ implies that C_1 does not occur, $q = 0$ that C_2 does not occur, and so on for other affixes.

Finally, on referring to § 5.4, it follows that

$$B \sim \operatorname{diag}\{C_1^p(\lambda), C_2^q(\lambda),...\}, \tag{16}$$

where $p+2q+...$ is the multiplicity of the root λ.

Summary. We have thus proved that each matrix $B_1,..., B_k$ of (1) may be transformed into the form (16). Moreover, in the light of § 5.2 (the local effect of H and G changes) and § 5.3 (how to effect an X change of a submatrix embedded in a larger, diagonal, matrix), the transformation of each B can be obtained by transforming the full matrix of which B is a submatrix. We have thus proved that, when J is the matrix (1),

$$J \sim \operatorname{diag}\{D_1, D_2,..., D_k\},$$

where each D_s is of the form given by

$$D = \operatorname{diag}\{C_1^p(\lambda), C_2^q(\lambda),...\}.$$

We have thereby proved the following theorem to be a consequence of Theorem 13.

THEOREM 14. *Let A have k distinct characteristic roots $\lambda_1,..., \lambda_k$ and let λ denote a typical root. Then A is c-equivalent to a matrix*

$$C \equiv \operatorname{diag}\{D_1, D_2,..., D_k\}, \tag{17}$$

in which $D_1,..., D_k$ are submatrices that correspond to the k distinct roots $\lambda_1,..., \lambda_k$, and each D is of the form given by

$$D \equiv \operatorname{diag}\{C_1^p(\lambda), C_2^q(\lambda),...\},$$

where $p+2q+...$ is the multiplicity of the root λ.

The form (17) is usually called the CLASSICAL CANONICAL FORM of the original matrix A.

5.9. *The G changes of* § 5.8

When we think of collineation as a change of variables, as in § 1.3 (*b*), it is all but obvious that a matrix

$$\operatorname{diag}\{C_l(\lambda), C_m(\lambda), C_n(\lambda),...\}$$

† The details are indicated in § 5.9.

can be transformed into one in which any C_1's come first, then C_2's, and so on, but it will round off our present treatment more satisfactorily if we make the actual transformation by means of the appropriate G changes.

We indicate elements $a_{12}, a_{34},\ldots$ by writing only the suffixes, thus: 12, 34,.... We consider

$$A = \left[\begin{array}{ccc|c} 11 & 12 & 13 & 14 \\ 21 & 22 & 23 & 24 \\ 31 & 32 & 33 & 34 \\ \hline 41 & 42 & 43 & 44 \end{array}\right].$$

As a little calculation will show†

$$G_{12} \quad G_{23} \quad G_{34} \quad A$$

transfers the element 44 to the top-left position: in fact

$$G_{12}\,G_{23}\,G_{34}\,A = \left[\begin{array}{c|ccc} 44 & 41 & 42 & 43 \\ \hline 14 & 11 & 12 & 13 \\ 24 & 21 & 22 & 23 \\ 34 & 31 & 32 & 33 \end{array}\right]. \tag{18}$$

This type of G change will bring any $C_1(0)$ in (14) to a top-left position in (15).

Again, let us consider, for example,

$$A = \left[\begin{array}{ccc|cc} 11 & 12 & 13 & 14 & 15 \\ 21 & 22 & 23 & 24 & 25 \\ 31 & 32 & 33 & 34 & 35 \\ \hline 41 & 42 & 43 & 44 & 45 \\ 51 & 52 & 53 & 54 & 55 \end{array}\right].$$

We see at once, from (18), that

$$G_{12}\,G_{23}\,G_{34}\,A = \left[\begin{array}{c|ccc|c} 44 & 41 & 42 & 43 & 45 \\ \hline 14 & 11 & 12 & 13 & 15 \\ 24 & 21 & 22 & 23 & 25 \\ 34 & 31 & 32 & 33 & 35 \\ \hline 54 & 51 & 52 & 53 & 55 \end{array}\right]$$

† We use $G_{34}A$ to indicate the G change $I_{34}AI_{34}$; and so for other suffixes.

and that

$$G_{23}G_{34}G_{45}(G_{12}G_{23}G_{34}A) = \begin{bmatrix} 44 & 45 & 41 & 42 & 43 \\ 54 & 55 & 51 & 52 & 53 \\ 14 & 15 & 11 & 12 & 13 \\ 24 & 25 & 21 & 22 & 23 \\ 34 & 35 & 31 & 32 & 33 \end{bmatrix}.$$

This type of G change will transfer any $C_2(0)$ in (14) to its appropriate place in (15); and the generalization for transferring a submatrix of order three or more is built up in the same way.

5.10. *Deduction of the canonical form from the number of 'chains' in a matrix*

We consider the submatrix B which corresponds [cf. (2)] to an r-ple root λ of the original matrix;

$$B \equiv \begin{bmatrix} \lambda & b_{12} & . & . & b_{1r} \\ 0 & \lambda & . & . & b_{2r} \\ . & . & . & . & . \\ 0 & 0 & . & . & \lambda \end{bmatrix}.$$

First modify this, as in § 5.4, to B^* by putting $\lambda = 0$. Next apply the appropriate H and G changes (as in §§ 5.5, 5.6) to obtain a transform of B^*

$$B^* \sim P = \begin{bmatrix} . & . & p_{1j} & & & \\ & & & & p_{2k} & \\ . & . & . & . & . & . \end{bmatrix}$$

in which each row contains at most one non-zero and the non-zeros move to the right as we come down the rows.

We know (from § 5.8) that this has a transform

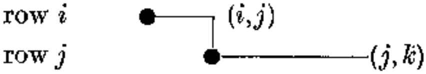
$$C = \text{diag}\{C_l(0), C_m(0), \ldots\}. \tag{19}$$

We can determine the values of l, m,... without recourse to the actual details of transformation.

We consider only the non-zero elements of P and denote a non-zero element in the ith row and jth column by (i,j). We notice that, since the non-zeros of P all lie above the principal diagonal, $j > i$. Two non-zero elements

$$(i,j), \qquad (j,k)$$

are said to be LINKED; diagrammatically,



the heavy dots marking the ith and jth diagonal positions. When m non-zero elements

$$(i,j), \quad (j,k), \quad \ldots, \quad (r,s), \quad (s,t)$$

are linked in succession, we say that they form a CHAIN. We use these links and chains to find the canonical form of P.

If the first row (first l rows) of P has all zeros, we can, by G changes, interchange rows and columns so that this row (rows) comes after a row containing a non-zero. Consider this done (if needed) and suppose that P has a non-zero element $(1,j)$.

(a) Let the jth row of P be all zeros. Make the one G change G_{2j} (i.e. interchange 2nd and jth rows, 2nd and jth columns). The resulting transform of P is

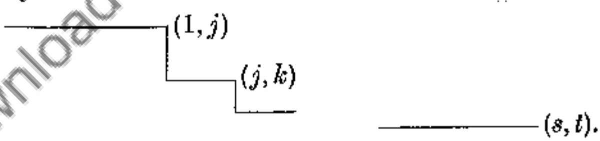
$$P_1 = \left[\begin{array}{cc|c} \cdot & (1,2) & O \\ & \cdot & \\ \hline & O & Q \end{array}\right],$$

which shows that the canonical form of P contains a $C_2(0)$.

(b) Let the non-zero $(1,j)$ be the start of a chain

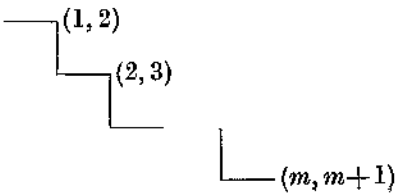
$$(1,j), \quad (j,k), \quad \ldots, \quad (s,t) \quad (m \text{ elements}),$$

the chain ending when the tth row has all zeros. Diagrammatically the chain is



The G changes $\quad G_{2j}, \quad G_{3k}, \quad \ldots, \quad G_{m+1,t}$

performed in succession (interchange 2nd and jth, 3rd and kth,...) change the diagram of the chain to



and give a transform of P

$$P_2 = \left[\begin{array}{cccc|c} \cdot & (1,2) & & & \\ & \cdot & (2,3) & & O \\ & & & (m, m+1) & \\ & & & \cdot & \\ \hline & & O & & R \end{array}\right],$$

which shows that the canonical form of P contains a $C_{m+1}(0)$. Thus

a chain of m non-zero elements in P means that the canonical form of P contains a

$$C_{m+1}(0).$$

All non-zero elements of P not accounted for in the chain just dealt with have been moved down to R, and the interchanges of rows and columns used to move them down cannot affect the linking of elements. A 'staircase' (i) can

(i) (ii)

become a staircase (ii), with a different set of lengths to the separate steps, but the number of steps and the elements linked remain unchanged. Thus, apart from the one chain in P accounted for by the $C_{m+1}(0)$, any chain in P exists as a chain in R. We can repeat the whole process in R and continue until we have used all the distinct chains that occur in P. The final result may be stated thus:

Let P, with n rows and columns, contain k (and only k) distinct chains, having $m_1, m_2,..., m_k$ non-zero elements respectively. Let

$$m_1+m_2+...+m_k = n-r-k.$$

Then the canonical form of P consists of

$$C_{m_1+1}(0), \quad C_{m_2+1}(0), \quad ..., \quad C_{m_k+1}(0)$$

together with the r elements each equal to $C_1(0)$ necessary to make up the order of P.

6. *The Segre characteristic of a matrix*

6.1. Let A be a matrix, of order n, whose distinct characteristic roots are λ_1, λ_2, λ_3,..., where λ_1 is an r-ple root, λ_2 an s-ple root, and so on. Then $r+s+t+...=n$. The classical canonical form of A, established in Theorem 14, may take many shapes, of which the simplest is

$$\mathrm{diag}\{C_r(\lambda_1), C_s(\lambda_2), C_t(\lambda_3),...\}. \tag{1}$$

In this matrix one classical submatrix corresponds to each root. The form (1) is specified, apart from the values of the λ's, by the symbol

$$[r\,s\,t\,...]. \tag{2}$$

When the submatrix D_1 of Theorem 14 is not the single elementary matrix $C_r(\lambda_1)$, but is itself compounded of two or more elementary matrices, say

$$A \sim \mathrm{diag}\{C_x(\lambda_1), C_y(\lambda_1), C_s(\lambda_2), C_t(\lambda_3)...\}, \tag{3}$$

we specify the form of matrix by a symbol

$$[(x\,y)\,s\,t\,...], \tag{4}$$

enclosing in a parenthesis the orders of the submatrices in (3) that correspond to one and the same root; and so for any root λ_k when the corresponding D_k involves more than one elementary matrix.

EXAMPLES. When $n=9$ and the characteristic roots are α, α, β, β, β, γ, γ, γ, γ, the symbol

$$[2\,(1\,2)\,(1\,1\,2)] \tag{5}$$

indicates the matrix

$$\mathrm{diag}\{C_2(\alpha), C_1(\beta), C_2(\beta), C_1(\gamma), C_1(\gamma), C_2(\gamma)\}, \tag{6}$$

while the symbol† $[2\,3\,4]$

indicates the matrix

$$\mathrm{diag}\{C_2(\alpha), C_3(\beta), C_4(\gamma)\}.$$

6.2. It is possible to write down in symbols all the types that can occur when $n=3$, 4, or any other definite number that is not too large. For example, when $n=3$ the matrix A

† It is convenient to put the smaller numbers first, but a symbol [4 3 2] or [4 2 3] indicates the same type of matrix as [2 3 4] apart from an obvious interchange of rows and columns.

has 3 characteristic roots, say α, β, γ, which may or may not be distinct.

(i) Let α, β, γ be all different. The classical form must be

$$\operatorname{diag}\{C_1(\alpha), C_1(\beta), C_1(\gamma)\}, \quad \text{i.e.} \quad [1\,1\,1].$$

(ii) Let $\alpha = \beta \neq \gamma$. The classical form is

either $$\operatorname{diag}\{C_2(\alpha), C_1(\gamma)\}, \quad \text{i.e.} \quad [2\,1],$$

or $$\operatorname{diag}\{C_1(\alpha), C_1(\alpha), C_1(\gamma)\}, \quad \text{i.e.} \quad [(1\,1)\,1].$$

(iii) Let $\alpha = \beta = \gamma$. The classical form may be any one of the three forms

$$C_3(\alpha), \quad \text{i.e.} \quad [3],$$

$$\operatorname{diag}\{C_1(\alpha), C_2(\alpha)\}, \quad \text{i.e.} \quad [(1\,2)],$$

$$\operatorname{diag}\{C_1(\alpha), C_1(\alpha), C_1(\alpha)\}, \quad \text{i.e.} \quad [(1\,1\,1)].$$

A symbol like those in (2), (4), (5), (6), or any of the symbols [1 1 1], [2 1],... above, is called the SEGRE CHARACTERISTIC of the matrix.

7. *A theorem concerning diagonal matrices*

Let $$M = \operatorname{diag}\{M_1, M_2,..., M_k\},$$

where the M_i are square submatrices, possibly single-element submatrices. Think of M as set out in full, each M_i being set out as a square array. In §§ 8 and 10 we shall be concerned with the determinants of minors got by deleting rows and columns from M when this is set out in full. It will be important to recognize at a glance that certain of these determinants are zero.

THEOREM 15. *When we delete t rows and columns from M in such a way that we delete more rows than columns from a given M_i, the determinant of the resulting minor of M is equal to zero.*

PROOF. We may, on changing the order of the rows and columns of M, suppose that M_i is M_1. Let the deletions from M be such that r rows and $r+s$ columns of M_1 remain. The determinant of the resulting minor takes the shape

$$\begin{vmatrix} a_{11} & . & . & a_{1r} & . & . & a_{1,r+s} & \\ & & & & & & & O \\ a_{r1} & . & . & a_{rr} & . & . & a_{r,r+s} & \\ & & & O & & & & N \end{vmatrix},$$

where each O represents a block of zeros and N is an array with more rows than columns. The Laplace expansion of this determinant by its first r rows is manifestly zero.

COROLLARY. *Every non-zero minor† of M is either a minor of a submatrix M_i or is the product of two or more minors of $M_1,..., M_k$.*

8. Elementary divisors and the Segre characteristic

8.1. Let C be a classical canonical form, of order n. We consider the matrix

$$\lambda I - C, \tag{1}$$

where I is the unit matrix.

Let the H.C.F.'s of the minors of orders 1, 2,..., n be $D_1(\lambda)$, $D_2(\lambda),..., D_n(\lambda)$. Then (Chap. III, § 6) the invariant factors of (1) are $E_1(\lambda)$, $E_2(\lambda),..., E_n(\lambda)$, where

$$E_1 \equiv D_1, \quad E_2 \equiv D_2/D_1, \quad ..., \quad E_n \equiv D_n/D_{n-1}. \tag{2}$$

We go on to determine these functions in terms of the Segre characteristic of C.

8.2. Take first the simple case when $C \equiv C_r(\alpha)$. Then

$$|\lambda I - C| = \begin{vmatrix} \lambda-\alpha & -1 & & & & & & \\ & \lambda-\alpha & & & & & & \\ . & . & . & . & . & . & . & . \\ & & & & & & \lambda-\alpha & -1 \\ & & & & & & & \lambda-\alpha \end{vmatrix} \tag{3}$$

and $D_r = (\lambda-\alpha)^r$. The minor obtained by deleting the first column and the last row of (3) has the value $(-1)^{r-1}$ and so the H.C.F. of minors of order $r-1$ is unity; in consequence, all minors of lower order have unity as their H.C.F. Hence

$$D_1 = D_2 = ... = D_{r-1} = 1, \qquad D_r = (\lambda-\alpha)^r,$$

and $$E_1 = E_2 = ... = E_{r-1} = 1, \qquad E_r = (\lambda-\alpha)^r.$$

Now consider

$$C = \text{diag}\{C_r(\alpha), C_s(\beta), C_t(\gamma),...\}.$$

The Segre characteristic of C is

$$[r\, s\, t\, ...],$$

† By minor we mean a determinant obtained by deleting an equal number of rows and columns.

of order n; the characteristic roots are $\alpha, \beta, \gamma,\ldots$. Hence

$$D_n = |\lambda I - C| = (\lambda-\alpha)^r(\lambda-\beta)^s(\lambda-\gamma)^t\ldots.$$

As we see from our discussion of (3), D_{n-1} cannot have $\lambda-\alpha$ as a factor; neither can it have $\lambda-\beta, \lambda-\gamma,\ldots$. Hence

$$D_1 = D_2 = \ldots = D_{n-1} = 1, \qquad D_n = (\lambda-\alpha)^r(\lambda-\beta)^s\ldots,$$

and

$$E_1 = E_2 = \ldots = E_{n-1} = 1, \qquad E_n = (\lambda-\alpha)^r(\lambda-\beta)^s\ldots.$$

The elementary divisors of $\lambda I - C$ (Chap. III, § 8) are the factors $(\lambda-\alpha)^r, (\lambda-\beta)^s,\ldots$ of E_n. We have proved then that

the elementary divisors of $\lambda I - [r\,s\,t\ldots]$ are

$$(\lambda-\alpha)^r, \quad (\lambda-\beta)^s, \quad (\lambda-\gamma)^t, \quad \ldots,$$

these being all associated with the one invariant factor E_n; and all other invariant factors $E_{n-1},\ldots, E_1$ are unity.

8.3. Next consider $C = [(x\,y)]$, wherein $x \leqslant y$, $x+y = r$ and the r-ple characteristic root is α. Then

$$|\lambda I - C| = \begin{vmatrix} C'_x(\lambda-\alpha) & O \\ O & C'_y(\lambda-\alpha) \end{vmatrix}, \tag{4}$$

where $C'_x(\lambda-\alpha)$, $C'_y(\lambda-\alpha)$ are of the same type as (3) and have x, y rows respectively. Here

$$D_r = |\lambda I - C| = (\lambda-\alpha)^{x+y}.$$

The minor got by deleting the first column and last row of $C'_y(\lambda-\alpha)$ will have no factor $\lambda-\alpha$ arising from the C'_y part [cf. § 8.2]; it will have the factor $(\lambda-\alpha)^x$ arising from the C'_x part. Moreover, every non-zero minor of order $n-1$ taken from $|\lambda I - C|$ must contain† either the whole of C'_x, when it has a factor $(\lambda-\alpha)^x$, or the whole of C'_y, when it has a factor $(\lambda-\alpha)^y$. Since $x \leqslant y$, the H.C.F. of $(n-1)$-rowed minors of (4) is $(\lambda-\alpha)^x$. Thus $D_{r-1} = (\lambda-\alpha)^x$ and

$$E_r = D_r/D_{r-1} = (\lambda-\alpha)^y.$$

Again, the minor of order $r-2$ obtained by deleting from (4) the first column and last row of both C'_x and C'_y is equal

† Theorem 15 shows that, if we are to have a non-zero minor, the deletion of a row from C'_y must be accompanied by the deletion of a column from C'_y.

to ± 1. Hence $D_{r-2} = 1$ and, consequently, $D_{r-3} = \ldots = D_1 = 1$. We have thus proved that

$$D_1 = \ldots = D_{r-2} = 1, \quad D_{r-1} = (\lambda-\alpha)^x, \quad D_r = (\lambda-\alpha)^{x+y},$$

and

$$E_1 = \ldots = E_{r-2} = 1, \quad E_{r-1} = (\lambda-\alpha)^x, \quad E_r = (\lambda-\alpha)^y.$$

Accordingly, the elementary divisors of $\lambda I - [(xy)]$ when $x \leqslant y$ and $x+y = r$ are

$$(\lambda-\alpha)^x, \qquad (\lambda-\alpha)^y,$$

the former being associated with the invariant factor E_{r-1}, the latter with E_r.

8.4. Now embed the $[(xy)]$ in a larger matrix, of order n, say

$$[(xy)\,s\,t\,\ldots],$$

the characteristic roots being α (r times; $x+y = r$), β (s times), and so on. Our preceding work shows that

$$D_n = (\lambda-\alpha)^{x+y}(\lambda-\beta)^s(\lambda-\gamma)^t\ldots,$$

$$D_{n-1} = (\lambda-\alpha)^x, \quad D_{n-2} = \ldots = D_1 = 1,$$

and that the invariant factors are, by (2),

$$E_1 = \ldots = E_{n-2} = 1, \qquad E_{n-1} = (\lambda-\alpha)^x,$$

$$E_n = (\lambda-\alpha)^y(\lambda-\beta)^s(\lambda-\gamma)^t\ldots.$$

Thus the elementary divisors of $\lambda I - [(xy)\,s\,t\,\ldots]$ are

$$(\lambda-\alpha)^x \quad \text{forming} \quad E_{n-1},$$

$$(\lambda-\alpha)^y, \quad (\lambda-\beta)^s, \quad (\lambda-\gamma)^t, \quad \ldots \quad \text{forming} \quad E_n.$$

8.5. In the light of Theorem 15 and its corollary, the extension to more complicated forms is immediate; for example, *the elementary divisors of*

$$\lambda I - [(xyz)\,(uv)\,p\,q\,\ldots]$$

with characteristic roots $\alpha, \beta, \gamma,\ldots$, *when*

$$x \leqslant y \leqslant z \quad \textit{and} \quad u \leqslant v,$$

are

$$(\lambda-\alpha)^x, \quad (\lambda-\alpha)^y, \quad (\lambda-\alpha)^z, \quad (\lambda-\beta)^u, \quad (\lambda-\beta)^v, \quad (\lambda-\gamma)^p, \quad \ldots;$$

moreover, the invariant factors $E_1, \ldots, E_{n-3}$ *are equal to unity while*

$$E_{n-2} = (\lambda-\alpha)^x,$$
$$E_{n-1} = (\lambda-\alpha)^y(\lambda-\beta)^u,$$
$$E_n = (\lambda-\alpha)^z(\lambda-\beta)^v(\lambda-\gamma)^p \ldots.$$

In all cases, the indices that occur in the elementary divisors of $\lambda I-C$ *are the numbers that occur in the Segre characteristic*† *of* C.

When one is a little practised in handling these canonical forms there is no need to state how the invariant factors are made up, once we are given the complete set of elementary divisors. A set of elementary divisors

$$(\lambda-\alpha)^r, \quad (\lambda-\beta)^s, \quad (\lambda-\gamma)^t, \quad \ldots$$

indicates of itself that all the divisors are associated with E_n and that all other invariant factors, $E_1, \ldots, E_{n-1}$ are unity. Again, the set of elementary divisors

$$(\lambda-\alpha)^x, \quad (\lambda-\alpha)^y, \quad (\lambda-\alpha)^z, \quad (\lambda-\beta)^u, \quad (\lambda-\beta)^v, \quad (\lambda-\gamma)^t, \quad \ldots,$$

where $x \leqslant y \leqslant z$ and $u \leqslant v$, indicates of itself that all invariant factors from E_1 to E_{n-3} are unity, that

$$E_{n-2} = (\lambda-\alpha)^x,$$
$$E_{n-1} = (\lambda-\alpha)^y(\lambda-\beta)^u,$$

and

$$E_n = (\lambda-\alpha)^z(\lambda-\beta)^v(\lambda-\gamma)^t \ldots.$$

On the other hand, some parts of the literature are hard to follow because the author is so familiar with the above facts that he never states them explicitly.

8.6. So far we have been concerned only with classical canonical forms but, in view of Theorem 11, the results can be carried over at once to any square matrix A.

THEOREM 16. *Let* A *be a square matrix and let the elementary divisors of* $\lambda I-A$ *be*

$$\left.\begin{array}{lllll} \text{(i)} & (\lambda-\alpha)^{x_1}, & (\lambda-\alpha)^{y_1}, & \ldots & x_1 \leqslant y_1 \leqslant \ldots, \\ & (\lambda-\beta)^{x_2}, & (\lambda-\beta)^{y_2}, & \ldots & x_2 \leqslant y_2 \leqslant \ldots, \\ & \cdot & \cdot & \cdot & \cdot \\ \text{(ii)} & (\lambda-\zeta)^{r}, & (\lambda-\eta)^{s}, & \ldots & \end{array}\right\} \qquad (5)$$

† In fact, some treatments of the matter define the Segre characteristic by the indices that occur in the elementary divisors.

Then the classical canonical form of A is

$$[(x_1 y_1 \ldots)(x_2 y_2 \ldots) \ldots r s \ldots]. \tag{6}$$

Conversely, when the classical canonical form of A is given by (6), *the elementary divisors of* $\lambda I - A$ *are given by* (5).

NOTE. There may be no factor $\lambda-\alpha$ with more than one elementary divisor, when the terms (i) will be absent from (5) and the terms in parentheses absent from (6): there may be no factor $\lambda-\zeta$ with only one elementary divisor, when the terms (ii) will be absent from (5) and the terms $r\ s\ldots$ absent from (6).

PROOF. Let C be the classical canonical form of A. Then C is a transform of A; that is, there is a matrix T for which $C = TAT^{-1}$. Hence, by Theorem 11, $\lambda I - C$ has the same elementary divisors as $\lambda I - A$.

Now let $\lambda I - A$ have the elementary divisors (5). Then $\lambda I - C$ has the same elementary divisors and, by § 8.5, C has the Segre characteristic (6).

Conversely, when C is given by (6), the elementary divisors of $\lambda I - C$ are, by § 8.5, given by (5). Hence, by Theorem 11, the elementary divisors of $\lambda I - A$ are also given by (5).

COROLLARY. *A necessary and sufficient condition that a matrix A have a transform which is purely diagonal is that all the elementary divisors of* $\lambda I - A$ *be linear.*

PROOF. Let the divisors be all linear, say

$$\lambda-\alpha, \quad \lambda-\beta, \quad \lambda-\gamma, \quad \ldots, \tag{7}$$

where $\alpha, \beta, \gamma,\ldots$ are not necessarily all different. The classical form C must then be composed entirely of C_1's: for if it contained $C_k(\lambda_r)$, where $k > 1$, there would be an elementary divisor $(\lambda-\lambda_r)^k$ with $k > 1$.

Conversely, when the classical form C is purely diagonal, it is

$$\operatorname{diag}\{C_1(\alpha), C_1(\beta),\ldots\},$$

where $\alpha, \beta,\ldots$ are not necessarily all different. The elementary divisors are then given, from § 8.5, by

$$\lambda-\alpha, \quad \lambda-\beta, \quad \lambda-\gamma, \quad \ldots.$$

EXAMPLE. Take the simple matrix $C = \mathrm{diag}(\alpha,\alpha,\alpha)$. The H.C.F. of minors of $\lambda I-C$ of orders one, two, three is $\lambda-\alpha$, $(\lambda-\alpha)^2$, $(\lambda-\alpha)^3$ respectively. In the notation of § 8.1,

$$D_1 = \lambda-\alpha, \quad D_2 = (\lambda-\alpha)^2, \quad D_3 = (\lambda-\alpha)^3,$$

and each of the invariant factors

$$E_1 = D_1, \quad E_2 = D_2/D_1, \quad E_3 = D_3/D_2$$

is equal to $\lambda-\alpha$.

9. A short proof of Theorem 14

Let A be a given square matrix and let the elementary divisors of $\lambda I-A$ be

$$\begin{array}{lllll} \text{(i)} & (\lambda-\alpha)^{x_1}, & (\lambda-\alpha)^{y_1}, & \ldots & (x_1 \leqslant y_1 \leqslant \ldots), \\ & (\lambda-\beta)^{x_2}, & (\lambda-\beta)^{y_2}, & \ldots & (x_2 \leqslant y_2 \leqslant \ldots), \\ & \cdot & \cdot & \cdot & \cdot \\ \text{(ii)} & (\lambda-\zeta)^{r}, & (\lambda-\eta)^{s}, & \ldots, & \end{array}$$

it being understood that divisors of type (i) or of type (ii) may not occur for a particular A.

Write down the matrix

$$C \equiv \mathrm{diag}\{C_{x_1}(\alpha), C_{y_1}(\alpha), \ldots, C_{x_2}(\beta), \ldots, C_r(\zeta), \ldots\}.$$

By the calculations of § 8, the elementary divisors of $\lambda I-C$ are precisely those of $\lambda I-A$. Hence, by Theorem 11, C is a transform of A and

every square matrix A has a transform of the type C.

10. The rational canonical form of a matrix

10.1. *Preliminary lemma*

LEMMA. *When*

$$f(\lambda) \equiv \lambda^r + p_1\lambda^{r-1} + \ldots + p_r$$

and
$$P = \begin{bmatrix} 0 & 1 & 0 & . & . & 0 \\ 0 & 0 & 1 & . & . & 0 \\ . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & 1 \\ -p_r & -p_{r-1} & -p_{r-2} & . & . & -p_1 \end{bmatrix},$$

the value of the determinant of the matrix $\lambda I-P$ is given by

$$|\lambda I - P| = f(\lambda); \tag{1}$$

moreover, the H.C.F. of minors of $\lambda I-P$ having $r-1$ rows or less is unity.

PROOF.

$$|\lambda I-P| = \begin{vmatrix} \lambda & -1 & 0 & . & . & 0 \\ 0 & \lambda & -1 & . & . & 0 \\ . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & -1 \\ p_r & p_{r-1} & p_{r-2} & . & . & \lambda+p_1 \end{vmatrix}. \qquad (2)$$

Now neither the value of (2) nor the H.C.F.'s of minors of various orders is changed when its first column is replaced by

$$\text{col. } 1+\lambda(\text{col. } 2)+\lambda^2(\text{col. } 3)+\dots.$$

When this is done the new first column consists of $r-1$ zeros followed by $f(\lambda)$. It follows that

$$|\lambda I-P| = (-1)^{r-1}.f(\lambda).(-1)^{r-1} = f(\lambda). \qquad (3)$$

Moreover, the minor got by deleting the first column and last row is equal to $(-1)^{r-1}$, and so the H.C.F. of minors having $r-1$ rows is unity.

DEFINITION 14†. *We say that P is the* MATRIX ASSOCIATED WITH THE POLYNOMIAL $f(\lambda)$.

10.2. Let A be a square matrix of order n with elements in a field F. Let the invariant factors of $\lambda I-A$ be

$$E_1(\lambda), \quad \dots, \quad E_s(\lambda), \quad \dots, \quad E_n(\lambda)$$

and let $E_s(\lambda)$ be the first of these to differ from unity. Let E_i be the matrix associated with $E_i(\lambda)$ for $i = s,\dots, n$ and let

$$E = \text{diag}\{E_s, E_{s+1},\dots, E_n\}. \qquad (4)$$

We shall show that E is a transform of A by showing that $\lambda I-E$ and $\lambda I-A$ have the same invariant factors (Theorem 11).

The first point is to show that E is of order n. By Theorem 12,

$$|\lambda I-A| = E_s(\lambda)\dots E_n(\lambda). \qquad (5)$$

The product on the right-hand side must be of degree n and

† We introduce the definition solely for convenience of diction in the next few sub-sections.

so the sum of the number of rows in $E_s, E_{s+1}, \ldots, E_n$ of (4) is also n, for the number of rows in E_i is the degree of $E_i(\lambda)$.

From the definition of E in (4),

$$\lambda I - E = \operatorname{diag}\{\lambda I - E_s, \ldots, \lambda I - E_n\}, \tag{6}$$

it being understood that the I's are unit matrices of appropriate orders, not necessarily all the same. Again, by (3),

$$|\lambda I - E_i| = E_i(\lambda) \tag{7}$$

and the product of the determinants

$$|\lambda I - E_s| \ldots |\lambda I - E_n| \quad \text{is} \quad E_s(\lambda) \ldots E_n(\lambda). \tag{8}$$

Consider (6) with the submatrices set out in full; let the result of deleting one row and one column of elements be a minor K, say. If a non-zero minor is to result, the row and column must come from one and the same submatrix (Theorem 15). If they come from $\lambda I - E_n$, the value of K is

$$E_s(\lambda) \ldots E_{n-1}(\lambda) M_n, \tag{9}$$

where M_n is a minor of $\lambda I - E_n$. By § 10.1, the H.C.F. of the M_n is 1 and so the H.C.F. of (9) is

$$E_s(\lambda) \ldots E_{n-1}(\lambda).$$

That is, when we delete a row and column from $\lambda I - E_n$ the H.C.F. of resulting minors of (6) loses the factor $E_n(\lambda)$ from the product (8). When we delete a row and column from $\lambda I - E_i$, the H.C.F. of resulting minors of (6) loses the factor $E_i(\lambda)$ from the product (8). But $E_i(\lambda)$ is a factor† of $E_n(\lambda)$ and no $E_i(\lambda)$ loss can be heavier than $E_n(\lambda)$. Hence the H.C.F. of minors got by deleting one row and one column from $\lambda I - E$ is

$$E_s(\lambda) \ldots E_{n-1}(\lambda). \tag{10}$$

On deleting two rows and two columns from $\lambda I - E$ we get the heaviest loss of factors from (8) when we lose $E_n(\lambda)$ by a deletion from $\lambda I - E_n$ and lose $E_{n-1}(\lambda)$ by a deletion from $\lambda I - E_{n-1}$. The H.C.F. of $(n-2)$-rowed minors of (6) is

$$E_s(\lambda) \ldots E_{n-2}(\lambda);$$

and so on for minors of $n-3, \ldots, s$ rows. By taking one row

† This is a basic property of the invariant factors: cf. Theorem 4.

and one column from each of the submatrices in (6) we can obtain a minor of $\lambda I - E$ whose value is ± 1: the H.C.F. of minors of $\lambda I - E$ with less than s rows is therefore unity.

Thus the H.C.F.'s of minors of various orders of (6) are

$$1,\quad 1,\quad \ldots,\quad 1,\quad E_s(\lambda),\quad E_s(\lambda)E_{s+1}(\lambda),\quad \ldots,$$
$$E_s(\lambda)E_{s+1}(\lambda)\ldots E_n(\lambda),$$

and so the invariant factors of (6) are

$$1,\quad 1,\quad \ldots,\quad 1,\quad E_s(\lambda),\quad E_{s+1}(\lambda),\quad \ldots,\quad E_n(\lambda).$$

Hence $\lambda I - E$ has the same invariant factors as $\lambda I - A$ and so, by Theorem 11, E is a transform of A. Moreover, the coefficients of the $E_i(\lambda)$ lie in the field F, so that the elements of E also lie in F. To sum up:

THEOREM 17. *When the invariant factors of $\lambda I - A$, other than unity, are*

$$E_s(\lambda),\quad E_{s+1}(\lambda),\quad \ldots,\quad E_n(\lambda),$$

the matrix $$E = \text{diag}\{M_s, M_{s+1},\ldots, M_n\},$$

where M_i is the matrix associated with the polynomial $E_i(\lambda)$, is c-equivalent to A.†

Moreover, when the elements of A lie in a given field F the elements of E also lie in F.

This matrix E is sometimes called the RATIONAL CANONICAL FORM of A: the term rational is used because the calculations are (theoretically at least) possible within the confines of a given field. The invariant factors can be determined by H.C.F. processes and without finding the roots of the equation $|\lambda I - A| = 0$.

11. The minimum function of a matrix‡

11.1. We noted in Chapter I, § 11, that when A is a square matrix and

$$|\lambda I - A| = f(\lambda) = \lambda^n + p_1\lambda^{n-1} + \ldots + p_n, \tag{1}$$

then also $$f(A) = A^n + p_1 A^{n-1} + \ldots + p_n I = 0. \tag{2}$$

For any integer k there either is or is not a polynomial

$$g(\lambda) = \lambda^k + g_1\lambda^{k-1} + \ldots + g_k$$

† We have changed the notation for the associated matrix from E_s to M_s.

‡ For an alternative treatment, see Chapter VII, § 1.

for which $g(A) = 0$. Let m be the least value of k for which there is such a polynomial: the equation (2) shows that this is an effective definition of m and, in fact, that $m \leqslant n$. Moreover *there is only one polynomial*

$$h(\lambda) = \lambda^m + h_1\lambda^{m-1} + ... + h_m,$$

of degree m, for which $h(A) = 0$. For, if there were two such, say $h_1(\lambda)$ and $h_2(\lambda)$, the polynomial $h_1(\lambda) - h_2(\lambda)$ would be of degree $m-1$ at most and $h_1(A) - h_2(A)$ would be zero. These facts justify the following definition

DEFINITION 15. *The polynomial of lowest degree*

$$h(\lambda) = \lambda^m + h_1\lambda^{m-1} + ... + h_m$$

for which $h(A) = 0$ *is called the* MINIMUM FUNCTION (*or the* REDUCED CHARACTERISTIC FUNCTION) *of the matrix A.*

11.2. *An algebraical detail*

Before going on we dispose of a detail that occurs in later proofs.

Let

$$g(\lambda) \equiv g_0 + g_1\lambda + ..., \qquad f(\lambda) \equiv f_0 + f_1\lambda + ...,$$
$$q(\lambda) \equiv q_0 + q_1\lambda + ..., \qquad r(\lambda) \equiv r_0 + r_1\lambda + ...,$$

be polynomials, one or more of which may be constants, that satisfy an identity

$$g(\lambda) \equiv f(\lambda)q(\lambda) + r(\lambda). \tag{3}$$

Then, on equating coefficients in (3),

$$g_0 = f_0 q_0 + r_0, \qquad g_1 = f_0 q_1 + f_1 q_0 + r_1,$$

etc. Hence, when A is any square matrix,

$$g(A) = f(A)q(A) + r(A). \tag{4}$$

11.3. THEOREM 18. *Let* $h(\lambda)$ *be the minimum function of a square matrix A and let* $g(\lambda)$ *be a polynomial in* λ. *Then a necessary and sufficient condition for* $g(A)$ *to be zero is that* $g(\lambda)$ *should contain* $h(\lambda)$ *as a factor.*

PROOF. (i) Let $\qquad g(\lambda) = h(\lambda)q(\lambda),$

where $q(\lambda)$ is a polynomial, possibly a constant (not zero). Then

$$g(A) = h(A)q(A).$$

But, by definition, $h(A)$ is the null matrix and therefore $g(A)$ is the null matrix.

(ii) Let $$g(A) = 0.$$

By the definition of $h(\lambda)$, the degree of $g(\lambda)$ is not less than the degree of $h(\lambda)$ and

EITHER $g(\lambda)$ contains $h(\lambda)$ as a factor, and the theorem is thereby proved,

OR there are polynomials $q(\lambda)$ and $r(\lambda)$ for which
$$g(\lambda) = h(\lambda)q(\lambda)+r(\lambda)$$
and the degree of $r(\lambda)$ is less than that of $h(\lambda)$.

In the latter event,
$$g(A) = h(A)q(A)+r(A)$$
and, since $g(A) = h(A) = 0$, $r(A) = 0$. But this gives a contradiction, since $r(A) \neq 0$, the polynomial $r(\lambda)$ being of lower degree than $h(\lambda)$. Hence the second alternative cannot arise.

11.4. *The matrix associated with an invariant factor*

Consider the matrix
$$P = \begin{bmatrix} 0 & 1 & 0 & . & . & 0 \\ 0 & 0 & 1 & . & . & 0 \\ . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & 1 \\ -p_r & -p_{r-1} & -p_{r-2} & . & . & -p_1 \end{bmatrix}$$
associated with the function
$$p(\lambda) = \lambda^r+p_1\lambda^{r-1}+...+p_r.$$
As we saw in § 10.1,
$$|\lambda I_r-P| = p(\lambda),$$
and, since any matrix satisfies its own characteristic equation,
$$p(P) = 0. \tag{5}$$

But, as a little calculation shows, the first row of each of the matrices
$$I_r, \quad P, \quad P^2, \quad ..., \quad P^{r-1}$$
consists of $n-1$ zeros and one unity, the element 1 appearing in the first place in I_r, in the second place in P,..., in the last place in P^{r-1}. Hence the matrix
$$a_0 I_r+a_1 P+a_2 P^2+...+a_{r-1} P^{r-1} \tag{6}$$

has for its first row

$$a_0 \quad a_1 \quad a_2 \quad \dots \quad a_{r-1}.$$

Accordingly (6) can be the null matrix only if all the a_i are zero. We have thus proved that the matrix P cannot satisfy an equation of degree less than r, that is, $p(\lambda)$ is the minimum function of P.

Now let A be a square matrix of order n, *let* $E_i(\lambda)$ *be an invariant factor of* $\lambda I - A$, *and let* M_i *be the matrix associated with* $E_i(\lambda)$. *Then*, by what we have just proved for the matrix P, $E_i(\lambda)$ *is the minimum function of* M_i and, by Theorem 18, a necessary and sufficient condition that $f(M_i)$ be the null matrix is that $f(\lambda)$ contain $E_i(\lambda)$ as a factor.

11.5. *A general theorem*

THEOREM 19. *Let A be a square matrix of order n and let the invariant factors of $\lambda I - A$ be $E_1(\lambda)$, $E_2(\lambda)$,..., $E_n(\lambda)$. Then the minimum function of the matrix A is $E_n(\lambda)$ and $g(A) = 0$ if and only if $g(\lambda)$ contains $E_n(\lambda)$ as a factor.*

PROOF. By Theorem 17, there is a non-singular matrix T for which

$$TAT^{-1} = E = \operatorname{diag}\{M_s, M_{s+1},..., M_n\}.$$

Moreover, if $g(\lambda)$ be any polynomial in λ,

$$g(E) = \operatorname{diag}\{g(M_s), g(M_{s+1}),..., g(M_n)\}$$

and

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

Thus $g(A)$ is the null matrix if and only if each of the submatrices $g(M_s)$,..., $g(M_n)$ is a null matrix. By § 11.4, this is so if and only if $g(\lambda)$ contains each of $E_s(\lambda)$,..., $E_n(\lambda)$ as a factor. Accordingly, since $E_s(\lambda)$,..., $E_{n-1}(\lambda)$ are factors of $E_n(\lambda)$, a necessary and sufficient condition for $g(A)$ to be the null matrix is that $g(\lambda)$ contain $E_n(\lambda)$ as a factor, which proves the theorem.

CHAPTER V

INFINITE SERIES AND FUNCTIONS OF MATRICES

In this chapter all numbers belong to the field of complex numbers. When occasion warrants, we refer to any such number as a SCALAR; we refer to

$$a_0+a_1x+a_2x^2+\ldots, \tag{1}$$

where the constants a_r and the variable x are complex numbers (including real numbers as a particular case), as a SCALAR POWER SERIES.

We shall be concerned with series of the type

$$a_0I+a_1A+a_2A^2+\ldots, \tag{2}$$

wherein A is a square matrix whose elements are scalars. We refer to such a series as a MATRIX POWER SERIES.

1. A convergent sequence of matrices

The natural way of attaching a meaning to the notation (2) is to follow the procedure of elementary convergence theory. Let

$$B_1,\quad B_2,\quad \ldots,\quad B_N,\quad \ldots \tag{3}$$

be a given sequence of matrices, each B_k having m rows and n columns. Let

$$b_{rs/N}$$

be the element in the rth row and sth column of B_N. When there is a matrix B, with elements b_{rs}, such that, for every r and s,

$$b_{rs/N}\to b_{rs}\quad \text{as}\quad N\to\infty,$$

we write

$$B_N\to B\quad \text{as}\quad N\to\infty.$$

We then say that B_N converges to B as N tends to infinity and refer to (3) as a convergent sequence.

In fact, without elaboration of detail, we carry over to matrices the terminology of elementary convergence theory. For example, when $A_1, A_2,\ldots$ are square matrices of the same order,

$$S_n=A_1+A_2+\ldots+A_n,$$

and $S_n\to S$ as $n\to\infty$, we write

$$S=\sum_{n=1}^{\infty}A_n.$$

Or again, just as we write

$$S(x) = \sum_{n=1}^{\infty} u_n(x)$$

when the terms of the series are functions of a scalar variable x, so we shall write

$$S(A) = \sum_{n=1}^{\infty} u_n(A)$$

when the terms of the series are functions of a matrix variable A. In particular, when the scalar power series

$$a_0 + a_1 x + a_2 x^2 + \dots$$

is the expansion of some function $f(x)$ in the neighbourhood of the origin, we use $f(A)$ to denote the sum of the matrix power series

$$a_0 + a_1 A + a_2 A^2 + \dots$$

We go on to consider the convergence of such series. The next two sections (§§ 2 and 3) are by way of preliminary to our main problem; some of the details are of interest in their own right.

2. The auxiliary unit matrix

2.1. Let U be a matrix having unity in each place of the superdiagonal and zero in all other places: for example, when dealing with matrices of order three,

$$U = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

In the sequel we suppose U to have r rows and columns.

Suppose A is the matrix

$$\begin{bmatrix} a_{11} & a_{12} & . & . & a_{1r} \\ a_{21} & a_{22} & . & . & a_{2r} \\ . & . & . & . & . \\ a_{r1} & a_{r2} & . & . & a_{rr} \end{bmatrix}.$$

Then

$$UA = \begin{bmatrix} a_{21} & a_{22} & . & . & a_{2r} \\ . & . & . & . & . \\ a_{r1} & a_{r2} & . & . & a_{rr} \\ 0 & 0 & . & . & 0 \end{bmatrix},$$

which we get from A by removing its first row, moving the other rows up one, and filling in the last row by zeros. In particular,

$$U^2 = \begin{bmatrix} 0 & 0 & 1 & . & . & 0 \\ & . & . & . & . & \\ 0 & 0 & 0 & . & . & 1 \\ 0 & 0 & 0 & . & . & 0 \\ 0 & 0 & 0 & . & . & 0 \end{bmatrix},$$

has its 1's two places to the right of the diagonal, U^3 its 1's three places to the right of the diagonal, and so on. The sequence effectively ends with U^{r-1}, which has a single 1 in the top right position; U^r has zeros everywhere.

Thus the ranks of U, U^2,..., U^{r-1} are $r-1$, $r-2$,..., 1; and U^r is the null matrix. Further,

$$U^{r+p} = 0 \quad (p \geqslant 0).$$

2.2. *Classical submatrices*

We recall the definition of $C_r(\lambda)$ from Chapter IV, § 5, and note that

$$C_r(\lambda) = \lambda I + U.$$

By what we have just shown in § 2.1, the matrix

$$\{C_r(\lambda) - \lambda I\}^k = U^k$$

is of rank $r-k$ when $1 \leqslant k < r$ and is the null matrix when $k \geqslant r$.

3. Collineation and convergence

3.1. THEOREM 20. *Let $f_n(x)$ be a given sequence of scalar polynomials and let the matrices A and C be transforms, with $C = TAT^{-1}$. Then the convergence of the sequence of matrices $f_n(A)$ to a limit matrix $f(A)$ implies the convergence of the sequence of matrices $f_n(C)$ to a limit matrix $f(C)$ given by*

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

PROOF. Denote a matrix by the element in its ith row and jth column and let

$$T = [t_{ij}], \quad f_n(A) = [a_{ij}^{(n)}], \quad T^{-1} = [\theta_{ij}].$$

Then, since
$$C = TAT^{-1}$$
and $f_n(x)$ is a polynomial (cf. Chap. IV, § 3),

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

Hence $$f_n(C) = [t_{il}\, a_{lk}^{(n)}\, \theta_{kj}], \tag{1}$$
where the repeated suffixes l, k imply the use of the summation convention.

Now suppose that the sequence $f_n(A)$ converges to a limit matrix $f(A)$; that is to say, for every l and k there is an a_{lk} such that
$$a_{lk}^{(n)} \to a_{lk} \tag{2}$$
and $f(A)$ denotes the matrix $[a_{ij}]$. Then, by (2),
$$t_{il}\, a_{lk}^{(n)}\, \theta_{kj} \to t_{il}\, a_{lk}\, \theta_{kj}$$
and, from (1), $$f_n(C) \to Tf(A)T^{-1}. \tag{3}$$
This proves the theorem.

COROLLARY 1. *If $\phi_n(x)$ is itself a limit of polynomials in x and $\phi_n(A) \to \phi(A)$, then $\phi_n(C) \to$ a limit $\phi(C)$ given by*
$$\phi(C) = T\phi(A)T^{-1}.$$

COROLLARY 2. *A necessary and sufficient condition for the convergence of a sequence of polynomial forms*
$$f_1(A),\quad f_2(A),\quad \ldots,\quad f_n(A),\quad \ldots$$
is that, for some transform C of A, the sequence
$$f_1(C),\quad f_2(C),\quad \ldots,\quad f_n(C),\quad \ldots$$
should converge.

For if C is a transform of A, A is a transform of C; in symbols, $C = TAT^{-1}$ implies $A = T^{-1}C(T^{-1})^{-1}$.

3.2. In the applications of this theorem, the particular transform of A considered is usually the classical canonical form of A. As we shall see, this enables us to reduce the problem of convergence of sequences of matrices to that of the convergence of sequences of scalars.

4. Diagonal matrices

Before considering our main theorem, we note a property of diagonal matrices on which our proof of that theorem will depend.

Let $$B = \mathrm{diag}\{B_1, B_2,\ldots, B_m\},$$

where each B_r is a square matrix. Then,† for any polynomial form $f_N(x)$,

$$f_N(B) = \operatorname{diag}\{f_N(B_1), f_N(B_2),..., f_N(B_m)\}.$$

The sequence $f_N(B)$ will converge to a limit $f(B)$ if, and only if, for $r = 1, 2,..., m$,

$$f_N(B_r) \to \text{a limit } f(B_r).$$

5. The convergence of a matrix power series

5.1. THEOREM 21. *Let the classical canonical form of a matrix A be*

$$C = \operatorname{diag}\{C_p(\lambda_1), C_q(\lambda_2),...\}. \tag{1}$$

Then it is sufficient for the convergence of the matrix power series

$$a_0 I + a_1 C + a_2 C^2 + ..., \tag{2}$$

and so (by Theorem 20) of the matrix power series

$$a_0 I + a_1 A + a_2 A^2 + ..., \tag{3}$$

that all the latent roots λ of the matrix A lie within the circle of convergence of the scalar power series

$$a_0 + a_1 x + a_2 x^2 + \dots. \tag{4}$$

PROOF. Let

$$f_N(C) = a_0 I + a_1 C + ... + a_{N-1} C^{N-1}.$$

Then, by (1) and § 4,

$$f_N(C) = \operatorname{diag}[f_N\{C_p(\lambda_1)\}, f_N\{C_q(\lambda_2)\},...].$$

Accordingly, we consider the sequence $f_N\{C_k(\lambda)\}$ where $C_k(\lambda)$ is a typical submatrix in (1). By § 2.2,

$$f_N\{C_k(\lambda)\} = f_N\{\lambda I + U\}, \tag{5}$$

where I and U are the unit and auxiliary unit matrices of order k.

Now, $f_N(x)$ being a polynomial in x of degree $N-1$, Taylor's expansion

$$f_N(\lambda + x) = f_N(\lambda) + x f'_N(\lambda) + ... + \frac{x^{N-1}}{(N-1)!} f_N^{(N-1)}(\lambda)$$

is an identity in x. This establishes the matrix identity

$$f_N(\lambda I + U) = I f_N(\lambda) + U f'_N(\lambda) + ... + \frac{U^{N-1}}{(N-1)!} f_N^{(N-1)}(\lambda).$$

† Cf. Chap. I, § 12 (b), equation (13).

But U^k, U^{k+1},... are all null matrices and so, when $N \geqslant k$,

$$f_N\{C_k(\lambda)\} = If_N(\lambda)+Uf'_N(\lambda)+\ldots+\frac{U^{k-1}}{(k-1)!}f_N^{(k-1)}(\lambda). \tag{6}$$

Now let λ lie within the circle of convergence of

$$f(z) = \sum_{n=0}^{\infty} a_n z^n.$$

Then, as $N \to \infty$, $f_N(\lambda)$, $f'_N(\lambda)$,..., $f_N^{(k-1)}(\lambda)$ tend to the limits $f(\lambda)$, $f'(\lambda)$,..., $f^{(k-1)}(\lambda)$. Hence

$$f_N\{C_k(\lambda)\} \to If(\lambda)+Uf'(\lambda)+\ldots+\frac{U^{k-1}}{(k-1)!}f^{(k-1)}(\lambda) = F_k(\lambda),$$

say. Remembering the results of § 2.1, we see that

$$F_k(\lambda) = \begin{bmatrix} f(\lambda) & f'(\lambda) & f''(\lambda)/2! & . & . & f^{(k-1)}(\lambda)/(k-1)! \\ 0 & f(\lambda) & f'(\lambda) & . & . & f^{(k-2)}(\lambda)/(k-2)! \\ . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & f(\lambda) \end{bmatrix}. \tag{7}$$

Accordingly, when all the latent roots of C lie within the circle of convergence of $f(z)$,

$$\begin{aligned} f(C) &= \lim f_N(C) \\ &= \lim \operatorname{diag}[f_N\{C_p(\lambda_1)\}, f_N\{C_q(\lambda_2)\}\ldots] \\ &= \operatorname{diag}\{F_p(\lambda_1), F_q(\lambda_2),\ldots\}. \end{aligned} \tag{8}$$

5.2. *A corollary.* By using known results in the theory of scalar power series we can refine the result enunciated in Theorem 21. A necessary and sufficient condition for the convergence of the series

$$a_0+a_1x+a_2x^2+\ldots$$

and of the series obtained by differentiating its terms once, twice,..., $k-1$ times is that the last of these series should converge.

Accordingly, *a necessary and sufficient condition for the series*

$$a_0+a_1C+a_2C^2+\ldots$$

to converge to the sum (8), *is that each of the series*

$$f^{(p-1)}(\lambda_1), \quad f^{(q-1)}(\lambda_2), \quad \ldots$$

should converge;† for this condition ensures that, in (6), $f_N(\lambda)$, $f'_N(\lambda)$,... will converge to $f(\lambda)$, $f'(\lambda)$,... as $N \to \infty$, so that the condition is sufficient for the convergence of $\sum a_n C^n$; while if we know that $\sum a_n C^n$ is convergent, each $f_N\{C_k(\lambda)\}$ must converge and so, again from (6), the series $f(\lambda)$, $f'(\lambda)$,..., $f^{(k-1)}(\lambda)$ must converge.

5.3. *An extension of theorem* 21

When
$$f(z) = \sum_{n=0}^{\infty} a_n(z-\alpha)^n$$
and
$$f(C) = \sum_{n=0}^{\infty} a_n(C-\alpha I)^n,$$
we still have, as in § 5.1,
$$f_N(\lambda I+U) = If_N(\lambda)+Uf'_N(\lambda)+...+\frac{U^{k-1}}{(k-1)!}f_N^{(k-1)}(\lambda)$$
and $f_N\{C_k(\lambda)\} \to F_k(\lambda)$, where $F_k(\lambda)$ is defined as in § 5.1. In fact, the argument of §§ 5.1, 5.2 is unaltered save in minor details when $x-\alpha$ replaces x: the conclusions are the same for power series in $x-\alpha$ as they are for power series in x.

6. Matrix power series expressed as polynomials

6.1. Let A be a given square matrix and let the equation of lowest degree satisfied by A be
$$H(A) = A^m+h_1 A^{m-1}+...+h_m I = 0. \tag{1}$$
Then we may use (1) to express A^{m+s} in the form
$$A^{m+s} = \alpha_{1s} A^{m-1}+\alpha_{2s} A^{m-2}+...+\alpha_{ms} I \quad (s \geqslant 0),$$
each α_{rs} being a function of the h's. If now we substitute these expressions for A^m, A^{m+1},... in a convergent power series
$$a_0 I+a_1 A+a_2 A^2+... \tag{2}$$
and collect like powers of A, we replace (2) by
$$g_0 I+g_1 A+...+g_{m-1} A^{m-1}, \tag{3}$$
where
$$g_r = a_r+\sum_{s=0}^{\infty} a_{m+s}\,\alpha_{m-r,s}.$$

† $f^{(p-1)}(\lambda)$ denotes the series obtained by differentiating $a_0+a_1\lambda+a_2\lambda^2+...$ term-by-term $p-1$ times.

Such a process does not, of itself, prove that the series for g_r converge and the calculation of the α_{rs} is too complicated to be profitable. The process does, however, indicate the possibility of expressing the infinite series (2) as a finite power series (3). We make a fresh start on this problem in § 6.2.

6.2. Let the characteristic equation of A, namely

$$|\lambda I - A| = 0,$$

have k distinct roots $\alpha, \beta, ..., \kappa$, possibly repeated. Then these are also the roots of $H(\lambda) = 0$, where

$$H(\lambda) \equiv \lambda^m + h_1 \lambda^{m-1} + ... + h_m \tag{4}$$

is the minimum function† of A; some of the roots are, necessarily, repeated roots when $k < m$.

The fact that $\alpha, \beta, ..., \kappa$ are the roots of $H(\lambda) = 0$ enables us to prove the result indicated in § 6.1.

THEOREM 22. *A convergent matrix power series*

$$f(A) = a_0 I + a_1 A + a_2 A^2 + ...$$

can be expressed as a polynomial in A.

When the minimum function $H(\lambda)$ is of degree m and has m distinct zeros $\alpha, \beta, ..., \kappa$, this polynomial is

$$\sum_{\alpha,...,\kappa} f(\alpha) \frac{(A-\beta I)...(A-\kappa I)}{(\alpha-\beta)...(\alpha-\kappa)}. \tag{5}$$

When $H(\lambda)$ has repeated zeros, this polynomial has a form derived from (5) by a limiting process.

PROOF. Suppose first that the equation

$$H(\lambda) \equiv \lambda^m + h_1 \lambda^{m-1} + ... + h_m = 0$$

has m distinct roots $\alpha, \beta, ..., \kappa$.

Let $f_N(x)$ denote the sum of the first N terms of the scalar power series

$$a_0 + a_1 x + a_2 x^2 + ...$$

Then, $f_N(x)$ being a polynomial, there is a unique polynomial $r(x)$, of degree $m-1$ or less, and a polynomial $q(x)$ such that

$$f_N(x) \equiv q(x)H(x) + r(x). \tag{6}$$

† Cf. Chap. IV, § 11. It is sufficient here to note that in many examples $H(\lambda) \equiv |\lambda I - A|$ and that in all examples $H(\lambda) \equiv E_n(\lambda)$, the last invariant factor of $\lambda I - A$ (see p. 86), which contains $(\lambda-\alpha)...(\lambda-\kappa)$ as a factor.

Now $H(x)$ is zero when $x = \alpha,\ldots, \kappa$ and so

$$r(\alpha) = f_N(\alpha), \quad \ldots, \quad r(\kappa) = f_N(\kappa). \tag{7}$$

When we write the polynomial $r(x)$ in the form†

$$r(x) \equiv r_0 + r_1 x + \ldots + r_{m-1} x^{m-1},$$

the equations (7) provide m linear equations in $r_0, r_1,\ldots, r_{m-1}$ of the type

$$r_0 + r_1 \alpha + \ldots + r_{m-1} \alpha^{m-1} = f_N(\alpha). \tag{8}$$

Further, the scalar identity (6) establishes the matrix identity

$$f_N(A) = q(A)H(A) + r(A),$$

and so, since $H(A) = 0$,

$$r_0 + r_1 A + \ldots + r_{m-1} A^{m-1} = f_N(A). \tag{9}$$

On eliminating‡ $r_0,\ldots, r_{m-1}$ from (9) and the m equations (8),

$$\begin{vmatrix} 1 & \alpha & . & . & \alpha^{m-1} & f_N(\alpha) \\ 1 & \beta & . & . & \beta^{m-1} & f_N(\beta) \\ . & . & . & . & . & . \\ 1 & \kappa & . & . & \kappa^{m-1} & f_N(\kappa) \\ 1 & A & . & . & A^{m-1} & f_N(A) \end{vmatrix} = 0. \tag{10}$$

We know (from § 5) that $f_N(\alpha),\ldots, f_N(\kappa) \to f(\alpha),\ldots, f(\kappa)$, provided that $f_N(A) \to$ a finite $f(A)$ as $N \to \infty$. On making N tend to infinity in (10), we obtain

$$\begin{vmatrix} 1 & \alpha & . & . & \alpha^{m-1} & f(\alpha) \\ 1 & \beta & . & . & \beta^{m-1} & f(\beta) \\ . & . & . & . & . & . \\ 1 & \kappa & . & . & \kappa^{m-1} & f(\kappa) \\ 1 & A & . & . & A^{m-1} & f(A) \end{vmatrix} = 0, \tag{11}$$

provided only that the series

$$f(A) \equiv a_0 + a_1 A + a_2 A^2 + \ldots$$

is convergent.

The form (5) given in the enunciation of our theorem follows when we expand the determinant in (11) by its last column.

† The coefficients depend on N, but we prefer to avoid the double suffix notation that would be required to give an explicit expression to this dependence.

‡ The matrix equation (9) is a conspectus of n^2 scalar equations equating individual elements of $r(A)$ to the corresponding individual elements of $f_N(A)$; the 'eliminant' (10) is a conspectus of the n^2 'eliminants' obtained from combining each of the original n^2 scalar equations with the equations (8).

In detail, on using $\zeta(abc\ldots jk)$ to denote the produ
differences

$$(a-b)\ldots(a-k)(b-c)\ldots(b-k)\ldots(j-k)$$

and, when A is a matrix, $\zeta(Abc\ldots jk)$ to denote

$$(A-bI)\ldots(A-kI)(b-c)\ldots(b-k)\ldots(j-k),$$

the expansion is

$$f(\alpha)\zeta(A\kappa\ldots\gamma\beta)-f(\beta)\zeta(A\kappa\ldots\gamma\alpha)+\ldots+(-1)^m f(A)\zeta(\kappa\ldots\beta$$

Thus the equation (11) gives

$$f(A)=(-1)^{m-1}f(\alpha)\frac{\zeta(A\kappa\ldots\gamma\beta)}{\zeta(\kappa\ldots\gamma\beta\alpha)}+(-1)^m f(\beta)\frac{\zeta(A\kappa\ldots\gamma\alpha)}{\zeta(\kappa\ldots\gamma\beta\alpha)}$$

$$=f(\alpha)\frac{\zeta(A\kappa\ldots\gamma\beta)}{\zeta(\alpha\kappa\ldots\gamma\beta)}+f(\beta)\frac{\zeta(A\kappa\ldots\gamma\alpha)}{\zeta(\beta\kappa\ldots\gamma\alpha)}+\ldots.$$

6.21. When $H(\lambda)=0$ has repeated roots, say α is a
root, we regard A as the limit, when $h\to 0$, of a mat
whose minimum function has zeros

$$\alpha+2h,\quad \alpha+h,\quad \alpha,\quad \beta,\quad \gamma,\quad \ldots.$$

The corresponding form of (10) is

$$\begin{vmatrix} 1 & \alpha+2h & . & . & (\alpha+2h)^{m-1} & f_N(\alpha+2h) \\ 1 & \alpha+h & . & . & (\alpha+h)^{m-1} & f_N(\alpha+h) \\ 1 & \alpha & . & . & \alpha^{m-1} & f_N(\alpha) \\ . & . & . & . & . & . \\ 1 & \kappa & . & . & \kappa^{m-1} & f_N(\kappa) \\ 1 & A_1 & . & . & A_1^{m-1} & f_N(A_1) \end{vmatrix}=0.$$

When $\phi(x)$ is a polynomial and $h\to 0$,

$$\frac{\phi(x+2h)-2\phi(x+h)+\phi(x)}{h^2}\to\phi''(x),\qquad \frac{\phi(x+h)-\phi(x)}{h}\to\phi$$

and so the limiting form of (12), after the appropriate man
tions of the first three rows of the determinant, is

$$\begin{vmatrix} 0 & 0 & . & . & (m-1)(m-2)\alpha^{m-3} & f''_N(\alpha) \\ 0 & 1 & . & . & (m-1)\alpha^{m-2} & f'_N(\alpha) \\ 1 & \alpha & . & . & \alpha^{m-1} & f_N(\alpha) \\ . & . & . & . & . & . \\ 1 & \kappa & . & . & \kappa^{m-1} & f_N(\kappa) \\ 1 & A & . & . & A^{m-1} & f_N(A) \end{vmatrix}=0.$$

We are here supposing that the nth invariant factor of A is

$$E_n(\lambda) \equiv (\lambda-\alpha)^3(\lambda-\beta)\dots(\lambda-\kappa);$$

the canonical form of A can, accordingly, contain only

$$C_1(\alpha),\quad C_2(\alpha),\quad C_3(\alpha),\quad C_1(\beta),\quad \dots,\quad C_1(\kappa)$$

and must contain a $C_3(\alpha)$ because $(\lambda-\alpha)^3$ is an elementary divisor [cf. Chap. IV, §§ 8.5, 8.6; pp. 78, 9]. Hence [cf. Chap. V, § 5.2; p. 92] the sequences $f''_N(\alpha)$, $f'_N(\alpha)$, $f_N(\alpha)$,..., $f_N(\kappa)$ tend to finite limits $f''(\alpha)$, $f'(\alpha)$, $f(\alpha)$,..., $f(\kappa)$ provided that the matrix sequence $f_N(A)$ is convergent. This establishes the theorem when the roots of $H(\lambda) = 0$ are α, α, α, β,..., κ and other examples of repeated roots can be dealt with in the same way.

Alternatively, when $H(\lambda)$ has $(\lambda-\alpha)^3$ as a factor, (6) gives $r'(\alpha) = f'_N(\alpha)$, $r''(\alpha) = f''_N(\alpha)$, and the above limit processes can be avoided.

6.3. The form (5) was noted by Sylvester and is sometimes called Sylvester's interpolation formula,† by analogy with Lagrange's interpolation formula.

7. Abel's theorem for matrix power series

We shall not attempt the systematic development of a full convergence theory for series of matrices, since many of its details could only be wearisome elaborations of the corresponding results for series of scalars. As an example of one type of result we note

THEOREM 23. *Let $f(A)$ denote the sum of the convergent matrix power series*

$$a_0 I + a_1 A + a_2 A^2 + \dots. \tag{1}$$

Then the series $\quad a_0 I + a_1 tA + a_2 t^2 A^2 + \dots$

is convergent when $0 < t < 1$ and its sum tends to $f(A)$ when $t \to 1$.

PROOF. Let the element in the rth row and sth column of A^k be $\alpha_{rs}^{(k)}$. Then

$$\sum_{k=0}^{\infty} a_k A^k$$

is the matrix $$\left[\sum_{k=0}^{\infty} a_k \alpha_{rs}^{(k)}\right], \tag{2}$$

† Cf. Turnbull and Aitken, *Canonical Matrices* (Blackie, 1932), pp. 73–8, where an alternative treatment of Theorem 22 is given.

In detail, on using $\zeta(abc\dots jk)$ to denote the product of differences

$$(a-b)\dots(a-k)(b-c)\dots(b-k)\dots(j-k)$$

and, when A is a matrix, $\zeta(Abc\dots jk)$ to denote

$$(A-bI)\dots(A-kI)(b-c)\dots(b-k)\dots(j-k),$$

the expansion is

$$f(\alpha)\zeta(A\kappa\dots\gamma\beta)-f(\beta)\zeta(A\kappa\dots\gamma\alpha)+\dots+(-1)^m f(A)\zeta(\kappa\dots\beta\alpha).$$

Thus the equation (11) gives

$$f(A)=(-1)^{m-1}f(\alpha)\frac{\zeta(A\kappa\dots\gamma\beta)}{\zeta(\kappa\dots\gamma\beta\alpha)}+(-1)^m f(\beta)\frac{\zeta(A\kappa\dots\gamma\alpha)}{\zeta(\kappa\dots\gamma\beta\alpha)}+\dots$$

$$=f(\alpha)\frac{\zeta(A\kappa\dots\gamma\beta)}{\zeta(\alpha\kappa\dots\gamma\beta)}+f(\beta)\frac{\zeta(A\kappa\dots\gamma\alpha)}{\zeta(\beta\kappa\dots\gamma\alpha)}+\dots.$$

6.21. When $H(\lambda)=0$ has repeated roots, say α is a triple root, we regard A as the limit, when $h\to 0$, of a matrix A_1 whose minimum function has zeros

$$\alpha+2h,\quad \alpha+h,\quad \alpha,\quad \beta,\quad \gamma,\quad \dots.$$

The corresponding form of (10) is

$$\begin{vmatrix} 1 & \alpha+2h & . & . & (\alpha+2h)^{m-1} & f_N(\alpha+2h) \\ 1 & \alpha+h & . & . & (\alpha+h)^{m-1} & f_N(\alpha+h) \\ 1 & \alpha & . & . & \alpha^{m-1} & f_N(\alpha) \\ . & . & . & . & . & . \\ 1 & \kappa & . & . & \kappa^{m-1} & f_N(\kappa) \\ 1 & A_1 & . & . & A_1^{m-1} & f_N(A_1) \end{vmatrix}=0. \qquad (12)$$

When $\phi(x)$ is a polynomial and $h\to 0$,

$$\frac{\phi(x+2h)-2\phi(x+h)+\phi(x)}{h^2}\to\phi''(x),\qquad \frac{\phi(x+h)-\phi(x)}{h}\to\phi'(x),$$

and so the limiting form of (12), after the appropriate manipulations of the first three rows of the determinant, is

$$\begin{vmatrix} 0 & 0 & . & . & (m-1)(m-2)\alpha^{m-3} & f''_N(\alpha) \\ 0 & 1 & . & . & (m-1)\alpha^{m-2} & f'_N(\alpha) \\ 1 & \alpha & . & . & \alpha^{m-1} & f_N(\alpha) \\ . & . & . & . & . & . \\ 1 & \kappa & . & . & \kappa^{m-1} & f_N(\kappa) \\ 1 & A & . & . & A^{m-1} & f_N(A) \end{vmatrix}=0.$$

We are here supposing that the nth invariant factor of A is

$$E_n(\lambda) \equiv (\lambda-\alpha)^3(\lambda-\beta)\dots(\lambda-\kappa);$$

the canonical form of A can, accordingly, contain only

$$C_1(\alpha),\quad C_2(\alpha),\quad C_3(\alpha),\quad C_1(\beta),\quad \dots,\quad C_1(\kappa)$$

and must contain a $C_3(\alpha)$ because $(\lambda-\alpha)^3$ is an elementary divisor [cf. Chap. IV, §§ 8.5, 8.6; pp. 78, 9]. Hence [cf. Chap. V, § 5.2; p. 92] the sequences $f''_N(\alpha)$, $f'_N(\alpha)$, $f_N(\alpha),\dots, f_N(\kappa)$ tend to finite limits $f''(\alpha), f'(\alpha), f(\alpha),\dots, f(\kappa)$ provided that the matrix sequence $f_N(A)$ is convergent. This establishes the theorem when the roots of $H(\lambda) = 0$ are α, α, α, $\beta,\dots$, κ and other examples of repeated roots can be dealt with in the same way.

Alternatively, when $H(\lambda)$ has $(\lambda-\alpha)^3$ as a factor, (6) gives $r'(\alpha) = f'_N(\alpha)$, $r''(\alpha) = f''_N(\alpha)$, and the above limit processes can be avoided.

6.3. The form (5) was noted by Sylvester and is sometimes called Sylvester's interpolation formula,† by analogy with Lagrange's interpolation formula.

7. Abel's theorem for matrix power series

We shall not attempt the systematic development of a full convergence theory for series of matrices, since many of its details could only be wearisome elaborations of the corresponding results for series of scalars. As an example of one type of result we note

THEOREM 23. *Let $f(A)$ denote the sum of the convergent matrix power series*

$$a_0 I + a_1 A + a_2 A^2 + \dots. \tag{1}$$

Then the series $a_0 I + a_1 tA + a_2 t^2A^2 + \dots$

is convergent when $0 < t < 1$ and its sum tends to $f(A)$ when $t \to 1$.

PROOF. Let the element in the rth row and sth column of A^k be $\alpha_{rs}^{(k)}$. Then

$$\sum_{k=0}^{\infty} a_k A^k$$

is the matrix

$$\Big[\sum_{k=0}^{\infty} a_k \alpha_{rs}^{(k)}\Big], \tag{2}$$

† Cf. Turnbull and Aitken, *Canonical Matrices* (Blackie, 1932), pp. 73–8, where an alternative treatment of Theorem 22 is given.

the series being convergent by hypothesis.

When $0 < t < 1$,

$$\sum_{k=0}^{\infty} a_k t^k A^k$$

is the matrix

$$\Big[\sum_{k=0}^{\infty} a_k t^k \alpha_{rs}^{(k)}\Big] \tag{3}$$

and the limit of this matrix as $t \to 1$ is, by Abel's theorem† applied to the series of scalars that form the elements of (3), the matrix (2). This proves the theorem.

8. Functions of a matrix

8.1. *Definition via power series*

Let $f(z)$ denote the sum of the power series

$$\sum_{r=0}^{\infty} a_r z^r$$

and let the series converge when $|z| < R$. Let A be a matrix whose latent roots lie within the circle $|z| = R$; let the classical canonical form of A be

$$C = \operatorname{diag}\{C_p(\lambda_1), C_q(\lambda_2),...\}$$

and

$$A = TCT^{-1}.$$

Then, by p. 92 (8), the matrix power series

$$\sum_{r=0}^{\infty} a_r C^r$$

converges to the sum $f(C)$ given by

$$f(C) = \operatorname{diag}\{F_p(\lambda_1), F_q(\lambda_2),...\},$$

where

$$F_k(\lambda) = \begin{bmatrix} f(\lambda) & f'(\lambda) & f''(\lambda)/2! & . & . & f^{(k-1)}(\lambda)/(k-1)! \\ 0 & f(\lambda) & f'(\lambda) & . & . & f^{(k-2)}(\lambda)/(k-2)! \\ . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & f(\lambda) \end{bmatrix}.$$

Further, by Theorem 20, the matrix power series

$$\sum_{r=0}^{\infty} a_r A^r$$

converges and, if its sum is denoted by $f(A)$,

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

† W. L. Ferrar, *A Text-book of Convergence* (Oxford, 1938), p. 79.

These facts define the function $f(A)$ of a matrix A when the latent roots of A lie within the circle of convergence of $f(z)$. We leave aside the refinements of § 5.2 that may, in particular examples, enable us to define $f(A)$ even when some of the latent roots of A lie on the circle of convergence of $f(z)$.

8.2. *Definition via the functional form*

When a latent root of A lies outside the circle of convergence, the method of § 8.1 ceases to be applicable. Suppose now that $\mathsf{F}(z)$ is a function of the complex variable z which is analytic in the neighbourhood of $z = 0$. For sufficiently small values of $|z|$ the *function* $\mathsf{F}(z)$ can be represented by the sum $f(z)$, say, of a convergent power series

$$\sum_{r=0}^{\infty} a_r z^r.$$

We distinguish between the function $\mathsf{F}(z)$ and the sum $f(z)$ of the power series.

Let the power series converge when $|z| < R$. Then

$$\mathsf{F}(z) \equiv f(z) \quad \text{when} \quad |z| < R. \tag{1}$$

Now let C be a canonical form given by

$$C = \text{diag}\{C_p(\lambda_1), C_q(\lambda_2),...\}, \tag{2}$$

where no latent root λ of C is a singular point of $\mathsf{F}(z)$. When $|t|$ is sufficiently small, the latent roots $t\lambda_1$, $t\lambda_2$,... of the matrix tC all satisfy the condition $|t\lambda| < R$ and the points λ_1, λ_2,... lie within the circle of convergence, $|z| = R$, of

$$f(tz) = \phi(t, z) = \sum_{r=0}^{\infty} a_r t^r z^r. \tag{3}$$

Hence

$$\sum a_r t^r C^r \tag{4}$$

is convergent and its sum is,† by § 8.1,

$$\text{diag}\{F_p(t, \lambda_1), F_q(t, \lambda_2),...\}, \tag{5}$$

† $\dfrac{\partial}{\partial z}\phi(t, z) = \sum r a_r t^r z^{r-1} = t f'(tz) = t\mathsf{F}'(tz)$.

where a typical $F_k(t,\lambda)$ is given by

$$F_k(t,\lambda) = \begin{bmatrix} \mathsf{F}(t\lambda) & t\mathsf{F}'(t\lambda) & . & . & t^{k-1}\mathsf{F}^{(k-1)}(t\lambda)/(k-1)! \\ 0 & \mathsf{F}(t\lambda) & . & . & t^{k-2}\mathsf{F}^{(k-2)}(t\lambda)/(k-2)! \\ . & . & . & . & . \\ 0 & 0 & . & . & \mathsf{F}(t\lambda) \end{bmatrix}. \quad (6)$$

Define $\mathsf{F}(tC)$ to be $\sum a_r t^r C^r$.

Now in (6) we are concerned with the function F as distinct from its series representation and, provided λ is not a singular point of the function, we may put $t = 1$ (but see note at the end in the above and so obtain a definition of $\mathsf{F}(C)$.

That is to say, we define

$$\mathsf{F}(C) = \mathrm{diag}\{F_p(\lambda_1), F_q(\lambda_2),...\}, \quad (7)$$

where a typical $F_k(\lambda)$ is given by

$$F_k(\lambda) = \begin{bmatrix} \mathsf{F}(\lambda) & \mathsf{F}'(\lambda) & . & . & \mathsf{F}^{(k-1)}(\lambda)/(k-1)! \\ 0 & \mathsf{F}(\lambda) & . & . & \mathsf{F}^{(k-2)}(\lambda)/(k-2)! \\ . & . & . & . & . \\ 0 & 0 & . & . & \mathsf{F}(\lambda) \end{bmatrix}. \quad (8)$$

Finally, when A is a transform of C and $A = TCT^{-1}$, we define $\mathsf{F}(A)$ by the equation†

$$\mathsf{F}(A) = T\mathsf{F}(C)T^{-1}. \quad (9)$$

NOTE. Strictly, in order to define $\mathsf{F}(C)$ from (6), we must envisage a domain D within which $\mathsf{F}(z)$ is an analytic function of z and the point $t\lambda$ must be made to tend to λ along a path lying wholly within D. In particular, if $\mathsf{F}(z)$ is not a single-valued function and $f(z)$ is a power series representing a particular branch of $\mathsf{F}(z)$ near $z = 0$, different paths of $t\lambda$ from points within the circle of convergence of $f(z)$ to the point λ may provide different values of $\mathsf{F}(\lambda)$, and so of $\mathsf{F}(C)$. For example,

$$f(z) = 1+\alpha.2z+\frac{\alpha(\alpha+1)}{2!}(2z)^2+... \quad (|z| < \tfrac{1}{2}),$$

when α is not an integer, represents a particular branch of $(1-2z)^{-\alpha}$. This function has a branch point at $z = \frac{1}{2}$; the domain D may be taken to be any finite part of the z-plane with the point $z = \frac{1}{2}$ excluded. If C has a latent root $\lambda = 1$, the value of the function $(I-2C)^{-\alpha}$ will depend on the path followed by t as it comes from points within the circle $|t| = \frac{1}{2}$ to the point $t = 1$.

We do not dwell on such points, which belong rather to the theory of functions of a complex variable.

† We return to this definition in § 8.6, where we prove that (9) provides a *unique* definition of $\mathsf{F}(A)$.

8.3. *A simple example of the definition*

For simplicity, take the canonical form C to be a single classical form $C_2(\alpha)$, where $|\alpha|$ is unrestricted in value. When $|\alpha| < 1$, the matrix power series

$$I+C+C^2+...$$

is convergent, its sum being denoted by $(I-C)^{-1}$, and [by (7) of § 5.1] it is represented in matrix form by

$$\begin{bmatrix} 1+\alpha+\alpha^2+... & 1+2\alpha+3\alpha^2+... \\ 0 & 1+\alpha+\alpha^2+... \end{bmatrix} \tag{1}$$

i.e.
$$\begin{bmatrix} (1-\alpha)^{-1} & (1-\alpha)^{-2} \\ 0 & (1-\alpha)^{-1} \end{bmatrix}. \tag{2}$$

The series form (1) ceases to be applicable when $|\alpha| \geqslant 1$ but, provided that $\alpha \neq 1$, i.e. provided that α is not a singularity of the *function* $(1-z)^{-1}$, the functional form (2) remains valid regardless of the value of $|\alpha|$.

It is easy to verify that when (2) is multiplied by $I-C$, that is by
$$\begin{bmatrix} 1-\alpha & -1 \\ 0 & 1-\alpha \end{bmatrix},$$

the result is I; and so (2) defines $(I-C)^{-1}$, *the* reciprocal of $I-C$ (there can be only one reciprocal).

8.4. *Rational functions*

The concluding remark of § 8.3 raises a general question; having two functions $f(z)$ and $h(z)$ of a scalar variable z whose product is $g(z)$, so that

$$f(z)h(z) = g(z),$$

and having defined the matrix functions $f(A)$, $g(A)$, $h(A)$ of a matrix A, as in § 8.1 or § 8.2, is it true that $f(A)h(A) = g(A)$? We shall show that it is true.

Let $f(z)$, $h(z)$ be analytic functions of z regular in the neighbourhood of the origin and of each latent root of a classical canonical matrix

$$C = \text{diag}\{C_p, C_q, ...\}, \tag{3}$$

and let $f(z)h(z) = g(z)$.

Let C_k be a typical submatrix in (3) and let the matrices $f(C_k)$, $h(C_k)$, $g(C_k)$ be defined† as in § 8.2. We shall prove that

$$f(C_k)h(C_k) = g(C_k). \tag{4}$$

The left-hand side of (4) is the product of the matrix

$$f(C_k) = \begin{bmatrix} f(\lambda) & f'(\lambda) & . & . & f^{(k-1)}(\lambda)/(k-1)! \\ 0 & f(\lambda) & . & . & f^{(k-2)}(\lambda)/(k-2)! \\ . & . & . & . & . \\ 0 & 0 & . & . & f(\lambda) \end{bmatrix}, \tag{5}$$

by a corresponding matrix having h instead of f. In this product the $(n+1)$th element of the first row [the order of multiplication is $f.h$] is‡

$$\frac{1}{n!}D^0f.D^nh+\frac{1}{(n-1)!}Df.D^{n-1}h+\frac{1}{(n-2)!\,2!}D^2fD^{n-2}h+...,$$

which, by Leibnitz's theorem, is equal to

$$\frac{1}{n!}D^n(f.h) = \frac{1}{n!}D^n(g).$$

That is to say, the first row of the product is

$$g(\lambda), \quad g'(\lambda), \quad \frac{1}{2!}g''(\lambda), \quad ....$$

Similarly, the second row of the product is

$$0, \quad g(\lambda), \quad g'(\lambda), \quad ...,$$

and so on for the other rows. Thus the product of $f(C_k)$ and $h(C_k)$ is $g(C_k)$ and the equation (4) is seen to be a deduction from Leibnitz's theorem. Now

$$f(C) = \text{diag}\{f(C_p), f(C_q),...\} \tag{6}$$

and so for $h(C)$ and $g(C)$. Accordingly§

$$f(C)h(C) = \text{diag}\{f(C_p)h(C_p), f(C_q)h(C_q),...\}$$

and this, by (4), gives

$$f(C)h(C) = \text{diag}\{g(C_p), g(C_q),...\} = g(C).$$

† This definition uses the function and includes the infinite series form of definition; e.g. to use $1+z+z^2+...$ requires all latent roots of C to lie within $|z| = 1$, while to use the function $(1-z)^{-1}$ requires only that no latent root be equal to unity.

‡ We use D^rf to denote $f^{(r)}(\lambda)$.

§ Cf. Chap. I, § 12 (b).

Finally, when $A = TCT^{-1}$ so that [§ 8.2 (9)] $f(A) = Tf(C)T^{-1}$, etc.,

$$f(A)h(A) = g(A). \tag{7}$$

In particular, if $h(A)$ is non-singular,

$$f(A) = g(A)\{h(A)\}^{-1}$$

and, when $g(z)$ and $h(z)$ are polynomials, $f(A)$ is a rational function of A; the preceding work of this subsection reconciles the general definition of $f(A)$ given in § 8.2 to the more elementary definition of a rational function of a matrix given in Chapter I, § 4(d) — it shows that $f(A)$ defined by means of (5) is the matrix obtained by multiplying $g(A)$ by the reciprocal of $h(A)$.

8.5. *Fractional powers of matrices*

We now show that (7) of § 8.2 serves also to define a matrix whose qth power is equal to a given matrix. Take first a single classical submatrix, $C_k(\lambda)$ say. Then, when m and n are positive integers, the work of § 5.1 shows that

$$C^m = \begin{bmatrix} \lambda^m & m\lambda^{m-1} & \frac{1}{2}m(m-1)\lambda^{m-2} & . & . & . \\ 0 & \lambda^m & m\lambda^{m-1} & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & \lambda^m \end{bmatrix},$$

while C^n is obtained by writing n for m. Moreover C^{m+n} is obtained by writing $m+n$ for m and, m and n being integers,

$$C^m . C^n = C^{m+n}. \tag{1}$$

But equation (1) is a conspectus of the scalar equations that equate coefficients of powers of λ in the individual elements of the product $C^m . C^n$ to the corresponding coefficients in the individual elements of the matrix C^{m+n}. Considered as equations in the variable m, each of these is of finite degree: when n is an integer they hold for $m = 1, 2, 3,...$ and are therefore true for *all* values of m provided n is an integer. Again, each equation is of finite degree in n; and, as we have just seen, each equation is true for any given m when $n = 1, 2, 3,...$. It is consequently† an identity in n. Thus the scalar equations

† The argument is the familiar 'double induction' often used to prove the validity of the binomial series; cf. Ferrar, *Convergence*, p. 95.

hold for all values of m and n. Consequently, if we *define* C^x when x is not an integer by

$$C^x = \begin{bmatrix} \lambda^x & x\lambda^{x-1} & \frac{1}{2}x(x-1)\lambda^{x-2} & . & . & . \\ 0 & \lambda^x & x\lambda^{x-1} & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & \lambda^x \end{bmatrix}, \tag{2}$$

which we may do provided we exclude $\lambda = 0$ when some $x-r$ is negative, the relation

$$C^x . C^y = C^{x+y} \tag{3}$$

will hold for all values of x and y.

It follows that, when n is an integer,

$$(C^x)^n = C^{xn} \tag{4}$$

and that an nth root of C is given by taking $x = 1/n$ in (2) provided that $\lambda \neq 0$.† There are, of course, n values of $\lambda^{1/n}$ and any one of these may be used: $C^{1/n}$ is an n-valued function of C.

When C is more complex, say

$$C = \text{diag}\{C_p(\lambda_1), C_q(\lambda_2),...\},$$

an nth root of C is given by

$$\text{diag}\{C_p^{1/n}(\lambda_1), C_q^{1/n}(\lambda_2),...\}, \tag{5}$$

provided that no λ_r is zero. In this expression each submatrix is capable of n determinations corresponding to the n different values of each of the functions $\lambda_1^{1/n}$, $\lambda_2^{1/n}$,... .

NOTE. It is not suggested that this sub-section is anything more than an indication of how matrices X can be defined so that $X^n = C$. It does not touch on the question 'What is the most general form of matrix X for which, say, $X^2 = C$?'

8.6. *Note on the definition* $\mathsf{F}(A) = T\mathsf{F}(C)T^{-1}$

When A and C are given matrices so related that we can determine a matrix T for which $A = TCT^{-1}$, this matrix T cannot be determined uniquely. In fact, when B is any non-singular matrix that commutes with A and we put

$$K = BT,$$

† We cannot define $\lambda^{(1-n)/n}$ when $\lambda = 0$ and it is because of this that we must exclude the zero value.

we find that, since T is non-singular and $AT = TC$,

(i) K is non-singular;

(ii) $KC = BTC = BAT$,

$AK = ABT = BAT$;

(iii) $AK = KC$,

so that $$A = KCK^{-1}.$$

It is not obvious that $f(A)$ defined as $Kf(C)K^{-1}$ will be the same as $Tf(C)T^{-1}$. When we were dealing with power series (§ 3.1) we proved that, assuming convergence,

$$\sum a_n A^n = T(\sum a_n C^n)T^{-1}$$

whenever $A = TCT^{-1}$. In proving this result we defined the left-hand side as the limit of the sum of N terms of the series and this definition is independent of T. Accordingly, when we are dealing with power series $f(z)$ and matrices A and C whose latent roots lie within the circle of convergence of $f(z)$, the *definition* of $f(A)$ by means of the equation

$$f(A) = Tf(C)T^{-1}$$

is a unique definition and independent of what particular T we use in the equation $A = TCT^{-1}$.

Now let $\mathsf{F}(z)$ be an analytic function regular in the neighbourhood of the origin. Let C be the canonical form of A and let

$$A = TCT^{-1}, \qquad A = KCK^{-1}.$$

Then, provided $|t|$ is sufficiently small, $\mathsf{F}(tA)$ and $\mathsf{F}(tC)$ are the sums of convergent matrix power series and, as we have just seen,

$$T\mathsf{F}(tC)T^{-1} \quad \text{and} \quad K\mathsf{F}(tC)K^{-1} \tag{1}$$

are one and the same matrix. The matrix $\mathsf{F}(tC)$ has elements $\mathsf{F}(t\lambda)$, $t\mathsf{F}'(t\lambda)$,... (§ 8.2), and the elements in the rth row and sth column of the two products in (1) are of the form

$$\phi_{rs}(t), \qquad \psi_{rs}(t),$$

where ϕ_{rs}, ψ_{rs} denote analytic functions of the variable t. We have proved above that

$$\phi_{rs}(t) = \psi_{rs}(t)$$

when $|t|$ is sufficiently small. It follows from the theory of analytic functions that the two remain equal throughout their domain of definition; in particular, when no latent root of C is a singular point of $\mathsf{F}(z)$, the equality holds when $t = 1$ and thus, from (1),

$$T\mathsf{F}(C)T^{-1} \quad \text{and} \quad K\mathsf{F}(C)K^{-1}$$

are one and the same matrix.

8.7. *Matrices $f(A)$ and $h(A)$ are commutative*

Let $f(z)$, $h(z)$ be analytic functions of z regular in the neighbourhood of the origin and of each latent root of a matrix A; let $f(z)h(z) = g(z)$ and let the matrices $f(A)$, $h(A)$, $g(A)$ be defined as in § 8.2.

We have already proved by means of Leibnitz's theorem (§ 8.4) that

$$f(A)h(A) = g(A).$$

When we reverse the order of f, h and write $h(z)f(z) = g(z)$, the same work gives

$$h(A)f(A) = g(A).$$

Hence
$$f(A)h(A) = h(A)f(A)$$

and *any two functions of a matrix A are commutative matrices.*

9. *The canonical form of $f(C)$*

Let
$$C = \text{diag}\{C_n(\lambda),..., ...\}$$

be a canonical matrix and $f(x)$ a function of the scalar variable x. Then

$$f(C) = \text{diag}[f\{C_n(\lambda)\},..., ...],$$

where

$$f\{C_n(\lambda)\} = \begin{bmatrix} f(\lambda) & f'(\lambda) & . & . & f^{(n-1)}(\lambda)/(n-1)! \\ 0 & f(\lambda) & . & . & f^{(n-2)}(\lambda)/(n-2)! \\ . & . & . & . & . \\ 0 & 0 & . & . & f(\lambda) \end{bmatrix}. \qquad (1)$$

In order to find the canonical form of the full matrix $f(C)$, it is sufficient to find the canonical form of a typical submatrix $f\{C_n(\lambda)\}$.

9.1. *When $f'(\lambda) \neq 0$*

When $f'(\lambda) \neq 0$, appropriate H changes (cf. Chap. IV, § 4.2) will give a transform of (1) in which the diagonal elements are $f(\lambda)$, the elements one place to the right are $f'(\lambda)$, and all other domain of definition; in particular, when no latent root of C is a singular point of $\mathsf{F}(z)$, the equality holds when $t = 1$ and thus, from (1),

$$T\mathsf{F}(C)T^{-1} \quad \text{and} \quad K\mathsf{F}(C)K^{-1}$$

are one and the same matrix.

8.7. *Matrices $f(A)$ and $h(A)$ are commutative*

Let $f(z)$, $h(z)$ be analytic functions of z regular in the neighbourhood of the origin and of each latent root of a matrix A; let $f(z)h(z) = g(z)$ and let the matrices $f(A)$, $h(A)$, $g(A)$ be defined as in § 8.2.

We have already proved by means of Leibnitz's theorem (§ 8.4) that

$$f(A)h(A) = g(A).$$

When we reverse the order of f, h and write $h(z)f(z) = g(z)$, the same work gives

$$h(A)f(A) = g(A).$$

Hence $$f(A)h(A) = h(A)f(A)$$

and *any two functions of a matrix A are commutative matrices.*

9. *The canonical form of $f(C)$*

Let $$C = \text{diag}\{C_n(\lambda),..., ...\}$$

be a canonical matrix and $f(x)$ a function of the scalar variable x. Then

$$f(C) = \text{diag}[f\{C_n(\lambda)\},..., ...],$$

where

$$f\{C_n(\lambda)\} = \begin{bmatrix} f(\lambda) & f'(\lambda) & . & . & f^{(n-1)}(\lambda)/(n-1)! \\ 0 & f(\lambda) & . & . & f^{(n-2)}(\lambda)/(n-2)! \\ . & . & . & . & . \\ 0 & 0 & . & . & f(\lambda) \end{bmatrix}. \qquad (1)$$

In order to find the canonical form of the full matrix $f(C)$, it is sufficient to find the canonical form of a typical submatrix $f\{C_n(\lambda)\}$.

9.1. *When $f'(\lambda) \neq 0$*

When $f'(\lambda) \neq 0$, appropriate H changes (cf. Chap. IV, § 4.2) will give a transform of (1) in which the diagonal elements are $f(\lambda)$, the elements one place to the right are $f'(\lambda)$, and all other elements are zero. The canonical form is, after an X change (cf. Chap. IV, § 5.3), simply $C_n\{f(\lambda)\}$.

9.2. *When* $f'(\lambda) = 0$

Let $f^{(k)}(\lambda)$ be the first of $f'(\lambda)$, $f''(\lambda)$,... to differ from zero. The canonical form of (1) is not quite so simple as in § 9.1, but

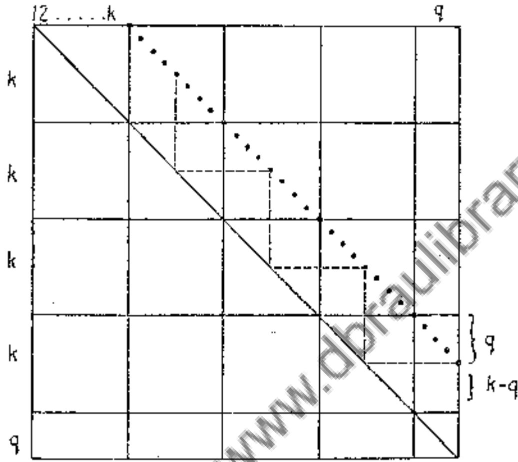


DIAGRAM 2.

the method of chains, given in Chapter IV, § 5.10, enables us to find the canonical form without too much difficulty.

The first row of (1) is

$$f(\lambda) \quad 0 \quad 0 \quad \ldots \quad f^{(k)}(\lambda)/k! \quad \ldots \quad f^{(n-1)}(\lambda)/(n-1)!.$$

Appropriate H changes give a transform of (1) in which all terms after $f^{(k)}(\lambda)/k!$ are replaced by zeros; and so for the other rows. On marking the positions only of the non-zeros that lie off the principal diagonal, the transform of (1) just obtained appears as

$$\begin{matrix} . & . & . & . & . & (1,k+1) & & \\ . & . & . & . & . & . & (2,k+2) & \\ & & & & & & & \ddots \\ & & & & & & & & (n-k,n). \end{matrix}$$

Let $n = pk+q$, where $0 \leqslant q < k$. Then (with $p = 4$ and $q > 0$) Diagram 2 above gives a picture of the complete matrix under consideration. In this diagram, the heavy dots mark the non-zeros that lie off the principal diagonal, while the broken line shows the chain formed by the linked non-zero elements

$$(q,k+q),\quad (k+q,2k+q),\quad \ldots\quad \{(p-1)k+q,pk+q\}.$$

This chain has p non-zero elements; moreover, there are q distinct chains of this length, beginning with the non-zero elements $(1,k+1)$, $(2,k+2)$,..., $(q,k+q)$ respectively. Hence the canonical form of the matrix contains $C_{p+1}\{f(\lambda)\}$ repeated q times. Again, there are $k-q$ chains, beginning with the non-zero elements $(q+1,k+q+1)$,..., $(k,2k)$ respectively, which have only $p-1$ non-zero elements: these chains end in the last square of side-length k because there are no non-zero elements available to carry the chain into the strip of width q on the right of the diagram. Thus the canonical form also contains $C_p\{f(\lambda)\}$ repeated $k-q$ times.

Since $$p(k-q)+(p+1)q = pk+q = n,$$

the n by n matrix can contain no other terms. Hence *the canonical form of $f\{C_n(\lambda)\}$ is given by the following rule:*†

If $f^{(k)}(\lambda)$ is the first of $f'(\lambda)$, $f''(\lambda)$,... to differ from zero, and $n = pk+q$, where $0 \leqslant q < k$, then

$$f\{C_n(\lambda)\} \sim \text{diag}[C_p\{f(\lambda)\},\ldots C_{p+1}\{f(\lambda)\},\ldots],$$

the C_p repeated $k-q$ times and C_{p+1} repeated q times.

9.3. *An alternative proof*

The same result can, with some thought and ingenuity, be derived directly from the H.C.F.'s of minors of a matrix of order $n = pk+q$. Consider a matrix with α in each diagonal place, unity in each place k to the right of the diagonal, and zeros elsewhere.

Let D_r be the H.C.F. of minors having r rows and columns. At sight, and using only the elements 1,

$$D_1 = D_2 = \ldots = D_{n-k} = 1.$$

To get a non-zero minor with $n-k+1$ rows and columns we must include an α; and to include one α is to be forced to

† The rule is given in D. E. Rutherford, *Proc. Edinburgh Math. Soc.* (2), 3 (1932), 135–43.

include an α from the next square and, with a little trouble over details, a total of α^p. Thus

$$D_{n-k+1} = \alpha^p$$

and, in the same way, an additional α^p at each step until we get to D_{n-q}; thereafter, we have to include $p+1$ extra α's to

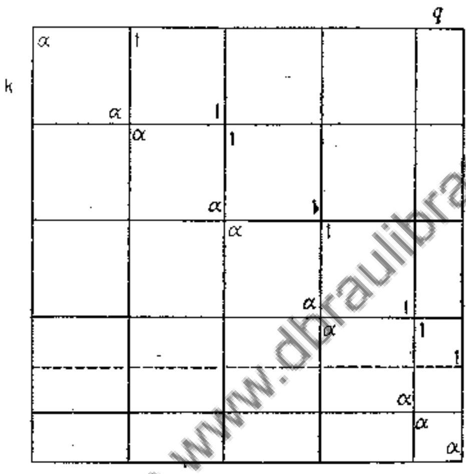


DIAGRAM 3.

complete a non-zero minor having an extra row and column and an additional α^{p+1} enters into the H.C.F.† Thus

$$D_{n-k+1} = \alpha^p, \quad D_{n-k+2} = \alpha^{2p}, \quad \ldots, \quad D_{n-q} = \alpha^{(k-q)p},$$
$$D_{n-q+1} = \alpha^{p+1} D_{n-q}, \ldots.$$

These results show that

$$E_1 = E_2 = \ldots = E_{n-k} = 1,$$
$$E_{n-k+1} = \ldots = E_{n-q} = \alpha^p,$$
$$E_{n-q+1} = \ldots = E_n = \alpha^{p+1},$$

and that the canonical form is

$$\text{diag}\{C_p(\alpha) \ldots (k-q) \text{ times}, \quad C_{p+1}(\alpha) \ldots q \text{ times}\}.$$

† The argument is indicated only: the reader may care to accept the challenge of *proving* the result this way.

10. Differentiation of matrix functions

10.1. *Differentiation with respect to a scalar variable*

When the elements of a matrix are functions of a scalar variable t, say
$$A(t) = [a_{rs}(t)],$$
we define dA/dt to be the limit as $h \to 0$ of
$$\{A(t+h)-A(t)\}/h.$$
This limit is well defined and is equal to the matrix
$$\left[\frac{d}{dt}a_{rs}(t)\right],$$
provided only that each $a_{rs}(t)$ is differentiable.

The matrices $A(t)$ and dA/dt are not necessarily commutative.

10.2. *Differentiation with respect to a matrix variable*

Let $\phi(A)$ be a given function (according to the definition in § 8.2) of a matrix A. Then we shall show that, as $h \to 0$,
$$\{\phi(A+hI)-\phi(A)\}/h \tag{1}$$
tends to a limit matrix: this matrix we call the derivative† of $\phi(A)$ with respect to A and we denote it by the notations common in scalar differentiation, such as
$$\phi'(A), \qquad \frac{d\phi(A)}{dA}.$$

In order to prove that (1) does tend to a limit as $h \to 0$ we first consider $\phi(C)$, where
$$C = T^{-1}AT$$
is the classical canonical form of A.

When
$$C = \text{diag}\{C_k(\lambda),...\},$$
the definition of § 8.2 defines $\phi(C)$, provided no latent root of C is a singularity of the scalar function $\phi(z)$, by the equation
$$\phi(C) = \text{diag}\{F_k(\lambda),...\}, \tag{2}$$
where
$$F_k(\lambda) = \begin{bmatrix} \phi(\lambda) & \phi'(\lambda) & . & . & \phi^{(k-1)}(\lambda)/(k-1)! \\ 0 & \phi(\lambda) & . & . & \phi^{(k-2)}(\lambda)/(k-2)! \\ . & . & . & . & . \\ 0 & 0 & . & . & \phi(\lambda) \end{bmatrix}. \tag{3}$$

† The only reference I have found in the literature is Born and Jordan, *Elementare Quantenmechanik* (Springer, Berlin, 1930), p. 38.

Further, $\phi(C+hI)$ is obtained from the above by writing $\lambda+h$ instead of λ and, as $h \to 0$,

$$\{\phi(C+hI)-\phi(C)\}/h$$

tends to a matrix† of which a typical submatrix is

$$F'_k(\lambda) = \begin{bmatrix} \phi'(\lambda) & \phi''(\lambda) & . & . & \phi^{(k)}(\lambda)/(k-1)! \\ 0 & \phi'(\lambda) & . & . & \phi^{(k-1)}(\lambda)/(k-2)! \\ . & . & . & . & . \\ 0 & 0 & . & . & \phi'(\lambda) \end{bmatrix}. \tag{4}$$

Thus $$\phi'(C) = \text{diag}\left\{\frac{dF_k(\lambda)}{d\lambda}, \quad ...\right\}. \tag{5}$$

We now prove that, as $h \to 0$,

$$\{\phi(A+hI)-\phi(A)\}/h$$

tends to a limit $\phi'(A)$ given by

$$\phi'(A) = T\phi'(C)T^{-1}. \tag{6}$$

Since $A = TCT^{-1}$ and $A+hI = T(C+hI)T^{-1}$, the functions $\phi(A+hI)$ and $\phi(A)$ are defined by

$$\phi(A+hI) = T\phi(C+hI)T^{-1}, \qquad \phi(A) = T\phi(C)T^{-1}.$$

Accordingly,

$$\frac{\phi(A+hI)-\phi(A)}{h} = T\,\frac{\phi(C+hI)-\phi(C)}{h}\,T^{-1}. \tag{7}$$

We have proved that the middle term on the right tends to a limit $\phi'(C)$ and, as in § 3.1 (with $h \to 0$ instead of $N \to \infty$), it follows‡ that the right-hand side of (7) tends to a limit

$$T\phi'(C)T^{-1}. \tag{8}$$

This proves (6).

10.3. *Differentiation of matrix power series*

Let $\sum a_n z^n$ converge when $|z|$ is sufficiently small. Then it represents in the neighbourhood of the origin an analytic

† Since λ is not a singularity of $\phi(z)$, $\{\phi(\lambda+h)-\phi(\lambda)\}h^{-1} \to \phi'(\lambda)$, and so for the other elements of $F_k(\lambda)$.

‡ In brief, when $T = [t_{ij}]$, $T^{-1} = [\theta_{ij}]$, and $\gamma(h) = [\gamma_{ij}(h)] \to [\gamma_{ij}] = \gamma$, it follows that $T\gamma(h)T^{-1} = [t_{ik}\,\gamma_{kl}(h)\theta_{lj}] \to [t_{ik}\,\gamma_{kl}\,\theta_{lj}] = T\gamma T^{-1}$.

10. Differentiation of matrix functions

10.1. *Differentiation with respect to a scalar variable*

When the elements of a matrix are functions of a scalar variable t, say
$$A(t) = [a_{rs}(t)],$$
we define dA/dt to be the limit as $h \to 0$ of
$$\{A(t+h)-A(t)\}/h.$$
This limit is well defined and is equal to the matrix
$$\left[\frac{d}{dt}a_{rs}(t)\right],$$
provided only that each $a_{rs}(t)$ is differentiable.

The matrices $A(t)$ and dA/dt are not necessarily commutative.

10.2. *Differentiation with respect to a matrix variable*

Let $\phi(A)$ be a given function (according to the definition in § 8.2) of a matrix A. Then we shall show that, as $h \to 0$,
$$\{\phi(A+hI)-\phi(A)\}/h \tag{1}$$
tends to a limit matrix: this matrix we call the derivative† of $\phi(A)$ with respect to A and we denote it by the notations common in scalar differentiation, such as
$$\phi'(A), \qquad \frac{d\phi(A)}{dA}.$$

In order to prove that (1) does tend to a limit as $h \to 0$ we first consider $\phi(C)$, where
$$C = T^{-1}AT$$
is the classical canonical form of A.

When
$$C = \text{diag}\{C_k(\lambda),...\},$$
the definition of § 8.2 defines $\phi(C)$, provided no latent root of C is a singularity of the scalar function $\phi(z)$, by the equation
$$\phi(C) = \text{diag}\{F_k(\lambda),...\}, \tag{2}$$
where
$$F_k(\lambda) = \begin{bmatrix} \phi(\lambda) & \phi'(\lambda) & . & . & \phi^{(k-1)}(\lambda)/(k-1)! \\ 0 & \phi(\lambda) & . & . & \phi^{(k-2)}(\lambda)/(k-2)! \\ . & . & . & . & . \\ 0 & 0 & . & . & \phi(\lambda) \end{bmatrix}. \tag{3}$$

† The only reference I have found in the literature is Born and Jordan, *Elementare Quantenmechanik* (Springer, Berlin, 1930), p. 38.

Further, $\phi(C+hI)$ is obtained from the above by writing $\lambda+h$ instead of λ and, as $h \to 0$,

$$\{\phi(C+hI)-\phi(C)\}/h$$

tends to a matrix† of which a typical submatrix is

$$F'_k(\lambda) = \begin{bmatrix} \phi'(\lambda) & \phi''(\lambda) & . & . & \phi^{(k)}(\lambda)/(k-1)! \\ 0 & \phi'(\lambda) & . & . & \phi^{(k-1)}(\lambda)/(k-2)! \\ . & . & . & . & . \\ 0 & 0 & . & . & \phi'(\lambda) \end{bmatrix}. \tag{4}$$

Thus
$$\phi'(C) = \text{diag}\left\{\frac{dF_k(\lambda)}{d\lambda}, \quad ...\right\}. \tag{5}$$

We now prove that, as $h \to 0$,

$$\{\phi(A+hI)-\phi(A)\}/h$$

tends to a limit $\phi'(A)$ given by

$$\phi'(A) = T\phi'(C)T^{-1}. \tag{6}$$

Since $A = TCT^{-1}$ and $A+hI = T(C+hI)T^{-1}$, the functions $\phi(A+hI)$ and $\phi(A)$ are defined by

$$\phi(A+hI) = T\phi(C+hI)T^{-1}, \qquad \phi(A) = T\phi(C)T^{-1}.$$

Accordingly,

$$\frac{\phi(A+hI)-\phi(A)}{h} = T\,\frac{\phi(C+hI)-\phi(C)}{h}\,T^{-1}. \tag{7}$$

We have proved that the middle term on the right tends to a limit $\phi'(C)$ and, as in § 3.1 (with $h \to 0$ instead of $N \to \infty$), it follows‡ that the right-hand side of (7) tends to a limit

$$T\phi'(C)T^{-1}. \tag{8}$$

This proves (6).

10.3. *Differentiation of matrix power series*

Let $\sum a_n z^n$ converge when $|z|$ is sufficiently small. Then it represents in the neighbourhood of the origin an analytic

† Since λ is not a singularity of $\phi(z)$, $\{\phi(\lambda+h)-\phi(\lambda)\}h^{-1} \to \phi'(\lambda)$, and so for the other elements of $F_k(\lambda)$.

‡ In brief, when $T = [t_{ij}]$, $T^{-1} = [\theta_{ij}]$, and $\gamma(h) = [\gamma_{ij}(h)] \to [\gamma_{ij}] = \gamma$, it follows that $T\gamma(h)T^{-1} = [t_{ik}\,\gamma_{kl}(h)\theta_{lj}] \to [t_{ik}\,\gamma_{kl}\,\theta_{lj}] = T\gamma T^{-1}$.

function $\phi(z)$. Let A be a square matrix and $C = T^{-1}AT$ its classical canonical form; let

$$C = \text{diag}\{C_k(\lambda),...\}. \tag{9}$$

In addition to these data we make the sole hypothesis '$\sum na_n A^{n-1}$ is convergent'. Then (Theorem 20) $\sum na_n C^{n-1}$ is also convergent. Further (corollary to Theorem 21), the series $\sum na_n \lambda^{n-1}$ and the series obtained by differentiating it once, twice,..., $k-1$ times are convergent and their sums are given by $\phi'(\lambda), \phi''(\lambda),..., \phi^{(k)}(\lambda)$; and that for each λ and corresponding k of (9). Hence, by § 5.2,

$$\sum na_n C^{n-1} = \text{diag}\{G_k(\lambda),...\}, \tag{10}$$

where

$$G_k(\lambda) = \begin{bmatrix} \phi'(\lambda) & \phi''(\lambda) & . & . & \phi^{(k)}(\lambda)/(k-1)! \\ 0 & \phi'(\lambda) & . & . & \phi^{(k-1)}(\lambda)/(k-2)! \\ . & . & . & . & . \\ 0 & 0 & . & . & \phi'(\lambda) \end{bmatrix}. \tag{11}$$

Again, the convergence of $\sum na_n \lambda^{n-1}$ to $\phi'(\lambda)$ implies the convergence of $\sum a_n \lambda^n$ to $\phi(\lambda)$. Thus, on using what we have already proved above, the series $\sum a_n \lambda^n$ and the series obtained by differentiating it once, twice,..., $k-1$ times are convergent, their sums being $\phi(\lambda)$, $\phi'(\lambda)$,..., $\phi^{(k-1)}(\lambda)$. Hence the series $\sum a_n C^n$ is convergent and its sum is equal to

$$\phi(C) = \text{diag}\{F_k(\lambda),...\},$$

where $F_k(\lambda)$ is defined as in (3). Moreover, $d\phi/dC$ is now obtained as in § 10.2 and, as a comparison of (4) and (11) shows,

$$\frac{d\phi(C)}{dC} = \text{diag}\{G_k(\lambda),...\} = \sum na_n C^{n-1}.$$

This proves
$$\phi'(C) = \sum na_n C^{n-1}.$$
Hence
$$T\phi'(C)T^{-1} = T(\sum na_n C^{n-1})T^{-1},$$
so that, by (6) and Theorem 20,

$$\phi'(A) = \sum na_n A^{n-1}.$$

We have thus proved

THEOREM 24. *Given that*

(i) $\sum a_n z^n$ *converges for some non-zero value of* z,
(ii) $\sum na_n A^{n-1}$ *is convergent,*

then $$\frac{d}{dA}\sum a_n A^n = \sum n a_n A^{n-1}. \tag{12}$$

NOTE. The argument is slightly simpler at certain points if we assume that all the latent roots of A lie within the circle of convergence of $\sum a_n z^n$, but there is some interest in proving (12) under the minimum hypothesis, that the right-hand side of (12) converges.

11. Some details in the algebra of power series

In this section we suppose $f(z)$, $g(z)$ to be analytic functions in a domain containing the origin: their expansions are taken to be

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \quad \text{with radius of convergence } R,$$

$$g(z) = \sum_{n=0}^{\infty} b_n z^n \quad \text{with radius of convergence } \rho.$$

When $|z| < R$, the mth power of $f(z)$ can be expanded as a power series in z and we suppose $a_{n,m}$ to be the coefficient of z^n in this expansion; thus

$$\Big(\sum_{n=0}^{\infty} a_n z^n\Big)^m = \sum_{n=0}^{\infty} a_{n,m} z^n.$$

11.1. *The product of two power series*

When z lies within the circles of convergence of $f(z)$ and $g(z)$,

$$f(z).g(z) = \sum_{n=0}^{\infty} (a_0 b_n + a_1 b_{n-1} + \ldots + a_n b_0) z^n$$
$$= \sum_{n=0}^{\infty} c_n z^n$$

say. Denote the sum of this series by $h(z)$.

Let C be a classical canonical form, say

$$C = \operatorname{diag}\{C_k(\lambda),\ldots\},$$

whose latent roots lie within the circles of convergence of $f(z)$ and $g(z)$. Then a typical submatrix of $f(C)$ is

$$f(C_k) = \begin{bmatrix} f(\lambda) & f'(\lambda) & . & . & f^{(k-1)}(\lambda)/(k-1)! \\ 0 & f(\lambda) & . & . & f^{(k-2)}(\lambda)/(k-2)! \\ . & . & . & . & . \\ 0 & 0 & . & . & f(\lambda) \end{bmatrix},$$

and a typical submatrix of $g(C)$ is got by replacing f by g.

Form the product $f(C_k).g(C_k)$. Then, by Leibnitz's theorem (as in § 8.4),

$$f(C_k).g(C_k) = h(C_k)$$

and so (again as in § 8.4)

$$f(C).g(C) = h(C).$$

On writing $A = TCT^{-1}$,

$$\begin{aligned} f(A).g(A) &= Tf(C)T^{-1}.Tg(C)T^{-1} \\ &= Tf(C)g(C)T^{-1} \\ &= Th(C)T^{-1} = h(A). \end{aligned}$$

That is to say,

THEOREM 25. *Provided the latent roots of A lie within the circles of convergence of the two power series* $\sum a_n z^n$ *and* $\sum b_n z^n$,

$$\Big(\sum_{n=0}^{\infty} a_n A^n\Big)\Big(\sum_{n=0}^{\infty} b_n A^n\Big) = \sum_{n=0}^{\infty} c_n A^n,$$

where $$c_n = a_0 b_n + a_1 b_{n-1} + \dots + a_n b_0.$$

COROLLARY. The result can be extended to the product of any finite number of series. In particular, *when the latent roots of A lie within the circle of convergence of* $\sum a_n z^n$,

$$\Big(\sum_{n=0}^{\infty} a_n A^n\Big)^m = \sum_{n=0}^{\infty} a_{n,m} A^n.$$

11.2. *Substitution of one power series in another*

Preliminary. Let $|\lambda| < R$, the radius of convergence of

$$f(z) = a_0 + a_1 z + a_2 z^2 + \dots.$$

Let I be the unit matrix, U the auxiliary unit matrix of order k, say; then $U^k = U^{k+1} = \dots = 0$, the null matrix. Thus

$$\begin{aligned} f(\lambda I + U) &= \lim_{N\to\infty} \sum_{n=0}^{N} a_n(\lambda I + U)^n \\ &= \lim_{N\to\infty} \sum_{n=0}^{N} a_n \times \\ &\qquad \times \Big\{\lambda^n I + n\lambda^{n-1} U + \dots + \frac{n(n-1)\dots(n-k+2)}{(k-1)!}\lambda^{n-k+1} U^{k-1}\Big\} \\ &= f(\lambda) I + f'(\lambda) U + \dots + f^{(k-1)}(\lambda)\frac{U^{k-1}}{(k-1)!}. \end{aligned}$$

Also, on writing f for $f(\lambda)$, f' for $f'(\lambda)$, f^m for $\{f(\lambda)\}^m$ and so on,

$$\{f(\lambda I+U)\}^m = f^m I + m f^{m-1} f' U + \ldots + \frac{d^{k-1}}{d\lambda^{k-1}}(f^m)\frac{U^{k-1}}{(k-1)!}.$$

11.21. *Scalar power series.* We first establish those results about scalar power series that we shall need in dealing with matrix power series and begin by quoting a standard theorem.

THEOREM 26. *Let*

$$g(y) = \sum_{m=0}^{\infty} b_m y^m,$$

with radius of convergence ρ. *Then, provided that the series* $\sum |a_n \lambda^n|$ *is convergent and has a sum* s *less than* ρ, $g\{f(\lambda)\}$ *may be expanded as a power series in* λ *by means of the steps*

$$g\{f(\lambda)\} = \sum_{m=0}^{\infty} b_m \Big(\sum_{n=0}^{\infty} a_n \lambda^n\Big)^m \tag{1a}$$

$$= \sum_{m=0}^{\infty} b_m \sum_{n=0}^{\infty} a_{n,m} \lambda^n$$

$$= \sum_{n=0}^{\infty} \Big(\sum_{m=0}^{\infty} b_m a_{n,m}\Big)\lambda^n$$

$$= \sum_{n=0}^{\infty} c_n \lambda^n, \tag{1b}$$

say. The steps remain valid when b_m, a_n *are replaced by their absolute values;* a fortiori, *the series* $\sum |c_n \lambda^n|$ *is convergent.*†

COROLLARY I. *The series* $\sum m b_m y^{m-1}$, $\sum m(m-1) b_m y^{m-2}$,... *have the same radius of convergence as the series* $\sum b_m y^m$. *Accordingly, under the conditions required for the theorem we may also obtain* $\sum m b_m (\sum a_n \lambda^n)^{m-1}$, $\sum m(m-1) b_m (\sum a_n \lambda^n)^{m-2}$,... *as power series in* λ *by expanding and collecting like powers of* λ. *The resulting series are*

$$\sum_{n=0}^{\infty} \Big(\sum_{m=0}^{\infty} m b_m a_{n,m-1}\Big)\lambda^n, \qquad \sum_{n=0}^{\infty} \Big(\sum_{m=0}^{\infty} m(m-1) b_m a_{n,m-2}\Big)\lambda^n,$$

and so on.

† Cf. K. Knopp, *Theory and Application of Infinite Series* (Blackie and Son, London, 1928), p. 180.

There is a further small refinement of Theorem 26 that we need to formulate explicitly. This is

COROLLARY 2. *If* $|\lambda| < R$ *and* $\sum |a_n \lambda^n| < \rho$, *then* λ *lies* WITHIN *the circle of convergence of the series* $\sum c_n z^n$.

PROOF. Let R_1 lie between $|\lambda|$ and R. Then the sum of $\sum |a_n z^n|$ is continuous and monotonic increasing in

$$0 \leqslant |z| \leqslant R_1 < R;$$

we can, therefore, choose a λ_1 such that $|\lambda| < |\lambda_1| < R_1$ and $\sum |a_n \lambda_1^n| < \rho$. By the theorem, $\sum c_n \lambda_1^n$ is absolutely convergent. Therefore, since $|\lambda| < |\lambda_1|$, λ lies *within* (and not on the boundary) of the circle of convergence of $\sum c_n z^n$.

11.22. *Application of the theorem.* We now suppose that $|\lambda| < R$ (to enable us to use Corollary 2) and $\sum |a_n \lambda^n| < \rho$. Then, by the theorem itself,

$$\sum_{m=0}^{\infty} b_m \{f(\lambda)\}^m = \sum_{n=0}^{\infty} c_n \lambda^n.$$

By differentiating this result,

$$\left\{\sum_{m=0}^{\infty} m b_m \{f(\lambda)\}^{m-1}\right\} f'(\lambda) = \sum_{n=0}^{\infty} n c_n \lambda^{n-1}, \tag{2}$$

the term-by-term differentiation being justified because

$$|f(\lambda)| \leqslant \sum |a_n \lambda^n| < \rho$$

and $|\lambda|$ is less than the radius of convergence of $\sum c_n z^n$.

By Corollary 1, the above gives

$$\left\{\sum_{n=0}^{\infty}\left(\sum_{m=0}^{\infty} m b_m a_{n,m-1}\right)\lambda^n\right\}\left\{\sum_{n=0}^{\infty} n a_n \lambda^{n-1}\right\} = \sum_{n=0}^{\infty} n c_n \lambda^{n-1}, \tag{3}$$

the product of the two power series in λ being formed as in Theorem 25. Similarly, we may prove, by differentiating (2) and using Corollary 1,

$$\left[\sum_{n=0}^{\infty}\left(\sum_{m=0}^{\infty} m b_m a_{n,m-1}\right)\lambda^n\right]\left[\sum_{n=0}^{\infty} n(n-1) a_n \lambda^{n-2}\right] + \\ + \left[\sum_{n=0}^{\infty}\left(\sum_{m=0}^{\infty} m(m-1) b_m a_{n,m-2}\right)\lambda^n\right]\left[\sum_{n=0}^{\infty} n a_n \lambda^{n-1}\right]^2 \\ = \sum_{n=0}^{\infty} n(n-1) c_n \lambda^{n-2};$$

and so on. Thus the 'expanded' form of†

$$\frac{d^r}{d\lambda^r}\Big\{\sum_{m=0}^{\infty} b_m f^m\Big\} = \sum_{n=0}^{\infty} n(n-1)...(n-r+1)c_n \lambda^{n-r} \quad (r = 1, 2,...).$$

11.23. *The classical canonical matrix in a power series.* Let $C(\lambda)$ be a classical submatrix of order k, so that

$$C(\lambda) = \lambda I + U,$$

where I, U are the unit, auxiliary unit matrices of order k. As our preliminary work showed, when $|\lambda| < R$,

$$\{f(\lambda I+U)\}^m = f^m I + mf^{m-1}f'U + ... + \frac{d^{k-1}}{d\lambda^{k-1}}(f^m)\frac{U^{k-1}}{(k-1)!}.$$

Now let $|\lambda| < R$ and $\sum |a_n \lambda^n| < \rho$. Then

$$\sum_{m=0}^{\infty} b_m \Big(\sum_{n=0}^{\infty} a_n C^n\Big)^m = \sum_{m=0}^{\infty} b_m \{f(\lambda I+U)\}^m$$

$$= \sum_{m=0}^{\infty} b_m \Big\{f^m I + mf^{m-1}f'U + ... + \frac{d^{k-1}}{d\lambda^{k-1}}(f^m)\frac{U^{k-1}}{(k-1)!}\Big\}$$

$$= \Big(\sum_{m=0}^{\infty} b_m f^m\Big)I + \Big(\sum_{m=0}^{\infty} mb_m f^{m-1}f'\Big)U + ...$$

By the results of § 11.22, this gives

$$\sum_{m=0}^{\infty} b_m \Big(\sum_{n=0}^{\infty} a_n C^n\Big)^m$$

$$= \sum_{n=0}^{\infty} c_n \Big\{\lambda^n I + n\lambda^{n-1}U + ... + \frac{n(n-1)...(n-k+2)}{(k-1)!}\lambda^{n-k+1}U^{k-1}\Big\}$$

and, on using the notation

$$\psi(z) = g\{f(z)\} = \sum c_n z^n,$$

this gives

$$\sum_{m=0}^{\infty} b_m \Big(\sum_{n=0}^{\infty} a_n C^n\Big)^m = \psi(\lambda)I + \psi'(\lambda)U + ... + \frac{\psi^{(k-1)}(\lambda)}{(k-1)!}U^{k-1} = \psi(C), \tag{4}$$

on recalling § 5.1.

We have thus proved that when

$$g\{f(z)\} = \psi(z)$$

† As in § 11.2, f^m denotes $\{f(\lambda)\}^m$.

and $C = \lambda I+U$ is a classical submatrix for which $|\lambda| < R$ and $\sum |a_n \lambda^n| < \rho$,
$$g\{f(C)\} = \psi(C). \tag{5}$$

The extension of this result to a matrix
$$C = \text{diag}\{C_k(\lambda),...\}$$
is immediate: for
$$C^n = \text{diag}[\{C_k(\lambda)\}^n,...],$$
$$\sum_{n=0}^{\infty} a_n C^n = \text{diag}\Big[\sum_{n=0}^{\infty} a_n\{C_k(\lambda)\}^n,...\Big],$$
$$\psi(C) = \text{diag}[\psi\{C_k(\lambda)\},...],$$
and the steps required for (4) can be carried out for each submatrix separately.†

Hence (5) holds when C is any classical canonical form provided that each latent root λ satisfies the two conditions
$$|\lambda| < R, \qquad \sum |a_n \lambda^n| < \rho.$$

11.24. *The general matrix in a power series.* Let A be a given square matrix, C its classical canonical form, and let
$$A = TCT^{-1}.$$
As before, let
$$f(z) = \sum a_n z^n \quad \text{converge when} \quad |z| < R,$$
$$g(z) = \sum b_n z^n \quad \text{converge when} \quad |z| < \rho.$$
Let
$$\phi(z) = \sum_{m=0}^{\infty} b_m\Big(\sum_{n=0}^{\infty} a_n z^n\Big)^m \tag{6}$$
and
$$\psi(z) = \sum c_n z^n,$$
the series obtained by expanding the terms of (6) and rearranging as a single power series in z.

† Perhaps the simplest process is this: let C be of order N and let (C_k) be the matrix of order N that consists of $C_k(\lambda)$ in its proper position and of zeros everywhere else. Then
$$C = (C_k)+(C_p)+(C_q)+..., \quad \text{say } M \text{ elements}.$$
Each product $(C_p)(C_q)$, with $p \neq q$, is the null matrix and the work of (4) is easily thought of as M calculations each of the type
$$\sum b_m\{\sum a_n(C_k)^n\}^m = \psi\{(C_k)\}.$$

Then, in succession, on using Theorem 20,

$$A = TCT^{-1}, \quad f(A) = Tf(C)T^{-1}, \quad \{f(A)\}^m = T\{f(C)\}^m T^{-1},$$

$$\phi(A) = \sum_{m=0}^{\infty} b_m\{f(A)\}^m = \sum_{m=0}^{\infty} b_m T\{f(C)\}^m T^{-1}$$

$$= T\Big[\sum_{m=0}^{\infty} b_m\{f(C)\}^m\Big]T^{-1} = T\phi(C)T^{-1},$$

and $$\psi(A) = T\psi(C)T^{-1},$$

the steps being subject to the sole condition that the infinite series which occur are convergent.

We have proved in (5) that

$$\phi(C) = \psi(C)$$

whenever each latent root λ satisfies the two conditions $|\lambda| < R$, $\sum |a_n \lambda^n| < \rho$. Hence, under the same conditions

$$\phi(A) = \psi(A).$$

We sum up our results in a formal theorem.

THEOREM 27. *Let*

$$f(z) = \sum a_n z^n \quad \textit{converge when} \quad |z| < R,$$

$$g(z) = \sum b_n z^n \quad \textit{converge when} \quad |z| < \rho,$$

and let A be a square matrix for which each latent root λ satisfies the two conditions

$$|\lambda| < R, \qquad \sum |a_n \lambda^n| < \rho. \tag{7}$$

Then $$\sum_{m=0}^{\infty} b_m\{f(A)\}^m$$

can be expanded in a power series in A by the steps

$$\sum_{m=0}^{\infty} b_m\Big(\sum_{n=0}^{\infty} a_n A^n\Big)^m = \sum_{m=0}^{\infty} b_m \sum_{n=0}^{\infty} a_{n,m} A^n = \sum_{n=0}^{\infty}\Big(\sum_{m=0}^{\infty} b_m a_{n,m}\Big)A^n.$$

In brief, under conditions (7), we may substitute $y = f(A)$ in $g(y)$ and rearrange the result as a power series in A.

11.3. *Power series whose sum is the null matrix*

Suppose that $$\sum_{n=0}^{\infty} a_n A^n$$

is convergent and that its sum is the null matrix for all matrices A whose latent roots are sufficiently small in modulus. Then (cf. § 5)

$$\sum_{n=0}^{\infty} a_n \lambda^n$$

is convergent when $|\lambda|$ is sufficiently small and its sum is zero. By a well-known result in the theory of functions†

$$a_n = 0 \quad (n = 0, 1, 2,...).$$

It follows that, if

$$\sum_{n=0}^{\infty} a_n A^n = \sum_{n=0}^{\infty} b_n A^n$$

for all matrices A whose latent roots are sufficiently small in modulus, then

$$a_n = b_n \quad (n = 0, 1, 2,...)$$

and we may write $\sum a_n A^n = \sum b_n A^n$ provided only that the series are convergent.

11.4. *The reciprocal of a power series*

When $a_0 \neq 0$ and $|z|$ is sufficiently small,

$$(a_0+a_1 z+a_2 z^2+...)^{-1} = a_0^{-1} \sum_{m=0}^{\infty} \left(-\frac{a_1 z+a_2 z^2+...}{a_0}\right)^m$$

and this series may, in virtue of Theorem 26, be expanded as a power series in z, say

$$\sum_{n=0}^{\infty} c_n z^n. \tag{8}$$

The existence of the convergent expansion (8) being thus established when $|z|$ is sufficiently small, we most easily determine c_n from the fact that

$$(\sum a_n z^n)(\sum c_n z^n) \equiv 1 \tag{9}$$

when $|z|$ is sufficiently small. This gives (by § 11.3)

$$a_0 c_0 = 1, \quad a_0 c_1 + a_1 c_0 = 0, \quad .... \tag{10}$$

† Or see W. L. Ferrar, *A Text-book of Convergence* (Oxford, 1938), p. 115, Theorem 47.

Thus, when $a_0 \neq 0$, there is a power series $\sum c_n z^n$, with coefficients given by (10), for which

$$(\sum a_n z^n)(\sum c_n z^n) = 1$$

whenever z lies within the circles of convergence of the two series (for this is the sole condition needed to justify forming the product of the two series as a single power series).

By § 11.1 this carries over at once to the reciprocal of a matrix power series. When $a_0 \neq 0$ and

$$f(A) = a_0 I + a_1 A + a_2 A^2 + \dots,$$

there is a set of coefficients c_0, c_1,..., given by (10), for which

$$c_0 I + c_1 A + c_2 A^2 + \dots = \{f(A)\}^{-1}$$

provided only that each latent root of A lies within the circles of convergence of $\sum a_n z^n$ and $\sum c_n z^n$.

11.5. *The reversion of a power series*

Let $a_1 \neq 0$ and let

$$y = a_0 + a_1 x + a_2 x^2 + \dots$$

be convergent when $|x| < R$. Then there is one and only one analytic function x of the variable y which tends to zero as y tends to a_0; when $y - a_0$ is sufficiently small, this value of x may be expressed as a power series in $y - a_0$, namely

$$x = b_1(y-a_0) + b_2(y-a_0)^2 + \dots, \tag{11}$$

where the b's are determined from the a's by equating coefficients of $(y-a_0)^k$ in the identity

$$y - a_0 \equiv \sum_{m=1}^{\infty} a_m \Big\{\sum_{n=1}^{\infty} b_n (y-a_0)^n\Big\}^m; \tag{12}$$

that is to say, the b's are determined by

$$\left.\begin{array}{l} 1 = a_1 b_1, \quad 0 = a_1 b_2 + a_2 b_1^2, \\ \quad 0 = a_1 b_3 + 2a_2 b_1 b_2 + a_3 b_1^3, \\ \dots\dots\dots\dots \end{array}\right\}. \tag{13}$$

The sole difficulty of the matter, as a problem in scalar power series, is to prove that there is a number η such that the formal work is logically justified when $|y-a_0| < \eta$. This difficulty is fully dealt with in the literature of scalar power series.†

Suppose now that X is a square matrix and that y is given by

$$y-a_0 I = a_1 X+a_2 X^2+\dots.$$

For convenience, denote $y-a_0 I$ by Y, so that

$$Y = a_1 X+a_2 X^2+\dots \quad (a_1 \neq 0).$$

When the latent roots of Y are sufficiently small (which amounts to saying 'when the latent roots of X are sufficiently small'—for when λ is a latent root of X, the corresponding latent root of Y is $a_1\lambda+a_2\lambda^2+\dots$), there is a power series

$$b_1 Y+b_2 Y^2+\dots,$$

with the b's defined by (13), such that

$$Y = \sum_{m=1}^{\infty} a_m\Big(\sum_{n=1}^{\infty} b_n Y^n\Big)^m. \tag{14}$$

The matrix result (14) follows from the scalar result (12) in virtue of Theorem 27, the conditions required by that theorem being certainly satisfied when the latent roots of Y are small enough.

Now let R_1 be the radius of convergence of $\sum b_n z^n$ and R_2 the radius of convergence of $\sum a_n z^n$. Let λ be a typical latent root of X and let $f(\lambda) = a_1\lambda+a_2\lambda^2+\dots$. Then $f(\lambda)$ is a typical latent root of Y and when, for each latent root,

$$|f(\lambda)| < R_2, \qquad \sum |b_n\{f(\lambda)\}^n| < R_1, \tag{15}$$

the series on the right-hand side of (14) can be expanded as a power series in Y and, by the relations (13), this power series reduces to the single term Y.

Hence, if

$$Y = a_1 X+a_2 X^2+\dots, \tag{16}$$

the b's are defined by the relations (13), and the conditions (15) are satisfied, then

$$Y = \sum_{m=1}^{\infty} a_m\Big(\sum_{n=1}^{\infty} b_n Y^n\Big)^m;$$

† See, for example, K Knopp, loc. cit., p. 184, or T. J. I'A. Bromwich, *An Introduction to the Theory of Infinite Series* (Macmillan, London, 1908), p. 138.

that is,
$$X = \sum_{n=1}^{\infty} b_n Y^n \tag{17}$$
satisfies (16).

NOTE 1. In general the b's are difficult to calculate explicitly and one must be content with the fact that the result is true when the latent roots of X are sufficiently small. Bromwich, loc. cit., p. 140, gives an explicit formula for a number μ which is less than or equal to the radius of convergence of $\sum b_n y^n$.

NOTE 2. Supposing Y to be given, the matrix X defined by (17) is not the only one to satisfy (16): it is that particular X which tends to the null matrix when Y tends to the null matrix; see, for example, Bromwich, loc. cit., p. 142, Example 1, where $y = x - ax^2$ and the x corresponding to the reversion process is

$$x = y + ay^2 + 2a^2y^3 + \dots,$$

or that root of the quadratic equation in x which is zero when y is zero; the other root is a^{-1} when $y = 0$.

NOTE 3. Any process valid for scalar power series will serve to determine the b's in terms of the a's: for example, Lagrange's theorem† shows that, under appropriate conditions, when

$$x = a + y(\sum b_m x^m)$$

or, what is the same thing,

$$y = (x-a)(\sum b_m x^m)^{-1} = \sum p_n x^n$$

the appropriate series for x in powers of y is

$$x = a + \sum_{n=1}^{\infty} \frac{y^n}{n!} \frac{d^{n-1}}{da^{n-1}} \{\sum b_m a^m\}^n.$$

In order to transcribe a scalar theorem into the appropriate matrix form what we have to ensure is that Theorem 27, or a theorem of that nature, will cover the transcription from scalars to matrices.

12. The differentiation of functions of matrices

We suppose $f(z)$, $g(z)$ to be functions of z analytic in the neighbourhood of $z = 0$ and we take their expansions to be

$$f(z) = \sum a_n z^n, \qquad g(z) = \sum b_n z^n.$$

We suppose that A is a square matrix and, whenever convenient, that its latent roots are sufficiently small in modulus to enable us to apply the 'algebra of power series' as developed in § 11. We shall see, in § 14, how this limitation on the latent

† See, for example, Whittaker and Watson, *Modern Analysis* (Cambridge, 1920), p. 133.

roots of A can be removed and the results freed of these conditions, which are imposed solely by the method of proof.

12.1. *Sums and products*

From the definition of § 10.2,

$$\frac{d}{dA}\{f(A)+g(A)\} = \frac{df}{dA}+\frac{dg}{dA}. \tag{1}$$

Again,

$$f(A+hI)g(A+hI)-f(A)g(A)$$
$$= f(A+hI)\{g(A+hI)-g(A)\}+\{f(A+hI)-f(A)\}g(A),$$

and it follows, with a little care over the details of the limiting processes (see Note in § 12.2), that

$$\frac{d}{dA}\{f(A)g(A)\} = f(A)\frac{dg}{dA}+\frac{df}{dA}g(A). \tag{2}$$

But (§ 8.7) any two functions of A are commutative and we can write the factors of the products that occur in (2) in any order we please: we can thus deduce Leibnitz's formula

$$D^n\{f.g\} = D^nf.g+nD^{n-1}f.Dg+...+f.D^ng, \tag{3}$$

where f, g are functions of the matrix A and D^r denotes the operation of differentiating r times with respect to A.

Again, when $a_0 \neq 0$, § 11.4 shows that there is a matrix $\{f(A)\}^{-1}$ given, when the latent roots of A are sufficiently small, by a convergent power series $\sum c_n A^n$, in which the c_n are obtained from the identity

$$I \equiv (\sum a_n A^n)(\sum c_n A^n).$$

On differentiating with respect to A, we get

$$0 \equiv f'(A)\{f(A)\}^{-1}+f(A)\frac{d}{dA}\{f(A)\}^{-1};$$

whence
$$\frac{d}{dA}\{f(A)\}^{-1} = -f'(A)/\{f(A)\}^2. \tag{4}$$

12.2. Note. There is no difficulty, merely tiresome detail, in extending to limits of matrices the well-known elementary theorems concerning limits of scalars. For example,

if $A_n \to A$ and $B_n \to B$, then $A_n B_n \to AB$.

Proof. Each element of $A_n \to$ the corresponding element of A and so

$$\epsilon, k > 0; \quad \exists N \text{ . when } n \geqslant N,$$

$$A = A_n + \epsilon k \Omega_n, \qquad B = B_n + \epsilon k O_n,$$

where no element of Ω_n or O_n exceeds unity in absolute value. Thus, when $n \geqslant N$,

$$A_n B_n = (A - \epsilon k \Omega_n)(B - \epsilon k O_n)$$
$$= AB - \epsilon k(\Omega_n B + A O_n) + \epsilon^2 k^2 \Omega_n O_n.$$

Now there is a finite number K such that no element of A or B exceeds K in absolute value. Supposing A and B to be square matrices of m rows and columns, the absolute value of an element of $\Omega_n B$ or $A O_n$ cannot exceed mK and the absolute value of an element of $\Omega_n O_n$ cannot exceed m. Thus, if we choose k to begin with so that

$$2kKm < 1, \qquad k^2 m < 1,$$

we have, when $n \geqslant N$,

$$A_n B_n = AB - \epsilon C_n + \epsilon^2 D_n,$$

where C_n and D_n are matrices whose elements do not exceed unity in absolute value. Hence

$$A_n B_n \to AB.$$

12.3. *The functions $f(tA)$, $f(A+tI)$*

We prove that, when t is a scalar constant,

(i) $$\frac{d}{dA} f(tA) = t f'(tA),$$

(ii) $$\frac{d}{dA} f(A+tI) = f'(A+tI).$$

Proof: (i)

$$\frac{d}{dA} f(tA) = \frac{d}{dA} \sum_{n=0}^{\infty} a_n t^n A^n = t \sum_{n=1}^{\infty} n a_n (tA)^{n-1} = t f'(tA).$$

(ii) Let $A = TCT^{-1}$, where C is a classical canonical form. Then [§ 10.2 (6)]

$$\frac{d}{dA} f(A+tI) = T \frac{d}{dC} f(C+tI) T^{-1}$$

$$= T \operatorname{diag}\left\{\frac{dF_k(\lambda+t)}{d\lambda}, \ldots\right\} T^{-1},$$

where†

$$F_k(\lambda+t) = \begin{bmatrix} f(\lambda+t) & f'(\lambda+t) & . & . & . \\ 0 & f(\lambda+t) & . & . & . \\ . & . & . & . & . \\ 0 & 0 & . & . & f(\lambda+t) \end{bmatrix}.$$

Thus a typical submatrix of $df(C+tI)/dC$ is

$$\begin{bmatrix} f'(\lambda+t) & f''(\lambda+t) & . & . & . \\ 0 & f'(\lambda+t) & . & . & . \\ . & . & . & . & . \end{bmatrix}$$

and hence $df(C+tI)/dC$ is given in the form of a power series by

$$\sum_{n=1}^{\infty} na_n(C+tI)^{n-1}.$$

Hence

$$\frac{d}{dA}f(A+tI) = T\sum_{n=1}^{\infty} na_n(C+tI)^{n-1}T^{-1}$$

$$= \sum_{n=1}^{\infty} na_n(A+tI)^{n-1}.$$

12.4. *Function of a function*

We shall prove that, when Z is a function of Y and Y is a a function of X,

$$\frac{dZ}{dX} = \frac{dZ}{dY}\frac{dY}{dX} \tag{5}$$

provided that the latent roots of the matrix X are of sufficiently small modulus (but see § 14 for the removal of this proviso).

We begin by considering the scalar result that corresponds to (5). Let

$$y = f(x), \qquad z = g(y),$$

where the analytic functions $f(x)$ and $g(y)$ have the power series expansions

$$f(x) = \sum_{n=0}^{\infty} a_n x^n \qquad \text{when} \quad |x| < R,$$

$$g(y) = \sum_{m=0}^{\infty} b_m(y-a_0)^m \qquad \text{when} \quad |y-a_0| < \rho.$$

Let

$$|x| < R, \qquad \sum_{n=1}^{\infty} |a_n x^n| < \rho, \tag{6}$$

conditions that can certainly be satisfied by taking $|x|$ small enough.

† See § 10.2 (5).

Then, as in § 11.21,
$$z = b_0 + \sum_{n=1}^{\infty} c_n x^n,$$
where†
$$c_n = \sum_{m=0}^{\infty} b_m a_{n,m};$$
moreover x lies *within* the circle of convergence of $\sum c_n x^n$. Hence
$$\frac{dz}{dx} = \sum_{n=0}^{\infty} nc_n x^{n-1}. \tag{7}$$
Further, $|y-a_0| \leqslant \sum_{n=1}^{\infty} |a_n x^n| < \rho$, so that
$$\frac{dz}{dy} = \sum_{m=0}^{\infty} mb_m (y-a_0)^{m-1} = \sum_{n=0}^{\infty} \beta_n x^n, \tag{8}$$
where
$$\beta_n = \sum_{m=0}^{\infty} mb_m a_{n,m-1}.$$
Finally,
$$\frac{dy}{dx} = \sum_{n=0}^{\infty} na_n x^{n-1}. \tag{9}$$

The conditions (6) ensure that the functions are analytic and so
$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx}.$$
Hence, for all x that satisfy (6),
$$\sum_{n=0}^{\infty} nc_n x^{n-1} \equiv \Big\{\sum_{n=0}^{\infty} \beta_n x^n\Big\}\Big\{\sum_{n=0}^{\infty} na_n x^{n-1}\Big\}$$
and, on equating coefficients,
$$nc_n = a_1\beta_{n-1} + 2a_2\beta_{n-2} + \dots + na_n\beta_0. \tag{10}$$

Now let X, Y, Z be matrices and let
$$Y = f(X), \qquad Z = g(Y),$$
the functional relation being defined (independently of its expression via power series) as in § 8.2. Let each latent root λ of X satisfy the conditions
$$|\lambda| < R, \qquad \sum_{n=1}^{\infty} |a_n \lambda^n| < \rho. \tag{11}$$

† $a_{n,m}$ is coefficient of x^n in $(a_1 x + a_2 x^2 + \dots)^m$; $a_{n,0} = 0$ when $n \geqslant 1$.

We may employ the power series representations

$$Y = \sum_{n=0}^{\infty} a_n X^n,$$

$$Z = \sum_{m=0}^{\infty} b_m (Y - a_0 I)^m,$$

and, by Theorem 27, substitute for $Y - a_0 I$ in terms of X to obtain

$$Z = \sum_{m=0}^{\infty} b_m (Y - a_0 I)^m = \sum_{n=0}^{\infty} c_n X^n.$$

Now, by § 12.3 (ii),

$$\frac{dZ}{dY} = \sum_{m=0}^{\infty} m b_m (Y - a_0 I)^{m-1}$$

and so, in the notation of (8) above,

$$\frac{dZ}{dY} = \sum_{n=0}^{\infty} \beta_n X^n.$$

Also (Theorem 26, Corollary 2) each latent root of X lies within the circle of convergence of $\sum c_n x^n$; hence

$$\frac{dZ}{dX} = \sum_{n=0}^{\infty} n c_n X^{n-1}. \tag{12}$$

Hence, under the conditions (11),

$$\frac{dZ}{dY}\frac{dY}{dX} = \Big(\sum_{n=0}^{\infty} \beta_n X^n\Big)\Big(\sum_{n=0}^{\infty} n a_n X^{n-1}\Big)$$

and we may multiply these series, in virtue of Theorem 25, to obtain

$$\sum_{n=0}^{\infty} (a_1 \beta_{n-1} + 2a_2 \beta_{n-2} + \ldots + n a_n \beta_0) X^{n-1}.$$

On using (10), this gives

$$\frac{dZ}{dY}\frac{dY}{dX} = \sum_{n=0}^{\infty} n c_n X^{n-1} = \frac{dZ}{dX}.$$

This proves (5) subject to the latent roots of X being small enough to ensure that the conditions (11) are satisfied.†

† I tried to prove (5) without resort to the algebra of power series and failed.

13. Taylor's expansion

13.1. Let C be a single classical submatrix of order k, and let $f(z)$ be a function of z which is analytic in a neighbourhood of λ, the latent root of C. Then, provided $\lambda+h$ lies within this neighbourhood, we may define $f(C+hI)$ by the equation (cf. § 8.2†)

$$f(C+hI) = \begin{bmatrix} f(\lambda+h) & f'(\lambda+h) & . & . & f^{(k-1)}(\lambda+h)/(k-1)! \\ . & . & . & . & . \\ 0 & 0 & . & . & f(\lambda+h) \end{bmatrix}.$$

With the above proviso satisfied, we may expand each element of the matrix by Taylor's theorem, giving

$$f(\lambda+h) = f(\lambda)+hf'(\lambda)+...,$$
$$f'(\lambda+h) = f'(\lambda)+hf''(\lambda)+...,$$

and so on. Thus the matrix written above is equal to

$$f(C)+hf'(C)+.... \tag{1}$$

When C is a general canonical form, say

$$C = \text{diag}\{C_k(\lambda),...\},$$

we may apply the argument to each $C_k(\lambda)$ separately, *provided that, for each* λ, *the point* $\lambda+h$ *lies in the neighbourhood of* λ *within which* $f(z)$ *is analytic.* Under these conditions

$$f(C+hI) = f(C)+hf'(C)+... \tag{2}$$

and, by using $A = TCT^{-1}$ as on previous occasions, the result holds for any square matrix A; the limitations on the point $\lambda+h$ for each latent root λ of A must, of course, be satisfied.

13.2. A more general result has been proved by H. B. Phillips.‡ He has proved that, when B is commutative with A,

$$f(A+B) = f(A)+Bf'(A)+.... \tag{3}$$

The result is subject to the condition that each latent root of $A+B$ lies in that neighbourhood of the 'corresponding' latent root of A within which $f(z)$ is analytic.

† The limitation that $f(z)$ be analytic near $z = 0$ is not essential in the present section.

‡ *Amer. J. of Math.* **41** (1919), 266–78.

The methods of the present chapter do not seem to be sufficient to prove this more general result.

14. On the identity of two matrix functions

14.1. We confine ourselves, to begin with, to functions of a canonical form C and, in following out details, to functions of a single classical canonical submatrix $C_k(\lambda)$. Suppose we have two functions of z, analytic in the neighbourhood of the origin, say $\mathsf{F}(z)$ and $\mathsf{G}(z)$. The functions $\mathsf{F}\{C_k(\lambda)\}$ and $\mathsf{G}\{C_k(\lambda)\}$ are then defined by

$$\mathsf{F}\{C_k(\lambda)\} = \begin{bmatrix} \mathsf{F}(\lambda) & \mathsf{F}'(\lambda) & . & . & \mathsf{F}^{(k-1)}(\lambda)/(k-1)! \\ 0 & \mathsf{F}(\lambda) & . & . & \mathsf{F}^{(k-2)}(\lambda)/(k-2)! \\ . & . & . & . & . \\ 0 & 0 & . & . & \mathsf{F}(\lambda) \end{bmatrix}, \qquad (1)$$

with a similar definition for $\mathsf{G}\{C_k(\lambda)\}$. The definition is valid provided only that λ is not a singularity of the function $\mathsf{F}(z)$ or $\mathsf{G}(z)$.

Near $z = 0$ the functions $\mathsf{F}(z)$ and $\mathsf{G}(z)$ can be represented by the sums of convergent power series: let these sums be denoted by $f(z)$ and $g(z)$, say

$$f(z) = \sum_{n=0}^{\infty} a_n z^n, \qquad g(z) = \sum_{n=0}^{\infty} b_n z^n.$$

When $|z|$ is sufficiently small, $\mathsf{F} \equiv f$ and $\mathsf{G} \equiv g$.

Now suppose that we have proved, e.g. the work of §§ 11 and 12, that

$$f\{C_k(\lambda)\} \equiv g\{C_k(\lambda)\},$$

when λ is sufficiently small. Then it follows† from (1) that

$$f^{(r)}(\lambda) \equiv g^{(r)}(\lambda) \quad (r = 0, 1,..., k-1)$$

when λ is sufficiently small, i.e. that the function $\mathsf{F}(z)$ and its first $k-1$ derivatives are identically equal to $\mathsf{G}(z)$ and its first $k-1$ derivatives when $|z|$ is sufficiently small. Since we are dealing with analytic functions of a complex variable z, this identity is preserved throughout the whole region common to the definitions of $\mathsf{F}(z)$ and $\mathsf{G}(z)$ as analytic functions. Thus, when λ lies in such a region,

$$\mathsf{F}^{(r)}(\lambda) \equiv \mathsf{G}^{(r)}(\lambda) \quad (r = 0, 1,..., k-1)$$

and, by (1),

$$\mathsf{F}\{C_k(\lambda)\} \equiv \mathsf{G}\{C_k(\lambda)\}.$$

† f and F, g and G are interchangeable when $|z|$ is small enough.

The argument extends to functions of any classical canonical matrix
$$C = \operatorname{diag}\{C_k(\lambda),...\}$$
and, by the usual $A = TCT^{-1}$ (cf. § 11.24), to any matrix A.

14.2. The arguments of § 11.3 and of § 14.1 are useful in extending the domains within which matrix identities can be proved valid. For example, consider § 12.1 (4) with
$$f(z) = 1+z+z^2.$$
What we there proved was that
$$\frac{d}{dA}\frac{1}{I+A+A^2} = -\frac{I+2A}{(I+A+A^2)^2}, \tag{2}$$
provided that the latent roots of A were all sufficiently small.

The result was essentially one about the power series representing the functions and the smallness of the latent roots was a condition needed by the algebra of the power series. The work of § 11.3 enables us to see at a glance that, regardless of the requirements of the steps of the original proof, the final result (2) will hold *as an identity of power series* provided the latent roots of A lie within the circle $|z| = 1$; for, subject to this condition, the two sides of (2) are, since
$$\frac{1}{I+A+A^2} = \frac{I-A}{I-A^3} = I-A+A^3-A^4+...,$$
$$\text{(i)} \quad -I+3A^2-4A^3+...,$$
$$\text{(ii)} \quad -(I+2A)(I-A+A^3-A^4+...)^2,$$
and the two sides are known to be equal when the latent roots λ are small enough. By § 11.3, the result (2) holds as an identity of power series provided only that the series converge.

We see further, by § 14.1, that (2) will hold *as an identity of functions* provided only that no latent root of A is an imaginary cube root of unity, for these two roots are the only singularities of the function $(1+z+z^2)^{-2}$. In this form there is no reference to infinite power series, and the fact proved by (2) is that
$$\lim_{h\to 0}\left\{\frac{1}{I+(A+hI)+(A+hI)^2}-\frac{1}{I+A+A^2}\right\}h^{-1}$$
$$= -(I+2A)(I+A+A^2)^{-2}.$$

15. Notes on functions of a matrix

15.1. *Maclaurin expansions*

We have, for the most part, confined our attention to functions $f(z)$ regular near $z = 0$ and to power series $\sum a_n z^n$. Such a limitation, though convenient, is not essential; what is essential to the whole argument is that the function $f(z)$ considered shall be analytic in a domain of the z-plane and that in the neighbourhood of some point z_0

$$f(z) = \sum a_n(z-z_0)^n. \tag{1}$$

The fact that we have based our arguments on Maclaurin expansions $\sum a_n z^n$ sometimes requires us to make minor modifications that would not be needed if we had used the more general expansion (1); for example, when

$$f(z) = z^m(a_0+a_1 z+...) = z^m g(z),$$

say, the work of § 12.1 cannot prove directly that

$$\frac{d}{dA}\{f(A)\}^{-1} = -f'(A)/\{f(A)\}^2. \tag{2}$$

This formula is, nevertheless, easily established; with $a_0 \neq 0$, § 12.1 shows that (2) holds when g replaces f and we have merely to differentiate $A^{-m}\{g(A)\}^{-1}$ by the product rule in order to prove (2) as it stands.

15.2. *Definitions of 'functions of a matrix'*

The definition via a power series $\sum a_n A^n$, given in § 8.1, is a commonplace of the literature and an account of it is given in many books†; the definition via the function $f(z)$, given in § 8.2, is not so well known. I developed it as a natural sequel to the series definition; the only reference I can find in the literature is a paper by M. Cipolla;‡ even such direct consequences of the definition as those of §§ 8.4 and 8.7 do not appear to have been published before.

† e.g. Turnbull and Aitken, *Canonical Matrices*, pp. 62, 74, 75.
‡ *Rend. Circ. mat. Palermo*, **56** (1931), 144–54.

15.3. *Dirac's definition of function*

In his *Principles of Quantum Mechanics*† Dirac has developed what, from the point of view of matrices, is tantamount to a general definition of 'function of a matrix' in terms of linear operators and observables. 'If two observables ξ and g are such that any linear operator that commutes with ξ also commutes with g, then g is a function of ξ.'

Turnbull and Aitken‡ have shown that a definition of matrix function

> '*If matrices Y and X are such that any matrix P for which $PX = XP$ also satisfies $PY = YP$, then Y is a function of X*'

leads to the theorem that Y is a polynomial in X. They show that when X has a given canonical form, Y has a definite number of degrees of freedom; they also find a general formula for Y.

This approach to the idea of function is essentially different from that of the present chapter.

15.4. *A matrix power series expressed as a polynomial*

As we showed in § 6, a convergent power series

$$a_0 I + a_1 A + a_2 A^2 + \dots \tag{3}$$

can be expressed as a polynomial

$$b_0 I + b_1 A + \dots + b_{m-1} A^{m-1}. \tag{4}$$

The particular character of the representation (4) is perhaps worth noting; the values of the coefficients $b_0, b_1, \dots$ depend not only on the coefficients $a_0, a_1, \dots$ but also on the elements of A itself, or rather on the minimum equation of A. If we replace A in (3) by a matrix D whose last invariant factor is different from that of A, we can indeed replace

$$a_0 I + a_1 D + a_2 D^2 + \dots$$

by a polynomial $\quad c_0 I + c_1 D + \dots + c_r D^r, \qquad (5)$

but the coefficients $c_0, c_1, \dots$ in (5) are not the same as the $b_0, b_1, \dots$ in (4).

† Oxford, Clarendon Press, 1935 (second edition): e.g. p. 57.

‡ *Canonical Matrices*, pp. 149, 150; in particular, Example 1, p. 150 gives the most general Y that is, by the above definition, a function of

$$X = \text{diag}\{C_3(\alpha), C_2(\alpha), C_3(\beta)\}.$$

As an instrument of investigation into the properties of a power series, the polynomial representation has obvious limitations.

15.5. *The differentiation of a matrix function*

So far as I can discover, no extensive use of the definition of § 10,

$$\frac{d}{dA}f(A) = \lim_{h\to 0}\frac{f(A+hI)-f(A)}{h},$$

occurs in the literature. The reference to Born and Jordan, given in § 10.2, was found only after I had completed most of the work of the present chapter and I have found no other. Various methods of 'differentiating' a matrix are given by C. C. MacDuffee in his chapter on Functions of Matrices.†

The developments from the above definition and the work on the algebra of matrix power series have not, I think, appeared previously in print.

† *The Theory of Matrices* (Chelsea Publishing Company, New York, 1946).

CHAPTER VI

CONGRUENCE

1. Introductory

1.1. In Chapter II we considered matrices A and B connected by a relation $B = RAS$, where R and S were non-singular matrices. In Chapter IV we studied the result of requiring R and S to be reciprocals: we now consider the result of requiring R and S to be transposes: that is, we consider matrices

$$H'AH$$

where H' is the transpose of H. We deal with the matter only in an elementary way and take A to be a symmetrical matrix whose elements are real or complex numbers.

The chief importance of the equivalence

$$B = H'AH \tag{1}$$

lies in the fact that when a quadratic form

$$\sum_{r=1}^{n} \sum_{s=1}^{n} a_{rs} x_r x_s \quad (a_{ij} = a_{ji}) \tag{2}$$

is transformed to variables $y_1,..., y_n$ by the transformation $x = Hy$, the new form is†

$$\sum_{r=1}^{n} \sum_{s=1}^{n} b_{rs} y_r y_s, \tag{3}$$

where the matrix $[b_{rs}]$ is given by (1). The first problem is so to choose H that B is a diagonal matrix: the form (3) then reduces to

$$b_{11} y_1^2 + b_{22} y_2^2 + ... + b_{nn} y_n^2.$$

We shall here treat the problem as one of matrix equivalence.‡

1.2. When the matrix B is related to the matrix A by a relation $B = H'AH$ we say that B IS CONGRUENT TO A. We note that the relation of congruence is

(i) *reflexive*, for $A = (H')^{-1}BH^{-1}$

$$= (H^{-1})'BH^{-1},$$

† F. 124.

‡ F. 148, Theorem 47, considers the quadratic form and not the matrix.

and (ii) *transitive*, for when $B = H'AH$ and $C = K'BK$,

$$C = K'H'AHK$$
$$= (HK)'A(HK).$$

In the sequel it will be important to recognize that when B is congruent to A, C to B,..., and finally L to M, then M is congruent to A.

1.3. We shall often use two particular types of matrix congruent to A:

(i) The matrix got from A by interchanging two rows and then interchanging two columns in the resulting matrix; this is, in the notation of Chapter II, § 1.3, the matrix

$$I'_{ij} A I_{ij}.$$

(ii) The matrix obtained by adding h times the ith column of A to its jth column, and then adding h times the ith row to the jth row in the resulting matrix; this is, in the notation of Chapter II, § 1.3, the matrix

$$(I+H_{ji})A(I+H_{ij}) = (I+H_{ij})'A(I+H_{ij}).$$

1.4. *Preliminary lemmas*

LEMMA 1. *Given a symmetrical matrix A, other than the null matrix, in which a_{11} is zero, there is a congruent matrix B in which b_{11} is not zero.*

PROOF. If any diagonal element, say a_{ii}, is not zero, the interchange of the first and ith rows followed by interchange of the first and ith columns gives a non-zero first element; the result is a matrix congruent to A and is the matrix B required.

If all a_{ii} are zero, then at least one a_{ij} $(i \neq j)$ is not zero. The matrix

$$C = (I+H_{ji})A(I+H_{ij}),$$

with $h = 1$ [add ith column to jth and then ith row to jth], has an element $a_{ij}+a_{ji} = 2a_{ij}$ at the jth place of the leading diagonal. We obtain B from C by interchanging the first and jth rows and then the first and jth columns.

NOTE. When we consider the wider domain of abstract algebra, it is possible for a field F to have 'CHARACTERISTIC p'; that is to say, for each element α of F

$$\alpha+\alpha+\ldots \text{ to } p \text{ terms}$$

is equal to zero.† Lemma 1, which turns on the fact that

$$a_{ij}+a_{ij} \neq 0,$$

may be false in a field of characteristic 2. The diagonal form of Theorem 28 is not always attainable in such a field.

References to the exclusion of fields of characteristic 2 at this point of the work are common in the literature of the subject, e.g., MacDuffee, p. 56, Theorem 34.1.

LEMMA 2. *When A is symmetrical and B is congruent to A, then B is also symmetrical.*

PROOF. Let $A' = A$ and $B = H'AH$. Then

$$B' = H'A'H = H'AH = B.$$

2. The diagonal form

THEOREM 28. *Given a symmetrical matrix A, other than the null matrix, there is a matrix H for which $C = H'AH$ is a diagonal matrix. Moreover, if the elements of A lie in a field F, so do the elements of H and C.*

PROOF. (i) If $a_{11} = 0$ we first find a congruent matrix B in which $b_{11} \neq 0$. If $a_{11} \neq 0$, we put $B \equiv A$.

(ii) We now have B congruent to A and $b_{11} \neq 0$. Add h times the first column of B to its jth column and in the result add h times the first row to the jth row. The resulting matrix is congruent to B and the jth element of its first row is

$$b_{1j}+hb_{11}.$$

A suitable choice of h makes this zero and a succession of such steps yields a matrix D congruent to A and having as its first row $d_{11} \neq 0$ followed by $n-1$ zeros. Since D is symmetrical, the first column is also d_{11} followed by $n-1$ zeros.

Moreover, any matrix I_{ij} used in step (i) has elements that are either zero or unity, so that the elements of B lie in the field F. Again in step (ii), the matrices $I+H_{ij}$ involved in effecting the congruences have elements that are either zero, unity, or $-b_{1j}/b_{11}$; hence their elements lie in F and so then do those of

$$(I+H_{ji})B(I+H_{ij}).$$

† W. V. D. Hodge and D. Pedoe, *Methods of Algebraic Geometry* (Cambridge, 1947), p. 14; A. A. Albert, *Modern Higher Algebra* (University of Chicago Science Series, 1936), p. 30.

and (ii) *transitive*, for when $B = H'AH$ and $C = K'BK$,

$$C = K'H'AHK$$
$$= (HK)'A(HK).$$

In the sequel it will be important to recognize that when B is congruent to A, C to B,..., and finally L to M, then M is congruent to A.

1.3. We shall often use two particular types of matrix congruent to A:

(i) The matrix got from A by interchanging two rows and then interchanging two columns in the resulting matrix; this is, in the notation of Chapter II, § 1.3, the matrix

$$I'_{ij}AI_{ij}.$$

(ii) The matrix obtained by adding h times the ith column of A to its jth column, and then adding h times the ith row to the jth row in the resulting matrix; this is, in the notation of Chapter II, § 1.3, the matrix

$$(I+H_{ji})A(I+H_{ij}) = (I+H_{ij})'A(I+H_{ij}).$$

1.4. *Preliminary lemmas*

LEMMA 1. *Given a symmetrical matrix A, other than the null matrix, in which a_{11} is zero, there is a congruent matrix B in which b_{11} is not zero.*

PROOF. If any diagonal element, say a_{ii}, is not zero, the interchange of the first and ith rows followed by interchange of the first and ith columns gives a non-zero first element; the result is a matrix congruent to A and is the matrix B required.

If all a_{ii} are zero, then at least one a_{ij} $(i \neq j)$ is not zero. The matrix

$$C = (I+H_{ji})A(I+H_{ij}),$$

with $h = 1$ [add ith column to jth and then ith row to jth], has an element $a_{ij}+a_{ji} = 2a_{ij}$ at the jth place of the leading diagonal. We obtain B from C by interchanging the first and jth rows and then the first and jth columns.

NOTE. When we consider the wider domain of abstract algebra, it is possible for a field F to have 'CHARACTERISTIC p'; that is to say, for each element α of F

$$\alpha+\alpha+\ldots \text{ to } p \text{ terms}$$

is equal to zero.† Lemma 1, which turns on the fact that

$$a_{ij}+a_{ij} \neq 0,$$

may be false in a field of characteristic 2. The diagonal form of Theorem 28 is not always attainable in such a field.

References to the exclusion of fields of characteristic 2 at this point of the work are common in the literature of the subject, e.g., MacDuffee, p. 56, Theorem 34.1.

Lemma 2. *When A is symmetrical and B is congruent to A, then B is also symmetrical.*

Proof. Let $A' = A$ and $B = H'AH$. Then

$$B' = H'A'H = H'AH = B.$$

2. The diagonal form

Theorem 28. *Given a symmetrical matrix A, other than the null matrix, there is a matrix H for which $C = H'AH$ is a diagonal matrix. Moreover, if the elements of A lie in a field F, so do the elements of H and C.*

Proof. (i) If $a_{11} = 0$ we first find a congruent matrix B in which $b_{11} \neq 0$. If $a_{11} \neq 0$, we put $B \equiv A$.

(ii) We now have B congruent to A and $b_{11} \neq 0$. Add h times the first column of B to its jth column and in the result add h times the first row to the jth row. The resulting matrix is congruent to B and the jth element of its first row is

$$b_{1j}+hb_{11}.$$

A suitable choice of h makes this zero and a succession of such steps yields a matrix D congruent to A and having as its first row $d_{11} \neq 0$ followed by $n-1$ zeros. Since D is symmetrical, the first column is also d_{11} followed by $n-1$ zeros.

Moreover, any matrix I_{ij} used in step (i) has elements that are either zero or unity, so that the elements of B lie in the field F. Again in step (ii), the matrices $I+H_{ij}$ involved in effecting the congruences have elements that are either zero, unity, or $-b_{1j}/b_{11}$; hence their elements lie in F and so then do those of

$$(I+H_{ji})B(I+H_{ij}).$$

† W. V. D. Hodge and D. Pedoe, *Methods of Algebraic Geometry* (Cambridge, 1947), p. 14; A. A. Albert, *Modern Higher Algebra* (University of Chicago Science Series, 1936), p. 30.

(iii) We now have a congruent matrix D, with elements in F, in which $d_{12},\ldots, d_{1n}$ and $d_{21},\ldots, d_{n1}$ are all zeros. When we interchange the ith and jth rows (or columns), where $i, j > 1$, or add multiples of row i to row j, we do not alter the zero values of the $n-1$ zeros in the first row or column. We can therefore, writing

$$D = \begin{bmatrix} d_{11} & o' \\ o & D_{n-1} \end{bmatrix},$$

work on D_{n-1} as before we worked on B without affecting the zeros of the first row: this will give a matrix

$$E = \begin{bmatrix} e_{11} & 0 & \\ 0 & e_{22} & O' \\ & O & E_{n-2} \end{bmatrix},$$

in which O and O' represent blocks of zeros. This matrix is congruent to A and may be written

$$E = \operatorname{diag}(e_{11}, e_{22}, E_{n-2}),$$

where e_{11}, e_{22} are non-zero elements of F and E_{n-2} is a matrix of order $n-2$ with its elements in F.

The theorem follows by continuing the process until we obtain

EITHER $\quad C = \operatorname{diag}(c_{11}, c_{22},\ldots, c_{nn}),$

OR $\quad C = \operatorname{diag}(c_{11}, c_{22},\ldots, c_{rr}, O),$

in which $r < n$ and O is the null matrix of order $n-r$.

COROLLARY. *The number of non-zero elements in the diagonal form is equal to the rank of the matrix A.*

PROOF. The matrices I_{ij}, $I+H_{ij}$ used to effect the change from A to C are non-singular, and so the rank is unchanged throughout.

3. Orthogonal and unitary matrices

3.1. *Preliminary*

The theorem to be stated in § 3.2 is a development of the following simple idea.

In two dimensions, let $\mathbf{x}$, $\mathbf{y}$ be perpendicular unit vectors and $\mathbf{v}$ any other vector. Then

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

where $(v.x)$ and $(v.y)$ are the lengths of the projections of $\mathbf{v}$ on $\mathbf{x}$ and $\mathbf{y}$ respectively. When $\mathbf{v}$ and $\mathbf{x}$ are given, we find a unit vector $\mathbf{y}$ perpendicular to $\mathbf{x}$ by observing that it is a numerical multiple of the vector

$$\mathbf{v}-(v.x)\mathbf{x}.$$

In three dimensions we may determine a unit vector $\mathbf{z}$ that is perpendicular to $\mathbf{x}$ and $\mathbf{y}$ by writing

$$\mathbf{v} = (v.x)\mathbf{x}+(v.y)\mathbf{y}+(v.z)\mathbf{z}$$

in the form $$(v.z)\mathbf{z} = \mathbf{v}-(v.x)\mathbf{x}-(v.y)\mathbf{y}.$$

3.2. In this section we use $\mathbf{x}_r$ to denote a vector (or single column matrix) with real components (or elements)

$$x_{1r},\quad x_{2r},\quad \ldots,\quad x_{nr}.$$

We use the notation $(x_r.y_s)$ to denote the SCALAR (or INNER) PRODUCT

$$x_{1r}y_{1s}+\ldots+x_{nr}y_{ns}$$

of the two vectors $\mathbf{x}_r$ and $\mathbf{y}_s$ and we say that $\mathbf{x}_r$ is a unit vector if $(x_r.x_r) = 1$. We say that $\mathbf{x}_r$, $\mathbf{y}_s$ are orthogonal if $(x_r.y_s) = 0$.

We say that $\mathbf{x}_1,\ldots, \mathbf{x}_m$ are linearly independent† when the only numbers $\lambda_1,\ldots, \lambda_m$ to satisfy

$$\lambda_1\mathbf{x}_1+\ldots+\lambda_m\mathbf{x}_m = 0$$

are given by $\lambda_1 = \ldots = \lambda_m = 0$.

THEOREM 29. *Let $\mathbf{x}_r$ $(r = 1,\ldots, n)$ be n given linearly independent vectors with real components. Then we can determine real constants α_{ij} so that the n vectors $\mathbf{z}_r$ given by*

$$\begin{aligned}\mathbf{z}_1 &= \alpha_{11}\mathbf{x}_1,\\ \mathbf{z}_2 &= \alpha_{12}\mathbf{x}_1+\alpha_{22}\mathbf{x}_2,\\ \mathbf{z}_3 &= \alpha_{13}\mathbf{x}_1+\alpha_{23}\mathbf{x}_2+\alpha_{33}\mathbf{x}_3,\\ &\ldots\ldots\ldots\end{aligned} \tag{1}$$

are linearly independent unit vectors which are mutually orthogonal; that is to say,

$$(z_r.z_r) = 1,\quad (z_r.z_s) = 0 \quad \textit{when} \quad r \neq s.$$

Moreover, no α_{rr} is equal to zero.

† Cf. F. 94. I make no attempt to repeat here the details of my previous book concerning linear dependence in a *field*.

PROOF. In succession, put†

$$\begin{aligned} \mathbf{z}_1 &= k_1\mathbf{x}_1, \\ \mathbf{z}_2 &= k_2[\mathbf{x}_2-(z_1 . x_2)\mathbf{z}_1], \\ \mathbf{z}_3 &= k_3[\mathbf{x}_3-(z_1 . x_3)\mathbf{z}_1-(z_2 . x_3)\mathbf{z}_2], \end{aligned} \tag{2}$$

and so on up to $\mathbf{z}_n$, choosing the k at each step so that the vector on the left is a unit vector. (This choice of k is always possible provided the vector on the right is not zero, and if the mth vector on the right were to be zero there would be a linear relation

$$\mathbf{x}_m+\alpha\mathbf{x}_{m-1}+...+\kappa\mathbf{x}_1 = 0,$$

which would contradict the hypothesis that the $\mathbf{x}_r$ are linearly independent.)

From their method of construction, the $\mathbf{z}$ vectors are orthogonal.‡ An independent proof of this fact is, in outline:

Suppose that for r, $s = 1, 2,..., R-1$

$$(z_r . z_s) = 0 \quad \text{when} \quad r \neq s. \tag{3}$$

Then, for $r \leqslant R-1$,

$$(z_r . z_R) = k_R[(z_r . x_R)-(z_r . x_R)(z_r . z_r)], \tag{4}$$

all other terms cancelling in virtue of (3). But $(z_r . z_r) = 1$ and so (4) is zero. The proof by induction that $(z_r . z_s) = 0$ when $r \neq s$ and $r, s = 1, 2,..., n$ is immediate.

The form (1) follows from (2) when we express the $\mathbf{z}$'s on the right of (2) in terms of the $\mathbf{x}$'s. Moreover $\alpha_{rr} = k_r \neq 0$.

3.3. *Orthogonal matrices*

DEFINITION 16. *When all the elements below the principal diagonal of a matrix are zero, we say that the matrix is* TRIANGULAR.

THEOREM 30. *Given a non-singular matrix X, with real elements, there is a real triangular matrix A for which*

$$Z = XA$$

is an orthogonal matrix; that is to say,

$$Z'Z = I, \qquad Z' = Z^{-1}.$$

† Usually referred to as Schmidt's orthogonalization process.

‡ The statement depends on a certain degree of familiarity with vectors in n dimensions.

PROOF. The columns of X form n vectors, say

$$\mathbf{x}_r \text{ with components } x_{1r}, x_{2r}, \ldots, x_{nr};$$

since X is non-singular, these vectors are linearly independent. Let A be

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} & . & . & \alpha_{1n} \\ 0 & \alpha_{22} & . & . & \alpha_{2n} \\ . & . & . & . & . \\ 0 & 0 & . & . & \alpha_{nn} \end{bmatrix},$$

where the α's are the numbers that occur in (1) of Theorem 29. Then the rth column of XA consists of the components of the vector $\mathbf{z}_r$ of Theorem 29. Denote these components by $z_{1r}, z_{2r}, \ldots, z_{nr}$. Then, on writing

$$XA = Z = \begin{bmatrix} z_{11} & . & . & z_{1n} \\ . & . & . & . \\ z_{n1} & . & . & z_{nn} \end{bmatrix},$$

we see that

$$Z'Z = \begin{bmatrix} z_{11} & z_{21} & . & . & z_{n1} \\ z_{12} & z_{22} & . & . & z_{n2} \\ . & . & . & . & . \\ z_{1n} & z_{2n} & . & . & z_{nn} \end{bmatrix} \times \begin{bmatrix} z_{11} & z_{12} & . & . & z_{1n} \\ z_{21} & z_{22} & . & . & z_{2n} \\ . & . & . & . & . \\ z_{n1} & z_{n2} & . & . & z_{nn} \end{bmatrix}$$

has the element $(z_r . z_s)$ in its rth row and sth column. Hence, by Theorem 29,

$$Z'Z = I.$$

3.4. *Unitary matrices*

As in § 3.2, use $\mathbf{x}_r$ to denote a vector, or a single column matrix, with elements

$$x_{1r}, \quad x_{2r}, \quad \ldots, \quad x_{nr};$$

but now think of these elements as, possibly, complex numbers and write

$$(x_r . y_s) = \bar{x}_{1r} y_{1s} + \ldots + \bar{x}_{nr} y_{ns}. \tag{5}$$

With this notation, $(x_r . y_s)$ is the conjugate complex of $(y_s . x_r)$; if either of them is zero, so is the other.

We continue to call two vectors $\mathbf{x}_r$ and $\mathbf{y}_s$ orthogonal when $(x_r . y_s) = 0$ and to call $\mathbf{x}_r$ a unit vector when $(x_r . x_r) = 1$.

THEOREM 31. *The results (and details of the proof) of Theorem 29 remain valid when the components of the vectors* $\mathbf{x}_r$ *are complex* *numbers and* $(x_r.y_s)$ *is defined by* (5) *above; the constants* α_{ij} *are no longer necessarily real.*

In setting out the proof of Theorem 29 we were careful to observe the order $(z_1.x_2)$, instead of the alternative order $(x_2.z_1)$. With this point attended to, the proof of Theorem 31 is a word-for-word copy of the proof of Theorem 29.

THEOREM 32. *Given a non-singular matrix* X *whose elements are complex numbers, there is a triangular matrix* A *for which*

$$Z = XA$$

is a unitary matrix; that is to say,

$$\bar{Z}'Z = I, \qquad \bar{Z}' = Z^{-1}.$$

PROOF. In the proof of Theorem 30 we need to change only the final step. We use Theorem 31 (instead of Theorem 29) to obtain the equations (1) of page 139 and then, with

$$XA = Z = \begin{bmatrix} z_{11} & . & . & z_{1n} \\ . & . & . & . \\ z_{n1} & . & . & z_{nn} \end{bmatrix},$$

we see that it is now [in virtue of definition (5)]

$$\bar{Z}'Z = \begin{bmatrix} \bar{z}_{11} & \bar{z}_{21} & . & . & \bar{z}_{n1} \\ . & . & . & . & . \\ \bar{z}_{1n} & \bar{z}_{2n} & . & . & \bar{z}_{nn} \end{bmatrix} \times \begin{bmatrix} z_{11} & z_{12} & . & . & z_{1n} \\ . & . & . & . & . \\ z_{n1} & z_{n2} & . & . & z_{nn} \end{bmatrix}$$

which has the element $(z_r.z_s)$ in its rth row and sth column.

4. Orthogonal congruence

4.1. *Reduction of a symmetrical matrix to a diagonal form*

Let A be a symmetrical matrix of real elements. Then its latent roots, say $\lambda_1,..., \lambda_n$, are all real numbers.† Since

$$|A-\lambda_1 I| = 0,$$

the equation $$A\mathbf{x} = \lambda_1 \mathbf{x} \tag{1}$$

has a non-zero solution, which may be taken to be a single-column *unit* vector with real elements

$$x_{11}, \quad x_{21}, \quad ..., \quad x_{n1}. \tag{2}$$

† F. 146.

Let X be a non-singular matrix with (2) as its first column and Z the orthogonal matrix derived from X by Theorem 30. Since $\mathbf{x}_1$ is itself a unit vector, Theorem 29 shows that $\mathbf{z}_1 = \mathbf{x}_1$ and that $\alpha_{11} = 1$: thus Z has (2) as its first column.

As in Chapter IV, § 4.1, (3),

$$Z^{-1}AZ = \begin{bmatrix} \lambda_1 & b' \\ o & B \end{bmatrix}, \tag{3}$$

wherein o is a column of zeros and B is a matrix of order $n-1$ whose characteristic roots are $\lambda_2,..., \lambda_n$. Since $Z^{-1} = Z'$ and A is symmetrical, the matrix $Z^{-1}AZ = Z'AZ$ is also symmetrical and b' is a row of zeros. Thus

$$Z^{-1}AZ = \text{diag}\{\lambda_1, B\}.$$

We can proceed as in Chapter IV, § 4.1, but now, at each step, using not any matrix with a given first column but an orthogonal matrix with a given unit vector as its first column. We obtain, in succession

$$A_1 = Z^{-1}AZ = \text{diag}\{\lambda_1, B\};$$

$$A_2 = Y^{-1}A_1 Y = \text{diag}\{\lambda_1, \lambda_2, C\},$$

where C has latent roots $\lambda_3,..., \lambda_n$; and so on until we reach

$$A_n = \text{diag}\{\lambda_1, \lambda_2,..., \lambda_n\}.$$

Also, each of Z, Y,... is an orthogonal matrix† of order n.

Now the product of two orthogonal matrices is itself orthogonal.‡ Accordingly, there is an orthogonal matrix K, the product of Z, Y,... in the above, for which

$$K'AK = K^{-1}AK = A_n = \text{diag}\{\lambda_1, \lambda_2,..., \lambda_n\}.$$

This proves

THEOREM 33. *Given a real symmetrical matrix A whose latent roots are $\lambda_1,..., \lambda_n$ (necessarily real, but not necessarily all different), there is a real orthogonal matrix K for which*

$$K'AK = K^{-1}AK = \text{diag}\{\lambda_1, \lambda_2,..., \lambda_n\}.$$

† When y is the $(n-1)$-rowed orthogonal matrix that makes $y^{-1}By$ a matrix with λ_2 as its leading element and zeros elsewhere in the first column, $Y = \text{diag}(1, y)$ is an n-rowed orthogonal matrix.

‡ When $Z'Z = Y'Y = I$, $(ZY)'(ZY) = Y'Z'ZY = Y'Y = I$.

4.2. *Reduction of a quadratic form to its canonical form*

The quadratic form†

$$\sum_{r=1}^{n}\sum_{s=1}^{n} a_{rs}x_r x_s \quad (a_{rs} = a_{sr})$$

may be written as the single-element matrix

$$x'Ax,$$

where x is a single-column matrix with elements $x_1,..., x_n$ and $A = [a_{rs}]$. Thus Theorem 33 gives at once

COROLLARY. *The orthogonal transformation $x = Ky$ changes the quadratic form $x'Ax$ into the canonical form*

$$y'(K'AK)y = \lambda_1 y_1^2+\lambda_2 y_2^2+...+\lambda_n y_n^2,$$

where $\lambda_1,..., \lambda_n$ are the latent roots of A.

When $x'Ax$ is a positive-definite form, all the λ's are positive‡ and a further change of variable $z_r = y_r\sqrt{\lambda_r}$ expresses $x'Ax$ as

$$z_1^2+z_2^2+...+z_n^2.$$

4.3. *Simultaneous reduction of two quadratic forms*

Theorem 33 also leads fairly quickly to a well-known result concerning the simultaneous reduction of two forms. This may be stated as follows:

THEOREM 34. *Given two real quadratic forms $x'Ax$ and $x'Cx$, of which the latter is positive-definite, there is a real, non-singular matrix H for which*

$$H'AH = \text{diag}(\lambda_1, \lambda_2,..., \lambda_n) \quad \textit{and} \quad H'CH = I,$$

where $\lambda_1,..., \lambda_n$ are the roots of the equation $|A-\lambda C| = 0$.

The transformation $x = Hz$ transforms

$$x'Ax \quad \textit{into} \quad \lambda_1 z_1^2+...+\lambda_n z_n^2$$

and

$$x'Cx \quad \textit{into} \quad z_1^2+...+z_n^2.$$

PROOF. By Theorem 33, there is a real orthogonal matrix K for which

$$K'CK = \text{diag}(\mu_1,..., \mu_n).$$

The μ's are all necessarily positive.§ Put

$$M = \text{diag}(\mu_1^{-\frac{1}{2}},..., \mu_n^{-\frac{1}{2}}).$$

† F. 122. ‡ F. 146. § F. 135, Theorem 40; or F. 146.

Then
$$M'K'CKM = I. \tag{4}$$

Suppose that $M'K'AKM = B$. Then B is symmetrical and there is, by Theorem 33, a real orthogonal matrix L for which

$$L'BL = \text{diag}(\lambda_1,..., \lambda_n),$$

where $\lambda_1,..., \lambda_n$ are the latent roots of B. Also

$$L'(M'K'CKM)L = L'IL = I.$$

Put $H = KML$. Then H is real and non-singular: also, since $H'AH = L'(M'K'AKM)L = L'BL$,

$$H'AH = \text{diag}(\lambda_1,..., \lambda_n), \qquad H'CH = I.$$

For all values of λ,

$$H'(A-C\lambda)H = \text{diag}(\lambda_1-\lambda,..., \lambda_n-\lambda),$$

so that $\lambda_1,..., \lambda_n$ are the roots† of $|A-C\lambda| = 0$.

5. Hermitian matrices

5.1. *Preliminary detail of notation*

We take as our standard Hermitian form‡

$$\sum_{r=1}^{n} \sum_{s=1}^{n} a_{rs} \bar{x}_r x_s \quad (a_{rs} = \bar{a}_{sr}), \tag{1}$$

so that, when x is a single-column matrix with elements x_1, $x_2,..., x_n$ and $A = [a_{rs}]$, the form is a single-element matrix

$$\bar{x}'Ax.$$

A change of variables $x = Hy$, accompanied by its conjugate complex, transforms (1) into $\bar{y}'By$, where

$$B = \bar{H}'AH.$$

5.2. *Conjunction*

When the matrix B is related to the matrix A by a relation

$$B = \bar{H}'AH,$$

we say that B IS CONJUNCTIVE WITH A. This relation, like congruence, is reflexive and transitive; also, when A is Hermitian (i.e. $\bar{A}' = A$), B is also Hermitian.

† F. 144, Theorem 43.

‡ In my former book [e.g. F. 3, 129] I took the standard form as $x'A\bar{x}$, with consequent minor differences of detail.

In particular (cf. § 1.3),

(i) $$\bar{I}'_{ij}AI_{ij}$$

effects an interchange of two rows of A followed by an interchange of two columns;

(ii) $$(I+\bar{H}_{ji})A(I+H_{ij})$$

is conjunctive with A and is obtained by adding h times the ith column of A to its jth column and, in the result, adding $\bar{h}$ times the ith row to the jth row.

As in § 1.4 we require a preliminary lemma.

Given a Hermitian matrix A, other than the null matrix, in which a_{11} is zero, there is a conjunctive matrix B in which b_{11} is not zero.

PROOF. When there is a non-zero a_{ii}, B is given by

$$\bar{I}'_{1i}AI_{1i}.$$

When all a_{ii} are zero, there is an a_{ij} $(i \neq j)$ which is not zero. The conjunctive matrix

$$C = (I+\bar{H}_{ji})A(I+H_{ij})$$

has an element c_{jj} given by

$$ha_{ji}+\text{its conjugate complex};$$

either the choice $h = 1$ makes $c_{jj} \neq 0$ or the choice $h = i$ will do so; and, when this is done, an interchange of the first and jth rows followed by an interchange of the first and jth columns gives the matrix B required.

5.21. *The diagonal form*

THEOREM 35. *Given a Hermitian matrix A, other than the null matrix, there is a matrix H for which $C = \bar{H}'AH$ is a diagonal matrix. Moreover, if the elements of A lie in a field F, so do the elements of H and C.*

The proof differs from that of Theorem 28 only in small details.

5.3. *Unitary conjunction*

Let A be a Hermitian matrix. Then its latent roots, say $\lambda_1,\ldots, \lambda_n$, are all real numbers.† The equation

$$A\mathbf{x} = \lambda_1\mathbf{x} \tag{1}$$

† F. 158, Example 14.

has as a solution a single-column *unit* vector with complex elements

$$x_{11}, \quad x_{21}, \quad \ldots, \quad x_{n1}. \tag{2}$$

Choose X a non-singular matrix with (2) as its first column and let Z be the unitary matrix derived from X by Theorem 32. Proceed as in § 4.1 and we obtain

THEOREM 36. *Given a Hermitian matrix A whose latent roots are $\lambda_1,\ldots, \lambda_n$ (necessarily real, but not necessarily all different), there is a unitary matrix K for which*

$$K^{-1}AK = \mathrm{diag}\{\lambda_1, \lambda_2,\ldots, \lambda_n\}.$$

The corollary analogous to that in § 4.2 follows at once.

The unitary transformation $x = Ky$ changes the Hermitian form $\bar{x}'Ax$ into the canonical form

$$\bar{y}'(\bar{K}'AK)y = \bar{y}'(K^{-1}AK)y = \lambda_1 \bar{y}_1 y_1 + \ldots + \lambda_n \bar{y}_n y_n,$$

where $\lambda_1,\ldots, \lambda_n$ are the latent roots of A.

5.4. *Simultaneous reduction of two Hermitian forms*

Finally, as in 4.3, we can prove

THEOREM 37. *Given two Hermitian forms $\bar{x}'Ax$ and $\bar{x}'Cx$, of which the latter is positive-definite, there is a non-singular matrix H for which*

$$\bar{H}'AH = \mathrm{diag}(\lambda_1, \lambda_2,\ldots, \lambda_n), \qquad \bar{H}'CH = I,$$

where $\lambda_1, \lambda_2,\ldots, \lambda_n$ are the roots of the equation $|A-\lambda C| = 0$.

The transformation $x = Hz$ transforms

$$\bar{x}'Ax \quad \textit{into} \quad \lambda_1 \bar{z}_1 z_1 + \ldots + \lambda_n \bar{z}_n z_n$$

and

$$\bar{x}'Cx \quad \textit{into} \quad \bar{z}_1 z_1 + \ldots + \bar{z}_n z_n.$$

5.5. *Note on Theorems 35–7*

It is a common practice in the literature of the subject to prove Theorems 35–7 at the same time as one proves Theorems 28, 33, and 34 by using a special notation $\tilde{H}'AH$ which is interpreted to be

$$H'AH \quad \text{when} \quad A \text{ is symmetrical,}$$

$$\bar{H}'AH \quad \text{when} \quad A \text{ is Hermitian.}$$

The resulting brevity has its advantages, but I have always preferred to deal with the real quadratic form and then see what minor adjustments are necessary to the Hermitian form.

6. Repeated characteristic roots and rank

6.1. *Symmetrical matrices*

Let A be a real symmetrical matrix with latent roots $\lambda_1,\ldots,\lambda_n$ and let λ_1 be an r-ple root. Then, by Theorem 33,

$$K^{-1}(A-\lambda I)K = \operatorname{diag}(\lambda_1-\lambda,\ldots,\lambda_n-\lambda).$$

When $\lambda = \lambda_1$ the latter matrix has just r zeros in the diagonal and so, since multiplication by non-singular matrices leaves rank unaltered, the rank of $A-\lambda_1 I$ is $n-r$. The equation

$$A\mathbf{x} = \lambda_1 \mathbf{x}$$

will have r linearly independent solutions† $x_1,\ldots, x_r$ each of which is a single-column matrix.

The same result holds good, *mutatis mutandis*, for repeated roots of $|A-C\lambda| = 0$ and linearly independent vectors $\mathbf{x}$ satisfying $A\mathbf{x} = \lambda C\mathbf{x}$.

An alternative treatment‡ of the whole matter is to prove directly that an r-ple root implies r linearly independent solutions and then make use of this fact in determining the matrix K of Theorem 33.

6.2. *Unsymmetrical matrices*

The result is quite different for matrices which are not symmetrical. Let us take, for simplicity, an unsymmetrical matrix A whose classical canonical form is $C_3(\alpha)$; then, for some non-singular matrix T,

$$T^{-1}AT = \begin{bmatrix} \alpha & 1 & 0 \\ 0 & \alpha & 1 \\ 0 & 0 & \alpha \end{bmatrix}$$

and the equation $|A-\lambda I| = 0$ is $(\lambda-\alpha)^3 = 0$. The matrix $A-\alpha I$ is

$$T\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}T^{-1}, \tag{1}$$

whose rank is 2 and so the equation

$$A\mathbf{x} = \alpha\mathbf{x} \tag{2}$$

has only one solution. Similarly, when $C_n(\alpha)$ replaces $C_3(\alpha)$ we

† Cf. Chap. I, § 9.

‡ J. A. Todd, 'A note on real quadratic forms', *Quart. J. of Math.* (Oxford), 18 (1947), 183–5; also the paper by W. L. Ferrar which follows it.

reach (1) with a matrix of rank $n-1$ and equation (2) again has only one solution.

The actual linear equations that make up (2) show how inevitable the result is. Take $A = C_3(\alpha)$ and $\mathbf{x} = \{x_1, x_2, x_3\}$. The scalar equations represented by (2) are

$$\text{(i)} \quad \alpha x_1 + x_2 = \alpha x_1,$$

$$\text{(ii)} \quad \alpha x_2 + x_3 = \alpha x_2,$$

$$\text{(iii)} \quad \alpha x_3 = \alpha x_3.$$

From (i), $x_2 = 0$ and x_1 is arbitrary; from (ii), $x_3 = 0$; and (iii) is automatically satisfied. The only non-zero solution is a multiple of

$$\mathbf{x} = \{1, 0, 0\}.$$

Considering now a more complex canonical form, we see that

$$C = \text{diag}\{C_p(\alpha), C_q(\beta), \ldots\}$$

when $p > 1$ and $\alpha \neq \beta, \gamma, \ldots$ gives only one solution of $C\mathbf{x} = \alpha\mathbf{x}$, while a form†

$$C = \text{diag}\{C_1(\alpha), C_1(\alpha), C_1(\alpha), C_q(\beta), \ldots\}$$

gives three linearly independent solutions of $C\mathbf{x} = \alpha x$ because the matrix $C - \alpha I$ is of rank $n-3$.

7. A direct proof of Theorem 34

The theorem to be proved is this:

Given two real symmetrical matrices A and C, the latter being such that the quadratic form $x'Cx$ is positive-definite, there is a non-singular matrix H for which

$$H'AH = \text{diag}(\lambda_1, \ldots, \lambda_n), \qquad H'CH = I,$$

where $\lambda_1, \ldots, \lambda_n$ are the roots of the equation $|A - \lambda C| = 0$.

The proof given in § 4.3 is somewhat indirect; the proof now to be given depends upon a direct step-by-step derivation of the matrix H. The details have some minor interest of their own.

† A form $\qquad C = \text{diag}\{C_p(\alpha), C_q(\alpha), \ldots, C_r(\beta), \ldots\}$
will give $\qquad C - \alpha I = \text{diag}\{C_p(0), C_q(0), \ldots, C_r(\beta - \alpha), \ldots\}$;
the ranks of $C_p(0)$, $C_q(0)$,... are $p-1$, $q-1$,..., and the ranks of $C_r(\beta-\alpha)$,... are r,... .

7.1. *Preliminary lemma*

LEMMA. *Let A, Z be square matrices of n rows and columns and let the r-th column of Z be the single-column matrix z_r. Then the element in the r-th row and s-th column of $Z'AZ$ is the single-element matrix $z_r' A z_s$; in symbols*

$$Z'AZ = [z_r' A z_s].$$

PROOF. We may write

$$Z = z_1 + z_2 + \ldots + z_n,$$

where each z_r is a single-column matrix accompanied by $n-1$ columns of zeros, and

$$Z' = z_1' + z_2' + \ldots + z_n',$$

where each z_r' is a single-row matrix accompanied by $n-1$ rows of zeros. It is now obvious that

$$Z'AZ = (z_1' + \ldots + z_n')A(z_1 + \ldots + z_n) = [z_r' A z_s].$$

7.2. *Schmidt's orthogonalization process*

Let $C = [c_{ij}]$ be the matrix of a positive-definite form, $\mathbf{x}_r$ a vector with components

$$x_{ir} \quad (i = 1, \ldots, n).$$

Define $(x_r . y_s)$ to be

$$x_r' C y_s = \sum_{i,j} c_{ij} x_{ir} y_{js}.$$

Then, since C is symmetrical and $c_{ij} = c_{ji}$,

$$(x_r . y_s) = (y_s . x_r).$$

We say that $\mathbf{x}_r$ and $\mathbf{y}_s$ are c-orthogonal when $(x_r . y_s) = 0$ and we call $\mathbf{x}_r$ a unit vector when $(x_r . x_r) = 1$.

Given a set of n linearly independent vectors $\mathbf{x}_r$ with real components, we define another set $\mathbf{z}_r$ by the equations

$$\mathbf{z}_1 = k_1 \mathbf{x}_1, \qquad \mathbf{z}_2 = k_2[\mathbf{x}_2 - (z_1 . x_2)\mathbf{z}_1],$$

and so on (as in (2) of § 3.2), choosing the k at each step so that the $\mathbf{z}$ is a unit vector. [When $\mathbf{x}_1$ is itself a unit vector, $k_1 = 1$.] These unit vectors are linearly independent and are mutually c-orthogonal [cf. § 3.2]; that is

$$z_r' C z_r = 1, \qquad z_r' C z_s = 0 \quad (r \neq s).$$

We may rewrite the above equations in the form

$$\mathbf{z}_1 = \alpha_{11}\mathbf{x}_1, \qquad \mathbf{z}_2 = \alpha_{12}\mathbf{x}_1 + a_{22}\mathbf{x}_2,$$

and so on, the α's being real and α_{11} being unity when $\mathbf{x}_1$ is a unit vector. Thus, given a real non-singular matrix X, there is (cf. Theorem 30) a real triangular matrix A such that

$$Z = XA$$

has columns $\mathbf{z}_1,..., \mathbf{z}_n$ which are mutually c-orthogonal unit vectors.

By the lemma of § 7.1,

$$Z'CZ = [z'_r Cz_s] = I,$$

since the vectors are mutually c-orthogonal unit vectors.

7.3. *First step of the reduction*

The roots of the equation $|A-\lambda C| = 0$ are necessarily real.†
Since $|A-\lambda_1 C| = 0$, the equation

$$A\mathbf{x} = \lambda_1 C\mathbf{x}$$

has a non-zero solution which may be taken to be a unit vector (in the sense of § 7.2) $\mathbf{x}_1$.

Let X be a non-singular matrix with $\mathbf{x}_1$ as its first column and Z the c-orthogonal matrix derived from it by the process of § 7.2. Then Z has a first column $\mathbf{z}_1$ satisfying

$$A\mathbf{z}_1 = \lambda_1 C\mathbf{z}_1. \tag{1}$$

By the lemma of § 7.1, the first column of $Z'AZ$ is

$$\mathbf{z}'_r A\mathbf{z}_1,$$

or, on using (1), $\lambda_1 \mathbf{z}'_r C\mathbf{z}_1$. Since the $\mathbf{z}_r$ are mutually c-orthogonal unit vectors, this first column is

$$\lambda_1, \quad 0, \quad 0, \quad ..., \quad 0.$$

The first row of $Z'AZ$, which is a symmetrical matrix, must also be λ_1 followed by $n-1$ zeros. Thus

$$Z'AZ = \text{diag}(\lambda_1, B),$$

† F. 145.

where B has $n-1$ rows and columns. Moreover,

$$Z'CZ = I.$$

Now the roots of $|A-\lambda C| = 0$ are the same as the roots of $|Z'(A-\lambda C)Z| = 0$; that is, they are the latent roots of

$$\text{diag}(\lambda_1-\lambda, B-\lambda I_{n-1}),$$

where I_{n-1} is the unit matrix of $n-1$ rows and columns. Hence $\lambda_2,\ldots, \lambda_n$ are the latent roots of B.

7.4. *Completing the reduction*

Since $|B-\lambda_2 I_{n-1}| = 0$, there is a non-zero unit† vector y_2, with elements

$$y_{22}, \quad y_{32}, \quad \ldots, \quad y_{n2}, \tag{2}$$

for which $By_2 = \lambda_2 y_2$; there is then an orthogonal matrix Y (of order $n-1$) having (2) as its first column. The first column of $Y'BY$ is

$$y'_r By_2 = y'_r \lambda_2 y_2.$$

Thus, since Y is an orthogonal matrix, the first column (and row, by symmetry) of $Y'BY$ is λ_2 followed by $n-2$ zeros.

Let
$$L = \text{diag}(1, Y).$$
Then
$$L'(Z'AZ)L = \begin{bmatrix} 1 & . \\ . & Y' \end{bmatrix} \times \begin{bmatrix} \lambda_1 & . \\ . & B \end{bmatrix} \times \begin{bmatrix} 1 & . \\ . & Y \end{bmatrix}$$
$$= \begin{bmatrix} \lambda_1 & . \\ . & Y'BY \end{bmatrix}$$
$$= \text{diag}(\lambda_1, \lambda_2, D)$$

where, by the previous argument, the latent roots of D are $\lambda_3,\ldots, \lambda_n$. Moreover,

$$L'(Z'CZ)L = L'IL = I.$$

The process continues step by step to the final forms

$$H'AH = \text{diag}(\lambda_1,\ldots, \lambda_n), \qquad H'CH = I.$$

7.5. *Hermitian forms*

Theorem 37 admits a similar proof; the definition of $(x_r \cdot y_s)$ related to the Hermitian form $c_{ij}\bar{x}_i x_j$ follows the pattern of § 3.4.

† 'Unit-vector' and 'orthogonal' here bear their usual meanings; the positive-definite form $c_{ij}x_i x_j$ has been replaced, since $Z'CZ = I$, by the simpler form $z_1^2+z_2^2+\ldots+z_n^2$.

CHAPTER VII

MATRIX EQUATIONS

1. The minimum equation

We have seen that a matrix A, of order n, satisfies its characteristic equation; that is, if

$$|A-\lambda I| \equiv f(\lambda) \equiv \sum_{r=0}^{n} p_r \lambda^r,$$

then

$$f(A) \equiv \sum_{r=0}^{n} p_r A^r = 0.$$

Further, we have proved that, for a given matrix A, there is a polynomial $h(\lambda)$ of minimum degree m such that

(i) $h(A) = 0$,

(ii) the matrix A cannot satisfy any equation of degree less than m,

(iii) $h(\lambda) \equiv E_n$, the nth invariant factor of $\lambda I - A$.

Before going on to other topics, we show that the results about functions of matrices developed in Chapter V render intuitive the facts concerning this MINIMUM FUNCTION† which we established in Chapter IV, § 11.

Let A be a given matrix, $C = TAT^{-1}$ its classical canonical form, and $g(\lambda)$ any polynomial in λ. Then

$$g(C) = Tg(A)T^{-1}$$

and $g(A) = 0$ if and only if $g(C) = 0$. Now, when

$$C = \operatorname{diag}\{C_k(\alpha),...\},$$

$$g(C) = \operatorname{diag}\{G_k(\alpha),...\},$$

where
$$G_k(\alpha) = \begin{bmatrix} g(\alpha) & g'(\alpha) & . & . & g^{(k-1)}(\alpha)/(k-1)! \\ 0 & g(\alpha) & . & . & g^{(k-2)}(\alpha)/(k-2)! \\ . & . & . & . & . \\ 0 & 0 & . & . & g(\alpha) \end{bmatrix}. \qquad (1)$$

Thus $g(C) = 0$ if and only if each $G_k(\alpha)$ is a null matrix and this is so if and only if $g(\alpha)$, $g'(\alpha)$,..., $g^{(k-1)}(\alpha)$ are all zero; this in turn happens if and only if $g(\lambda)$ contains $(\lambda-\alpha)^k$ as a factor.

† Sometimes called the '*reduced characteristic function*' or the R.C.F. (cf. Turnbull and Aitken, loc. cit.).

Thus when C has a Segre characteristic

$$[(xy\ldots z)(uv\ldots w)\ldots rs\ldots]$$

in which $x \leqslant y \leqslant \ldots \leqslant z$, $u \leqslant v \leqslant \ldots \leqslant w$,... and

$$C = \mathrm{diag}\{C_x(\alpha),\ldots, C_z(\alpha), C_u(\beta),\ldots, C_w(\beta),\ldots C_r(\zeta)\ldots\},$$

$g(C)$ is the null matrix if and only if $g(\lambda)$ contains

$$(\lambda-\alpha)^z(\lambda-\beta)^w\ldots(\lambda-\zeta)^r\ldots$$

as a factor.

Since E_n, the nth invariant factor of $\lambda I-C$ is given (Chap. IV, § 8) by
$$E_n = (\lambda-\alpha)^z(\lambda-\beta)^w\ldots(\lambda-\zeta)^r\ldots,$$
the same result is expressed in other words if we say that $g(A) = 0$ if and only if $g(\lambda)$ contains E_n as a factor (cf. Theorem 19).

2. Solutions of a given scalar equation

We suppose given a scalar polynomial,

$$g(x) \equiv x^N + g_1 x^{N-1} + \ldots + g_N$$

say, and we seek matrices A that satisfy the equation

$$g(A) = 0. \tag{1}$$

We suppose always that, by ordinary algebra, $g(x)$ has been expressed in the standard form†

$$g(x) = Q_1(x)\{Q_2(x)\}^2\ldots\{Q_k(x)\}^k \tag{2}$$

in which no $Q_r(x)$ has a repeated factor and no factor of one Q is a factor of any other Q.

2.1. *Solution in classical canonical form*

We seek classical canonical matrices C for which

$$g(C) = 0. \tag{3}$$

Let
$$Q_r(x) \equiv (x-\alpha_{r1})(x-\alpha_{r2})\ldots. \tag{4}$$

Then, as we saw in § 1, a matrix

$$C = \mathrm{diag}\{C_p(\lambda),\ldots\} \tag{5}$$

cannot satisfy (3) unless each λ has one of the values

$$\alpha_{rs}\ (r = 1, 2,\ldots, k;\ s = 1, 2,\ldots).$$

† W. L. Ferrar, *Higher Algebra* (Oxford, 1948), p. 229, Theorem 40.

Further, *the form* $C = \text{diag}\{C_p(\alpha_{rs}),...\}$, (6)

where now the λ*'s of* (5) *are restricted to be roots of* $g(x) = 0$, *will satisfy* (3) *if and only if* $p \leqslant r$ *for each submatrix that occurs in* (6). For if $p \leqslant r$, $g(\lambda)$ contains $(\lambda-\alpha_{rs})^p$ as a factor for each p and α_{rs} of (6); while if $p > r$ for any one submatrix of (6), $g(\lambda)$ contains $\lambda-\alpha_{rs}$ only to the power r and the corresponding $G_p(\alpha_{rs})$ in § 1, (1) cannot be a null matrix.

Thus (6), with $p \leqslant r$ at each entry, gives the general solution of (3); further, the general solution† of

$$g(A) = 0 \tag{7}$$

is given by $$A = TCT^{-1},$$

where C is the general solution of (3) and T is an arbitrary, non-singular matrix of the same order as C.

By taking $g(x) = Q_1(x)$ in (2) we see that *when the equation* $g(x) = 0$ *has no repeated root, any canonical matrix* C *satisfying* $g(C) = 0$ *is a purely diagonal matrix* [for in (6) the only possible value of p is then $p = 1$].

2.2. *Elementary examples*

Take $$g(x) \equiv x^2-1.$$

The roots of $x^2-1 = 0$ are ± 1 and the general solution of

$$C^2 = I$$

is
$$\begin{aligned} C &= \text{diag}\{C_1(\pm 1), C_1(\pm 1),...\} \\ &= \text{diag}\{\pm 1, \pm 1,...\}, \end{aligned} \tag{8}$$

there being as many entries as we need to make up the desired order of C and the signs being completely arbitrary.

The general solution‡ of $A^2 = I$ is given by

$$A = T\,\text{diag}(\pm 1, \pm 1,...)T^{-1},$$

where T is non-singular, but otherwise arbitrary.

Again, take $$g(x) \equiv (x^2-1)^2.$$

The general solution of $g(C) = 0$ is

$$C = \text{diag}\{C_1(\pm 1),..., C_2(\pm 1),...\},$$

† If C is the canonical form of A and $g(A) = 0$, then $g(C) = 0$.

‡ If C is the canonical form of A and $A^2 = I$, then $C^2 = I$.

there being sufficient entries to make up any desired order for C. For example, when C is required to be of order 4, one solution of $(C^2-I)^2 = 0$ is

$$C = \operatorname{diag}\{1, -1, C_2(1)\}.$$

It may be noted that

$$C^2 = \operatorname{diag}\left\{1, \quad 1, \quad \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}\right\},$$

so that
$$C^2-I = \operatorname{diag}\left\{0, \quad 0, \quad \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix}\right\}.$$

The square of C^2-I is zero, though, of course, C^2-I is not itself zero.

Finally, take
$$g(x) = x^2.$$

The only root of $x^2 = 0$ is a repeated root, zero. The general solution of $C^2 = 0$ is, therefore,

$$C = \operatorname{diag}\{C_1(0),..., C_2(0),...\}.$$

The general solution of the equation

$$A^2 = 0$$

is given by

$$A = T \operatorname{diag}\left\{0, \quad ..., \quad \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad ...\right\} T^{-1},$$

there being as many entries of either kind as we wish in the diagonal matrix and T being an arbitrary non-singular matrix of the same order as the diagonal matrix. The matrix C above will be itself the null matrix unless it contains a $C_2(0)$; if $C \neq 0$ and $C^2 = 0$, the form C above must contain at least one $C_2(0)$.

The reader will see for himself that the general solutions of $A^m = I$ or of $A^m = 0$ along these lines is merely a matter of writing out the details.

2.21. *Idem-potent and nil-potent matrices.* A matrix whose square is zero is said to be NIL-POTENT. A matrix A which satisfies the equation $A^2 = A$ is said to be IDEM-POTENT; since the roots of

$$x^2-x = 0$$

are 0, 1 and are not repeated; the general solution of

$$C^2-C = 0$$

is
$$C = \operatorname{diag}\{C_1(0),\ldots, C_1(1),\ldots\}$$
and of
$$A^2 - A = 0$$
is
$$A = TCT^{-1} = T\operatorname{diag}\{C_1(0),\ldots, C_1(1),\ldots\}T^{-1}.$$

2.3. *Solutions in rational canonical form*

We recall from Chapter IV, § 10, that when
$$p(\lambda) \equiv \lambda^r + p_1\lambda^{r-1} + \ldots + p_r$$
and
$$P = \begin{bmatrix} 0 & 1 & 0 & . & . & 0 \\ 0 & 0 & 1 & . & . & 0 \\ . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & 1 \\ -p_r & -p_{r-1} & -p_{r-2} & . & . & -p_1 \end{bmatrix},$$
the matrix P is said to be ASSOCIATED WITH the polynomial $p(\lambda)$ and
$$|\lambda I - P| = p(\lambda).$$

We proved, in Theorem 17, that when a square matrix A is given and the invariant factors, other than unity, of $\lambda I - A$ are
$$E_s(\lambda),\quad E_{s+1}(\lambda),\quad \ldots,\quad E_n(\lambda),$$
the matrix
$$E = \operatorname{diag}(M_s,\ldots, M_n), \tag{9}$$
where M_i is the matrix associated with the polynomial $E_i(\lambda)$, is a transform of A; that is, for some matrix T,
$$A = TET^{-1}.$$

Given a polynomial $g(x)$, we seek a matrix E, of type (9), that shall be of a given order M and satisfy the equation
$$g(E) = 0. \tag{10}$$
When we have found such a solution, the matrix $A = TET^{-1}$ will satisfy the equation $g(A) = 0$.

Now, as we saw in § 1, the matrix $g(E)$ is zero if and only if $g(\lambda)$ contains $E_n(\lambda)$ as a factor. Thus, to obtain a solution of type (9) that satisfies $g(E) = 0$ we take

$E_n(\lambda)$ to be a factor of $g(\lambda)$,

$E_{n-1}(\lambda)$ to be a factor of $E_n(\lambda)$,

$E_{n-2}(\lambda)$ to be a factor of $E_{n-1}(\lambda)$, and so on; let E_s be the last E_{n-k} to be other than unity. Then the matrix

$$E = \text{diag}(M_s, M_{s+1},..., M_n)$$

is

(i) a matrix in rational canonical form,

(ii) a solution of the equation $g(X) = 0$.

The order of E is the sum of the degrees in λ of the polynomials $E_s(\lambda),..., E_n(\lambda)$. If it is possible so to choose the degrees of these factors that their sum is M, we have then obtained a solution of the order required; otherwise no solution in rational canonical form is possible.

For example, take F to be the field of real numbers and

$$g(x) \equiv x^2+1.$$

The matrix associated with

$$p(\lambda) \equiv \lambda^2+0.\lambda+1$$

is

$$P = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

This matrix P, of order two, satisfies the equation $P^2+I = 0$. So too does the matrix, of order 4, obtained by taking

$$E_n(\lambda) = \lambda^2+1, \quad E_{n-1}(\lambda) = \lambda^2+1, \quad E_{n-2} = 1:$$

this matrix is

$$Q = \text{diag}(P, P)$$

and $Q^2+I = 0$; but, since λ^2+1 has no linear factor in the field of real numbers, it is not possible thus to find, in rational canonical form, a matrix of odd order† that will satisfy

$$X^2+I = 0.$$

The whole process, though theoretically complete, is much more tentative and ragged than the corresponding solution in classical canonical form.

2.4. *Note on odd and even order*

In reflecting on the difficulty of obtaining directly a third-order matrix to satisfy $X^2 = -I$ one is tempted to put in an extra factor and to solve $X(X^2+I) = 0$. This would lead us to take

$$E_n(\lambda) = \lambda^3+\lambda$$

† The method gives a matrix of odd order satisfying $g(X) = 0$ only if $g(x)$ has, for factorizations within the field F, a factor of odd degree.

and
$$E = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}.$$

The reader will find, however, that E satisfies $E^3 = -E$ but does not satisfy $E^2 = -I$. This is necessarily so; for E^2+I can be zero only if $E_n(\lambda)$ is a factor of λ^2+1 and therefore, as soon as we put in an additional λ and take $E_n(\lambda) = \lambda(\lambda^2+1)$, we make it impossible for $E_n(\lambda)$ to be a factor of λ^2+1.

3. The equation $g(X) = A$

Let A be a given square matrix and $C = TAT^{-1}$ its classical canonical form. We seek a matrix X to satisfy the equation

$$g(X) = A. \tag{1}$$

Our first step is to put $X = T^{-1}YT$, so that (1) becomes

$$T^{-1}g(Y)T = T^{-1}CT,$$

that is
$$g(Y) = C. \tag{2}$$

We shall indicate, rather than develop to the last detail, methods of determining whether (2) has a solution and of finding the solutions when they exist.†

3.1. *Rutherford's method*

We have to find Y so that, $g(x)$ being a given polynomial in x and C a given canonical form,

$$g(Y) = C.$$

Let Z be any canonical matrix, say

$$Z = \text{diag}\{C_r(\lambda),...\}. \tag{3}$$

Then $g(Z)$ is given by

$$g(Z) = \text{diag}\{G_r(\lambda),...\}, \tag{4}$$

where
$$G_r(\lambda) = \begin{bmatrix} g(\lambda) & g'(\lambda) & . & . & g^{(r-1)}(\lambda)/(r-1)! \\ . & . & . & . & . \\ 0 & 0 & . & . & g(\lambda) \end{bmatrix}. \tag{5}$$

The matrix (4) is not a classical canonical form, save in very

† C. C. MacDuffee, *The Theory of Matrices* (Chelsea Publishing Co., 1946), p. 96, gives several references.

special circumstances. Let its classical canonical form be G. Then, for some non-singular F,

$$G = Fg(Z)F^{-1} = g(FZF^{-1}).$$

If, by an appropriate choice of the λ's and r's in (3) we can identify G with C, we shall have found a matrix $Y = FZF^{-1}$ that satisfies $g(Y) = C$. If such a choice is impossible, $g(Y) = C$ has no solution.

When $g^{(k)}(\lambda)$ is the first of $g'(\lambda)$, $g''(\lambda)$,... to differ from zero, we denote the classical canonical form of (5) by $C_r\{g(\lambda)\}_k$. In this notation

$$G = \text{diag}[C_r\{g(\lambda)\}_k,\ldots] \tag{6}$$

and solutions of (2) are sought by identifying (6) with C, remembering that $C_r(\alpha)_k$ is simply $C_r(\alpha)$ when $k = 1$ and that, when $k > 1$ and we write $r = pk+q$ with $0 \leqslant q < k$, $C_r(\alpha)_k$ consists of $C_p(\alpha)$ repeated $k-q$ times and $C_{p+1}(\alpha)$ repeated q times.†

Further details, including a worked example, may be found in the original paper by D. E. Rutherford.‡

3.2. *The equation* $X^2 = A$

We consider

$$Y^2 = C,$$

where C is the classical canonical form of A, say

$$C = \text{diag}\{C_r(\mu),\ldots\}. \tag{7}$$

METHOD 1. Provided that C is non-singular, our work on functions of matrices enables us to write down a value of Y. As in Chapter V, § 8.5, let

$$\{C_r(\mu)\}^x = \begin{bmatrix} \mu^x & x\mu^{x-1} & \frac{1}{2}x(x-1)\mu^{x-2} & . & . & . \\ & \mu^x & x\mu^{x-1} & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & \mu^x \end{bmatrix},$$

there being, of course, r rows and columns. Then one solution of $Y^2 = C$ is given by§

$$Y = \text{diag}[\{C_r(\mu)\}^{\frac{1}{2}},\ldots].$$

† Cf. Chapter V, § 9.

‡ *Proc. Edinburgh Math. Soc.* (2) **3** (1932), 135–43.

§ For a more general solution see § 3.3 (*post*).

This throws no light on what happens when C is singular, but it does show that

Given a non-singular square matrix A, there is a matrix X whose square is A.

METHOD 2. This follows Rutherford's method and when C is non-singular gives, in fact, the same solution as Method 1.

(i) Let C be non-singular; then no μ in (7) is zero.

Put $$Z = \operatorname{diag}\{C_r(\sqrt{\mu}),...\}. \tag{8}$$

Then $$Z^2 = \operatorname{diag}\{Z_r(\mu),...\},$$

where $$Z_r(\mu) = \begin{bmatrix} \mu & 2\sqrt{\mu} & 1 & . & . & . \\ 0 & \mu & 2\sqrt{\mu} & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & \mu \end{bmatrix}. \tag{9}$$

Since $\mu \neq 0$, the canonical form of (9) is $C_r(\mu)$ and the canonical form of Z^2 is $\operatorname{diag}\{C_r(\mu),...\}$. If F is a matrix which transforms Z^2 into its canonical form, that is, if F is chosen so that

$$FZ^2F^{-1} = \operatorname{diag}\{C_r(\mu),...\} \tag{10}$$

and we put $$FZF^{-1} = Y,$$

then $$Y^2 = FZ^2F^{-1} = \operatorname{diag}\{C_r(\mu),...\}$$

and Y satisfies the equation $Y^2 = C$.

To show that this gives the same solution as Method 1, we proceed thus:

The canonical form of

$$\operatorname{diag}[\{C_r(\mu)\}^{\frac{1}{2}},...] \quad \text{is} \quad \operatorname{diag}\{C_r(\sqrt{\mu}),...\},$$

because no μ is equal to zero. Let

$$\operatorname{diag}[\{C_r(\mu)\}^{\frac{1}{2}},...] = F\operatorname{diag}\{C_r(\sqrt{\mu}),...\}F^{-1} = FZF^{-1},$$

where Z is defined by (8). On squaring this,

$$\operatorname{diag}\{C_r(\mu),...\} = FZ^2F^{-1}.$$

That is to say, a matrix F which changes Z^2 into its canonical form $\operatorname{diag}\{C_r(\mu),...\}$ makes FZF^{-1}, which is the solution given by Method 2, equal to $\operatorname{diag}[\{C_r(\mu)\}^{\frac{1}{2}},...]$, the solution given by Method 1.

(ii) Let C be singular, say

$$C = \operatorname{diag}\{C_{\rho_1}(0),\ldots, C_{\rho_m}(0), C_\pi(\mu),\ldots\}, \tag{11}$$

where $C_\pi(\mu)$ is a typical submatrix with $\mu \neq 0$. Let $\{C_\pi(\mu)\}^{\frac{1}{2}}$ be defined as in Method 1 and let

$$Z = \operatorname{diag}[C_{2r}(0),\ldots, C_{2s+1}(0),\ldots, \{C_\pi(\mu)\}^{\frac{1}{2}},\ldots],$$

where $2r$ denotes a typical even, $2s+1$ a typical odd suffix. Now the canonical forms of the squares of $C_{2r}(0)$, $C_{2s+1}(0)$ are†

$$\operatorname{diag}\{C_r(0), C_r(0)\}, \qquad \operatorname{diag}\{C_s(0), C_{s+1}(0)\}$$

respectively. Hence, *there is a solution of $Y^2 = C$ if, and only if, in the form* (11) *for* C

(i) *m is even*

and (ii) *the suffixes $\rho_1,\ldots, \rho_m$ can be arranged in pairs in such a way that the two members of each pair are either equal or differ by unity.*

As a simple example, let us determine a matrix whose square is equal to

$$C = \operatorname{diag}\{C_1(0), C_2(0), C_p(\lambda)\} \quad (\lambda \neq 0).$$

Let
$$Z = \operatorname{diag}[C_3(0), \{C_p(\lambda)\}^{\frac{1}{2}}];$$

then
$$Z^2 = \operatorname{diag}\left\{\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad C_p(\lambda)\right\}.$$

The first submatrix needs only a collineation, interchange of the first two rows and then of the first two columns, to identify Z^2 with C: to obtain a solution of $Y^2 = C$ we subject Z to the same collineation and write

$$Y = \operatorname{diag}\left\{\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \{C_p(\lambda)\}^{\frac{1}{2}}\right\},$$

the first submatrix being the result of interchanging the first two rows and then the first two columns in $C_3(0)$. A direct

† Chapter V, § 9, using $k = 2$ and (i) $p = r$, $q = 0$, (ii) $p = s$, $q = 1$; or the result is evident at once from the chains of non-zeros in (9) when $\mu = 0$ and (i) r is $2r$, (ii) r is $2s+1$.

verification gives

$$Y^2 = \operatorname{diag}\left\{\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 1\\ 0 & 0 & 0\end{bmatrix},\quad C_p(\lambda)\right\}$$

$$= \operatorname{diag}\{C_1(0), C_2(0), C_p(\lambda)\}.$$

In a way similar to (ii) above we can find conditions that, given a singular matrix C, there should be a matrix X for which $X^k = C$; when C is non-singular, one solution of $X^k = C$ can be written down by using $x = 1/k$ in Method 1.

METHOD 3. Let $\qquad |\lambda I - A| \equiv f(\lambda).$

Then it can be proved, by the processes of ordinary algebra, that, provided the constant term in $f(\lambda)$ is not zero, there is a polynomial $g(\lambda)$ for which $\{g(\lambda)\}^2 - \lambda$ contains $f(\lambda)$ as a factor. There is then an identity

$$\{g(\lambda)\}^2 - \lambda \equiv f(\lambda)q(\lambda)$$

and, since $f(A) = 0$, it follows that

$$\{g(A)\}^2 - A = 0.$$

The method is given in greater detail in at least two well-known books.†

3.3. *A general solution of $g(Y) = C$*

Suppose that we have found any one matrix Y that satisfies the equation $g(Y) = C$. Let K be any non-singular matrix that commutes‡ with C; then $KC = CK$, that is

$$KCK^{-1} = C.$$

Put $M = KYK^{-1}$. Then

$$g(M) = Kg(Y)K^{-1} = KCK^{-1} = C$$

and M is also a solution of the given equation.

3.4. *The reversion of a power series*

One solution of the equation

$$g_0 I + g_1 Y + \ldots + g_m Y^m = C$$

† L. E. Dickson, *Modern Algebraic Theories*, p. 120; W. V. D. Hodge and D. Pedoe, *Methods of Algebraic Geometry* (Cambridge, 1947), pp. 96, 137.

‡ The general form for such a matrix K is given by Turnbull and Aitken, loc. cit., pp. 146, 147.

can be obtained by expressing Y as a power series in $C-g_0 I$ provided $g_1 \neq 0$ and the latent roots of $C-g_0 I$ are sufficiently small (cf. Chap. V, § 11.5). This solution tallies with the solution given by Rutherford's method (§ 3.1) in the following way:

If $g_1 \neq 0$ and $g(x) = g_1 x+...+g_m x^m$, then

$$g'(x) = g_1+2g_2 x+... \quad \text{and} \quad g'(x) \neq 0$$

when $|x|$ is sufficiently small. Let

$$C-g_0 I = Z = \text{diag}\{C_p(\lambda),...\},$$

and take $X = \text{diag}\{x,...\}$, where x is the root of

$$g_1 x+...+g_m x^m = \lambda$$

which tends to zero as λ tends to zero. The canonical form of the matrix

$$g_1 X+g_2 X^2+...+g_m X^m$$

is then also $\text{diag}\{C_p(\lambda),...\}$ provided each $g'(x) \neq 0$. Thus F can be determined so that

$$F(g_1 X+...+g_m X^m)F^{-1} = Z = \text{diag}\{C_p(\lambda),...\}$$

and $Y = FXF^{-1}$ gives $g(Y) = C$.

Equally, as Rutherford's method shows, there may be no solution once we relax the condition that the latent roots of $C-g_0 I$ are 'sufficiently small'; for when we relax this condition, $g'(x)$ may become zero and the canonical form of

$$g_1 X+...+g_m X^m$$

become different from the canonical form of $C-g_0 I$.

4. Scalar equations deduced from matrix equations

It is well known that a square matrix A satisfies its own characteristic equation. One method of proving this can be generalized† so as to prove

THEOREM 38. *Let $C_0,..., C_k$ be given n-th order matrices, and A an n-th order matrix for which*

$$C_0 A^k+C_1 A^{k-1}+...+C_k = 0. \tag{1}$$

Let $P(\lambda)$ be the expanded form of the determinant

$$|C_0 \lambda^k+C_1 \lambda^{k-1}+...+C_k|. \tag{2}$$

Then $P(A) = 0$.

† MacDuffee, loc. cit., pp. 17, 18.

We defer the proof until § 4.1. When $k = 1$, $C_0 = I$ and $C_k = -A$, equation (1) is simply

$$IA-A = 0,$$

the determinant (2) is $|\lambda I - A|$, and $P(\lambda) = 0$ is the characteristic equation of A. Thus Theorem 38 contains, as a particular example, the fact that any square matrix satisfies its own characteristic equation.

In its turn, Theorem 38 is a particular example of a still more general theorem, which we shall now prove.

4.1. *Phillips' Theorem*

This theorem was first proved by H. B. Phillips,† who applied it to establish other theorems. I know of no later applications, but the generality of the theorem is, in itself, striking.

THEOREM 39. *Let $A = [a_{ik}]$,..., $P = [p_{ik}]$ be n-th order matrices and let λ,..., ρ be scalars. Let A',..., P' be matrices that are commutative with each other and let*

$$AA'+...+PP' = 0. \tag{3}$$

Then A',..., P' satisfy the n-th degree equation obtained by writing

$$|a_{ik}\lambda+...+p_{ik}\rho| = 0 \tag{4}$$

and, in the expanded form of the determinant, replacing λ,..., ρ by the matrices A',..., P'.

PROOF. Let E_{ik} be the nth order matrix which has unity at the cross of the ith row and kth column and zeros elsewhere. Then

$$E_{ij}E_{jk} = E_{ik}, \qquad E_{ij}E_{lk} = 0 \quad (l \neq j). \tag{5}$$

Now

$$A = \sum_{i,k} a_{ik}E_{ik}$$

and so (3) may be written as

$$\sum_{i,k} E_{ik}(a_{ik}A'+...+p_{ik}P') = 0.$$

Multiply this (on the left) by E_{11},..., E_{1n} in turn and use (5): the results are

$$E_{11}(a_{11}A'+...+p_{11}P')+...+E_{1n}(a_{1n}A'+...+p_{1n}P') = 0,$$
$$E_{11}(a_{21}A'+...+p_{21}P')+...+E_{1n}(a_{2n}A'+...+p_{2n}P') = 0,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$E_{11}(a_{n1}A'+...+p_{n1}P')+...+E_{1n}(a_{nn}A'+...+p_{nn}P') = 0.$$

† *Amer. J. of Math.* **41** (1919), 266–78.

Let Δ denote the polynomial in $A',\ldots, P'$ obtained by expanding the determinant $$|\lambda a_{ik}+\ldots+\rho p_{ik}|$$
and replacing $\lambda,\ldots, \rho$ by $A',\ldots, P'$. Then, as in scalar algebra, we obtain† from the above n linear equations in $E_{11},\ldots, E_{1n}$ the results $$E_{11}\Delta = 0, \quad E_{12}\Delta = 0, \quad \ldots, \quad E_{1n}\Delta = 0. \tag{6}$$

Now in (6) Δ is a matrix: the first equation, $E_{11}\Delta = 0$, implies that the first row of Δ is composed solely of zeros; the second, $E_{12}\Delta = 0$, implies that the second row of Δ is all zeros; and so on. Hence, by (6), every row of Δ is a row of zeros and Δ is the null matrix, which proves the theorem.

Theorem 38 is obtained by writing $A^k, A^{k-1},\ldots, I$ in place of $A', B',\ldots, P'$ and $C_0, C_1,\ldots, C_k$ in place of $A, B,\ldots, P$.

5. A general type of matrix equation

Let $A_1,\ldots, A_p$ be matrices having m rows and n columns and X a matrix of n rows and columns. Then a type of matrix equation is
$$\sum_{i=0}^{p} A_i X^{p-i} = 0.$$
Its solution has been considered by W. E. Roth.‡ We shall not develop an account of the various types of matrix equation which have yielded to treatment, but refer the reader to Roth's paper and to the account of matrix equations given by MacDuffee.§ Our own treatment of the topic, in §§ 1–4, has been limited to what comes easily and naturally from the work of earlier chapters.

6. A note on linear equations

The proof given by H. B. Phillips of Theorem 39 leads one to consider a set of n linear equations
$$\begin{aligned} X_1 A_{11}+\ldots+X_n A_{1n} &= 0, \\ \cdots\cdots & \\ X_1 A_{n1}+\ldots+X_n A_{nn} &= 0, \end{aligned} \tag{1}$$

† The fact that the matrices $A',\ldots, P'$ are commutative among themselves enables us to multiply by the 'cofactors' of the first, second,... columns of 'coefficients' in the linear equations and add just as we should do if $A',\ldots, P'$ were scalars.

‡ *Trans. Amer. Math. Soc.* **32** (1929), 61–80.

§ *The Theory of Matrices* (Chelsea Publishing Company, New York, 1946).

where the A_{rs} are $n \times n$ matrices that are commutative among themselves and the X_r matrices of m rows and n columns. This set of equations, with its obvious analogy to a set of linear scalar equations, sets the problem 'How far does the analogy go?' This problem may repay investigation, but I have not pursued it here.†

The one easy result that a first consideration of the problem gives is this:

Let Δ be the matrix obtained by writing A_{rs} for a_{rs} in the expanded form of the determinant $|a_{rs}|$. Then, if (1) is satisfied by a set of matrices $X_1,..., X_n$ containing among them n linearly independent rows, $\Delta = 0$. The proof is immediate: the equations (1) imply

$$X_1\Delta = X_2\Delta = ... = X_n\Delta = 0, \tag{2}$$

and, if Y is the $n \times n$ matrix formed by any n rows selected from among the mn rows of $X_1,..., X_n$, (2) implies

$$Y\Delta = 0;$$

if the n selected rows forming Y are linearly independent, Y has a reciprocal Y^{-1} and $\Delta = 0$.

† It might be suitable as part of a B.Sc. thesis: it is unlikely to yield results of sufficient weight for a doctorate thesis.

CHAPTER VIII

MISCELLANEOUS NOTES

1. The resolvent of a matrix

LET A be a given matrix of order n. Then the matrix

$$R(\lambda) = (\lambda I - A)^{-1}, \tag{1}$$

where λ is a scalar variable, is called the RESOLVENT of A. This function and a related function using $I - \lambda A$ have been extensively used in the study of matrices, linear equations, and bilinear forms. Our own development of these subjects does not use the resolvent. It may be of interest, however, to note some of the properties of this function; the best known are

(I) *When $|\lambda|$ is greater than the absolute value of any latent root of A,*

$$\frac{1}{\lambda I - A} = \frac{1}{\lambda}\left\{I + \frac{A}{\lambda} + \frac{A^2}{\lambda^2} + \dots\right\}. \tag{2}$$

(II) *Provided that λ is not equal to a latent root of A, the matrix $(\lambda I - A)^{-1}$ may be written as a sum of partial fractions.*

When E_n, the nth invariant factor of $\lambda I - A$, is of degree m ($\leqslant n$) in λ and has m distinct zeros $\alpha, \dots, \kappa$ the partial fraction form is

$$\frac{1}{\lambda I - A} = \frac{R_\alpha}{\lambda - \alpha} + \dots + \frac{R_\kappa}{\lambda - \kappa}, \tag{3}$$

wherein $R_\alpha, \dots, R_\kappa$ are polynomials in the matrix A.

When E_n has repeated zeros, the partial fraction form is that associated with repeated factors in scalar algebra; for example, if

$$E_n \equiv (\lambda - \alpha)^3 (\lambda - \beta) \dots (\lambda - \kappa),$$

$$\frac{1}{\lambda I - A} = \frac{R_{\alpha 1}}{\lambda - \alpha} + \frac{R_{\alpha 2}}{(\lambda - \alpha)^2} + \frac{R_{\alpha 3}}{(\lambda - \alpha)^3} + \frac{R_\beta}{\lambda - \beta} + \dots, \tag{4}$$

where each numerator is a polynomial in A.

(III) *When $\alpha, \dots, \kappa$ are distinct, as in* (3) *above, each numerator in* (3) *is an idem-potent matrix (i.e. $X^2 = X$) and the product of two different numerators is zero; for example,*

$$R_\alpha^2 = R_\alpha, \qquad R_\alpha R_\beta = 0.$$

This is sometimes referred to as 'the orthogonal property' of the numerators.

(IV) *Let* $\mathbf{y}$ *be a given single-column matrix,* A *a given square matrix, and* λ *a scalar. Then the single-column matrix* $\mathbf{x}$ *satisfying*

$$\lambda\mathbf{x}-A\mathbf{x} = \mathbf{y}$$

is given by

$$\mathbf{x} = (\lambda I-A)^{-1}\mathbf{y}.$$

(V) *When* $R(\lambda) = (\lambda I-A)^{-1}$,

$$(\lambda-\mu)R(\lambda)R(\mu) = R(\mu)-R(\lambda).$$

1.1. *Proof of (I)*

When $|z| < |\lambda|$,

$$\frac{1}{\lambda-z} = \frac{1}{\lambda}\left\{1+\frac{z}{\lambda}+\frac{z^2}{\lambda^2}+\dots\right\}.$$

Hence, by Chapter V, § 8.1, if the latent roots of A are less than $|\lambda|$ in absolute value, the series

$$\frac{1}{\lambda}\sum_{r=0}^{\infty}(A^r/\lambda^r)$$

is convergent and its sum is the matrix $(\lambda I-A)^{-1}$.

1.2. *Proof of (II)*

This property is an example of the general theorem that a power series in a matrix A can be expressed as a polynomial in A. This theorem proves the result (II) when $|\lambda|$ is sufficiently large to permit $(\lambda I-A)^{-1}$ to be expressed as a power series in A and an elementary 'permanence of algebraic identities' argument then shows that the result remains true provided only that λ is not a latent root of A.

(i) Let E_n, the nth invariant factor of $\lambda I-A$, be of degree m and have m distinct factors $\lambda-\alpha,\dots,\lambda-\kappa$. When $|\lambda|$ exceeds each of $|\alpha|,\dots,|\kappa|$,

$$\frac{1}{\lambda I-A} = \frac{1}{\lambda}\left\{I+\frac{A}{\lambda}+\frac{A^2}{\lambda^2}+\dots\right\}. \tag{5}$$

Put $f(A) = (\lambda I-A)^{-1}$ in Theorem 22 (p. 94); this gives

$$\frac{1}{\lambda I-A} = \sum_{\alpha}\frac{1}{\lambda-\alpha}\frac{(A-\beta I)\dots(A-\kappa I)}{(\alpha-\beta)\dots(\alpha-\kappa)}. \tag{6}$$

Thus (II) is proved when $|\lambda|$ is sufficiently large and, for such values of $|\lambda|$, (6) gives

$$I = (\lambda I - A)\sum_{\alpha}\frac{1}{\lambda-\alpha}\,\frac{(A-\beta I)...(A-\kappa I)}{(\alpha-\beta)...(\alpha-\kappa)}. \tag{7}$$

For convenience of writing, let

$$\psi(\lambda) = (\lambda-\alpha)(\lambda-\beta)...(\lambda-\kappa),$$

so that $$\psi'(\alpha) = (\alpha-\beta)...(\alpha-\kappa);$$

and let $$\psi(\lambda)_\alpha = (\lambda-\beta)...(\lambda-\kappa).$$

The equation (7) then is

$$\psi(\lambda)I = (\lambda I-A)\sum_{\alpha}\psi(\lambda)_\alpha\,\psi(A)_\alpha/\psi'(\alpha), \tag{8}$$

which, as an equation in λ, is of degree m. It is satisfied by an infinity of values of λ and is therefore† true for all λ. When $\lambda \neq \alpha,..., \kappa$, the matrix $\lambda I - A$, which has latent roots $\lambda-\alpha,...,$ $\lambda-\kappa$, is non-singular and, on dividing (8) by $(\lambda I-A)\psi(\lambda)$,

$$\frac{1}{\lambda I-A} = \sum_{\alpha}\frac{1}{\lambda-\alpha}\,\frac{(A-\beta I)...(A-\kappa I)}{(\alpha-\beta)...(\alpha-\kappa)}.$$

(ii) Let $E_n = (\lambda-\alpha)^3(\lambda-\beta)...(\lambda-\kappa)$. When $|\lambda|$ is large enough, expansion (5) holds and, by Chapter V, § 6.21 (p. 96),

$$\begin{vmatrix} 0 & 0 & . & . & (m-1)(m-2)\alpha^{m-3} & 2(\lambda-\alpha)^{-3} \\ 0 & 1 & . & . & (m-1)\alpha^{m-2} & (\lambda-\alpha)^{-2} \\ 1 & \alpha & . & . & \alpha^{m-1} & (\lambda-\alpha)^{-1} \\ . & . & . & . & . & . \\ 1 & \kappa & . & . & \kappa^{m-1} & (\lambda-\kappa)^{-1} \\ 1 & A & . & . & A^{m-1} & (\lambda I-A)^{-1} \end{vmatrix} = 0,$$

where m is the degree in λ of E_n.

On expanding this determinant by its last column we obtain (4), which is thus proved when $|\lambda|$ is sufficiently large. The proof for all λ not equal to $\alpha, \beta,..., \kappa$ follows as in (i).

1.3. *Proof of (III)*

When $$E_n = (\lambda-\alpha)...(\lambda-\kappa)$$

† This is the argument of elementary scalar algebra applied to each of the n^2 scalar equations which make up the one matrix equation (8).

the matrix A satisfies the equation†

$$(A-\alpha I)...(A-\kappa I) = 0. \tag{9}$$

The property (III) is an immediate consequence of this and of the fact that the matrices $A-\alpha I,..., A-\kappa I$ are commutative.

We have seen, in (6) of § 1.2, that

$$R_\alpha = \frac{(A-\beta I)...(A-\kappa I)}{(\alpha-\beta)...(\alpha-\kappa)}, \quad R_\beta = \frac{(A-\alpha I)(A-\gamma I)...}{(\beta-\alpha)(\beta-\gamma)...}, \quad ....$$

The product of any two of R_α, R_β,... contains

$$(A-\alpha I)...(A-\kappa I)$$

as a factor, and so is the null matrix. Again

$$\frac{(x-\beta)...(x-\kappa)}{(\alpha-\beta)...(\alpha-\kappa)} - 1 \equiv (x-\alpha)P(x),$$

where $P(x)$ is a polynomial in x. Hence $R_\alpha - I$ contains $A-\alpha I$ as a factor and, therefore,

$$R_\alpha(R_\alpha - I)$$

contains the factor $(A-\alpha I)...(A-\kappa I)$, which is the null matrix. Hence $R_\alpha^2 = R_\alpha$.

1.4. *Proof of (IV) and (V)*

(IV) When $$\lambda\mathbf{x} - A\mathbf{x} = \mathbf{y}$$

and λ is a scalar, we can at once write

$$(\lambda I - A)\mathbf{x} = \mathbf{y}.$$

Provided that λ is not a latent root of A, this equation has a unique solution $$\mathbf{x} = (\lambda I - A)^{-1}\mathbf{y}.$$

(V) Our work on functions of matrices, and particularly the observations of Chapter V, §§ 8.4 and 8.7, shows that the result

$$(\lambda-\mu)R(\lambda)R(\mu) = R(\mu) - R(\lambda)$$

is obtained by handling the algebraic processes in

$$\frac{1}{\mu I - A} - \frac{1}{\lambda I - A} = \frac{(\lambda-\mu)I}{(\lambda I - A)(\mu I - A)}$$

as though the matrices I and A were scalars.

† Cf. Chap. VII, § 1, p. 153.

A proof that avoids such processes is almost as simple. By definition,

$$(\lambda I - A)R(\lambda) = I, \qquad (\mu I - A)R(\mu) = I;$$

whence

$$\lambda R(\lambda) - I = AR(\lambda),$$
$$\mu R(\mu) - I = AR(\mu),$$

and so, the matrices involved being commutative,

$$(\lambda - \mu)R(\lambda)R(\mu) - R(\mu) + R(\lambda) = 0.$$

2. Positive-definite quadratic forms subject to linear conditions

We propose to find a set of necessary and sufficient conditions that the real quadratic form

$$\sum_{r,s=1}^{n} a_{rs} x_r x_s \tag{1}$$

shall be positive when the x_r are subject to one or more given linear conditions of the type

$$p_1 x_1 + \ldots + p_n x_n = 0, \tag{2}$$

and at least one x_r is not zero. The general form of these conditions is 'well known' in the sense that one finds references to it in books on Differential Calculus,† in mathematics for economists, and in examination papers. I have never found a printed proof of the result for an arbitrary positive integer n, even with only one restrictive condition (2). I therefore include among these notes‡ a proof in as simple a form as possible. I first work with one linear condition and then indicate, by considering two linear conditions, how the problem is solved when there are m linear conditions.

2.1. *One linear condition*

We consider (1) when the x_r are subject to the condition

$$p_1 x_1 + \ldots + p_n x_n = 0.$$

Unless the problem is to be unaffected by the condition, one at least of the p_r is not zero. We assume that $p_n \neq 0$.

† T. W. Chaundy, *The Differential Calculus* (Oxford, 1935), p. 258. See also (21) on p. 262 for the way in which m linear conditions introduce a sign-factor $(-1)^m$.

‡ S. N. Afriat, *Proc. Camb. Phil. Soc.* **47** (1951), deals with the problem in a general form.

If $x_1,..., x_{n-1}$ are all zero, then also $x_n = 0$; the requirement that one at least of $x_1,..., x_n$ be not zero is fulfilled only if one of $x_1,..., x_{n-1}$ is not zero.

Make the transformation

$$X_r = x_r \quad (r = 1,..., n-1), \tag{3a}$$

$$X_n = p_1 x_1 + ... + p_n x_n \tag{3b}$$

and let (1) become

$$\sum_{r,s=1}^{n} \alpha_{rs} X_r X_s. \tag{4}$$

Since $X_n \equiv 0$ and one at least of $X_1,..., X_{n-1}$ is not zero, our problem reduces to that of finding conditions that the form in $n-1$ variables

$$\sum_{r,s=1}^{n-1} \alpha_{rs} X_r X_s \tag{5}$$

be positive-definite. One set of conditions, necessary and sufficient for (5) to be positive-definite, is

$$\alpha_{11} > 0, \quad \begin{vmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{vmatrix} > 0, \quad ..., \quad \begin{vmatrix} \alpha_{11} & . & . & \alpha_{1,n-1} \\ . & . & . & . \\ \alpha_{n-1,1} & . & . & \alpha_{n-1,n-1} \end{vmatrix} > 0.$$

It remains only to express these conditions in terms of the a_{rs} and p_r. Introduce a further variable x_{n+1} and consider the form in $n+1$ variables

$$\sum_{r,s=1}^{n} a_{rs} x_r x_s + 2(p_1 x_1 + ... + p_n x_n) x_{n+1}. \tag{6}$$

Transform this by means of (3) and $X_{n+1} = x_{n+1}$. In view of (4) it becomes

$$\sum_{r,s=1}^{n} \alpha_{rs} X_r X_s + 2X_n X_{n+1}. \tag{7}$$

The modulus of the transformation is p_n and so†

$$p_n^2 \begin{vmatrix} a_{11} & . & . & a_{1n} & p_1 \\ . & . & . & . & . \\ a_{n1} & . & . & a_{nn} & p_n \\ p_1 & . & . & p_n & 0 \end{vmatrix} = \begin{vmatrix} \alpha_{11} & . & . & \alpha_{1n} & 0 \\ . & . & . & . & . \\ \alpha_{n1} & . & . & \alpha_{nn} & 1 \\ 0 & . & . & 1 & 0 \end{vmatrix}$$

$$= -\begin{vmatrix} \alpha_{11} & . & . & \alpha_{1,n-1} \\ . & . & . & . \\ \alpha_{n-1,1} & . & . & \alpha_{n-1,n-1} \end{vmatrix}. \tag{8}$$

† F. 127, Theorem 39.

Now consider (6) with $x_{n-1} = 0$. It is a form in the n variables $x_1,\ldots, x_{n-2}, x_n, x_{n+1}$. The transformation (3), leaving out $x_{n-1} = X_{n-1}$, has modulus p_n and transforms (6) with $x_{n-1} = 0$ into (7) with $X_{n-1} = 0$. Hence

$$p_n^2 \begin{vmatrix} a_{11} & . & . & a_{1,n-2} & a_{1,n} & p_1 \\ . & . & . & . & . & . \\ a_{n-2,1} & . & . & a_{n-2,n-2} & a_{n-2,n} & p_{n-2} \\ a_{n,1} & . & . & a_{n,n-2} & a_{n,n} & p_n \\ p_1 & . & . & p_{n-2} & p_n & 0 \end{vmatrix}$$

$$= \begin{vmatrix} \alpha_{11} & . & . & \alpha_{1,n-2} & \alpha_{1,n} & 0 \\ . & . & . & . & . & . \\ \alpha_{n-2,1} & . & . & \alpha_{n-2,n-2} & \alpha_{n-2,n} & 0 \\ \alpha_{n,1} & . & . & \alpha_{n,n-2} & \alpha_{n,n} & 1 \\ 0 & . & . & 0 & 1 & 0 \end{vmatrix}$$

$$= -\begin{vmatrix} \alpha_{11} & . & . & \alpha_{1,n-2} \\ . & . & . & . \\ \alpha_{n-2,1} & . & . & \alpha_{n-2,n-2} \end{vmatrix}.$$

Similarly for the other determinants, the results being

$$p_n^2 \begin{vmatrix} a_{11} & a_{1n} & p_1 \\ a_{n1} & a_{nn} & p_n \\ p_1 & p_n & 0 \end{vmatrix} = -\alpha_{11},$$

$$p_n^2 \begin{vmatrix} a_{11} & a_{12} & a_{1n} & p_1 \\ a_{21} & a_{22} & a_{2n} & p_2 \\ a_{31} & a_{32} & a_{3n} & p_n \\ p_1 & p_2 & p_n & 0 \end{vmatrix} = -\begin{vmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{vmatrix},$$

and so on until we reach (8). The condition that (1) be positive whenever (2) is satisfied and at least one x_r is not zero is that the determinants on the left be all negative.

2.2. *Two linear conditions*

Let there be two distinct linear conditions

$$p_1 x_1 + \ldots + p_n x_n = 0, \tag{2a}$$

$$q_1 x_1 + \ldots + q_n x_n = 0, \tag{2b}$$

and let $p_n \neq 0$. Then, as we shall prove, one at least of

$$p_n q_r - p_r q_n \quad (r = 1,\ldots, n-1)$$

is not zero. For if they are all zero, each q_r is given by $p_r(q_n/p_n)$ and every $p_r q_s - p_s q_r$ is zero; this is contrary to the hypothesis that (2 *a*) and (2 *b*) are distinct conditions. We shall assume that $p_n q_{n-1} - p_{n-1} q_n \neq 0$. That is, we shall assume

$$p_n \neq 0, \qquad p_n q_{n-1} - p_{n-1} q_n \neq 0. \tag{9}$$

If (9) holds and also conditions (2 *a*) and (2 *b*) hold, at least one of $x_1,..., x_{n-2}$ is not equal to zero.

We may now proceed on the lines laid down in § 2.1. Make the transformation

$$X_r = x_r \quad (r = 1,..., n-2),$$
$$X_{n-1} = q_1 x_1 + ... + q_n x_n,$$
$$X_n = p_1 x_1 + ... + p_n x_n.$$

The modulus of the transformation is $p_n q_{n-1} - p_{n-1} q_n$ and is not zero. Let the form

$$\sum_{r,s=1}^{n} a_{rs} x_r x_s$$

become

$$\sum_{r,s=1}^{n} \alpha_{rs} X_r X_s.$$

In this, $X_n = X_{n-1} = 0$ and at least one of $X_1,..., X_{n-2}$ is not zero. A set of necessary and sufficient conditions for the form to be positive is

$$\alpha_{11} > 0, \quad \begin{vmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{vmatrix} > 0, \quad ..., \quad \begin{vmatrix} \alpha_{11} & \cdot & \cdot & \alpha_{1,n-2} \\ \cdot & \cdot & \cdot & \cdot \\ \alpha_{n-2,1} & \cdot & \cdot & \alpha_{n-2,n-2} \end{vmatrix} > 0.$$

To express these in terms of a, p, q we proceed as in § 2.1; introduce new variables x_{n+1}, x_{n+2} and the form

$$\sum_{r,s=1}^{n} a_{rs} x_r x_s + 2x_{n+1}(q_1 x_1 + ... + q_n x_n) +$$
$$+ 2x_{n+2}(p_1 x_1 + ... + p_n x_n).$$

Transform it as we transformed $\sum a_{rs} x_r x_s$, making $X_{n+1} = x_{n+1}$ and $X_{n+2} = x_{n+2}$. It becomes

$$\sum_{r,s=1}^{n} \alpha_{rs} X_r X_s + 2X_{n-1} X_{n+1} + 2X_n X_{n+2}.$$

The modulus of the transformation is $p_n q_{n-1} - p_{n-1} q_n = M$, say, where $M \neq 0$. Hence

$$M^2 \begin{vmatrix} a_{11} & . & . & a_{1n} & q_1 & p_1 \\ . & . & . & . & . & . \\ a_{n1} & . & . & a_{n,n} & q_n & p_n \\ q_1 & . & . & q_n & 0 & 0 \\ p_1 & . & . & p_n & 0 & 0 \end{vmatrix} = \begin{vmatrix} \alpha_{11} & . & . & \alpha_{1n} & 0 & 0 \\ . & . & . & . & . & . \\ \alpha_{n-1,1} & . & . & \alpha_{n-1,n} & 1 & 0 \\ \alpha_{n,1} & . & . & \alpha_{n,n} & 0 & 1 \\ 0 & . & . & 1 & 0 & 0 & 0 \\ 0 & . & . & 0 & 1 & 0 & 0 \end{vmatrix}$$

$$= + \begin{vmatrix} \alpha_{11} & . & . & \alpha_{1,n-2} \\ . & . & . & . \\ \alpha_{n-2,1} & . & . & \alpha_{n-2,n-2} \end{vmatrix}. \tag{10}$$

We deal with the other α_{rs} minors by making, in succession, $x_{n-2} \equiv 0$, $x_{n-3} \equiv 0, \ldots, x_2 \equiv 0$. The results are obtained by deleting in succession all suffixes $n-2$, all suffixes $n-3, \ldots$, all suffixes 2; the last result in the sequence is

$$M^2 \begin{vmatrix} a_{11} & a_{1,n-1} & a_{1,n} & q_1 & p_1 \\ a_{n-1,1} & a_{n-1,n-1} & a_{n-1,n} & q_{n-1} & p_{n-1} \\ a_{n,1} & a_{n,n-1} & a_{n,n} & q_n & p_n \\ q_1 & q_{n-1} & q_n & 0 & 0 \\ p_1 & p_{n-1} & p_n & 0 & 0 \end{vmatrix} = \alpha_{11}.$$

Provided that we have so arranged the variables $x_1, \ldots, x_n$ that $p_n \neq 0$ and $p_n q_{n-1} - p_{n-1} q_n \neq 0$, the form $\sum a_{rs} x_r x_s$ is positive-definite under the linear conditions $\sum p_r x_r = 0$, $\sum q_r x_r = 0$ when the bordered determinants indicated above are all positive.

3. Sets of anti-commutative matrices

In this concluding section we give an elementary introduction† to matrices $E_1, E_2, \ldots$ which satisfy the equations

$$E_r^2 = -I, \qquad E_r E_s = -E_s E_r \quad (r \neq s). \tag{1}$$

We do this partly to call the attention of the reader to a particular set of properties and partly to indicate, if only by one example, how much of interest still remains after the main lines of a central theory have been covered. Before we consider

† It is an introduction to the work of Eddington, Newman, and others, not an account of that work. See list of references at the end of this section.

particular matrices we note the following general property, assuming that some set of matrices has been found to satisfy (1). We have, at the outset,

(i) a set of matrices E_r satisfying (1).

We leave aside for the moment how many matrices there may be in a set. First we note that the product of any two distinct matrices of (i) is a matrix whose square is $-I$; for example,

$$(E_r E_s)(E_r E_s) = -E_s E_r E_r E_s = E_s^2 = -I.$$

We thus have a second set

(ii) $$E_r E_s; \qquad (E_r E_s)^2 = -I.$$

It may or may not happen that some of these products are matrices that have already appeared in (i). Looking at the commutative property we find, by simple calculation, that, for a given r

$$E_r, \quad E_r E_s, \quad E_r E_t \quad \ldots \text{ are anti-commutative,}$$

but

$$E_r \quad \text{and} \quad E_s E_t \text{ are commutative.}$$

The product of any three members of (i) is either itself a (i) multiplied by ± 1 or its square is I and not $-I$.

3.1. *Matrices of order two*

We use Greek letters to denote matrices of order two and Arabic letters to denote their elements, real or complex numbers; i is, of course, $\sqrt{(-1)}$.

The canonical form of a matrix whose square is $-I$ must be

$$\operatorname{diag}(\pm i, \pm i),$$

but if two like signs are taken we have a matrix that is commutative with any other matrix of order two and is therefore not relevant to our problem. So let us begin with

$$\alpha_1 = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}.$$

Let

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

satisfy the condition $\alpha\alpha_1 = -\alpha_1\alpha$. Then

$$\begin{bmatrix} ai & -bi \\ ci & -di \end{bmatrix} = -\begin{bmatrix} ai & bi \\ -ci & -di \end{bmatrix},$$

and hence $a = d = 0$. Further, with

$$\alpha = \begin{bmatrix} 0 & b \\ c & 0 \end{bmatrix},$$

$\alpha^2 = \text{diag}(bc, bc)$, so that $\alpha^2 = -I$ requires $bc = -1$. Thus
matrix α_r satisfying

$$\alpha_r^2 = -I, \qquad \alpha_1\alpha_r = -\alpha_r\alpha_1,$$

is given by $$\alpha_r = \begin{bmatrix} 0 & b_r \\ -b_r^{-1} & 0 \end{bmatrix}.$$

If now we take two such matrices and require that

$$\alpha_r\alpha_s = -\alpha_s\alpha_r,$$

we find that $$b_r b_s^{-1} = -b_s b_r^{-1},$$

i.e. $$b_r/b_s = \pm i. \qquad ($$

If we choose a definite real value for b_r, any b_s is pure imagi
ary; and conversely. Taking $b_2 = -1$, $b_3 = i$, we have *thr*
matrices α_1, α_2, α_3 satisfying (1) given by

$$\alpha_1 = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \quad \alpha_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \alpha_3 = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}.$$

The condition (2) ensures that we can have three and no mo
in such a set. Moreover,

$$\alpha_2\alpha_3 = -\alpha_1, \quad \alpha_3\alpha_1 = -\alpha_2, \quad \alpha_1\alpha_2 = -\alpha_3.$$

3.2. *Matrices of order four*

If we look for solutions of (1) in which E_r is a matrix of order fo
the most systematic way is probably to follow Newman's meth
to show that the canonical form of any E_r must be diag($i, i, -i,$ -
and to take E_1 as diag($iI_2, -iI_2$). In writing this introductory accou
I thought it might be interesting to build up the set of five matri
given by Jeffreys and Jeffreys in *Methods of Mathematical Physic*

We use capital letters to denote matrices of order four a
Greek letters to denote matrices of order two. Some details

† p. 138.

subsidiary calculations are omitted: the reader will need to fill these in for himself.

The matrices $\text{diag}(\pm\alpha_r, \pm\alpha_r)$ offer themselves at once as possible members of an anti-commutative set of square roots of $-I_4$. We start with†

$$E_1 = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_1 \end{bmatrix}, \qquad E_3 = \begin{bmatrix} \alpha_3 & 0 \\ 0 & \alpha_3 \end{bmatrix},$$

leaving E_2 for the moment.

A matrix
$$E = \begin{bmatrix} 0 & \beta \\ \gamma & 0 \end{bmatrix}$$

will be anti-commutative with both E_1 and E_3 if β and γ are anti-commutative with both α_1 and α_3. To find such matrices β and γ let

$$\beta = \begin{bmatrix} 0 & b \\ d & 0 \end{bmatrix}.$$

Then (by § 3.1) β is anti-commutative with α_1; it is anti-commutative with α_3 if $b = -d$.

We see now that we cannot take $E_2 = \text{diag}(\alpha_2, \alpha_2)$ as one of our set to satisfy (1): for this will be anti-commutative with E above only if β and γ are anti-commutative with α_2; and $\alpha_2\beta = -\beta\alpha_2$ requires $b = d$. We therefore take

$$E_2 = \text{diag}(\alpha_2, -\alpha_2).$$

With this choice for E_2, any matrix

$$E = \begin{bmatrix} 0 & \beta \\ \gamma & 0 \end{bmatrix},$$

where
$$\beta = \begin{bmatrix} 0 & b \\ -b & 0 \end{bmatrix}, \qquad \gamma = \begin{bmatrix} 0 & c \\ -c & 0 \end{bmatrix},$$

is anti-commutative with E_1, E_2, and E_3.

The square of E is $\text{diag}(\beta\gamma, \beta\gamma)$ and $\beta\gamma = \text{diag}(-bc, -bc)$. Hence
$$E^2 = \text{diag}(-bc, -bc, -bc, -bc)$$

† The 0's denote null sub-matrices of order two and α_1, α_2, α_3 are the matrices of the previous sub-section.

satisfy the condition $\alpha\alpha_1 = -\alpha_1\alpha$. Then

$$\begin{bmatrix} ai & -bi \\ ci & -di \end{bmatrix} = -\begin{bmatrix} ai & bi \\ -ci & -di \end{bmatrix},$$

and hence $a = d = 0$. Further, with

$$\alpha = \begin{bmatrix} 0 & b \\ c & 0 \end{bmatrix},$$

$\alpha^2 = \mathrm{diag}(bc, bc)$, so that $\alpha^2 = -I$ requires $bc = -1$. Thus a matrix α_r satisfying

$$\alpha_r^2 = -I, \qquad \alpha_1\alpha_r = -\alpha_r\alpha_1,$$

is given by

$$\alpha_r = \begin{bmatrix} 0 & b_r \\ -b_r^{-1} & 0 \end{bmatrix}.$$

If now we take two such matrices and require that

$$\alpha_r\alpha_s = -\alpha_s\alpha_r,$$

we find that

$$b_r b_s^{-1} = -b_s b_r^{-1},$$

i.e.

$$b_r/b_s = \pm i. \tag{2}$$

If we choose a definite real value for b_r, any b_s is pure imaginary; and conversely. Taking $b_2 = -1$, $b_3 = i$, we have *three* matrices α_1, α_2, α_3 satisfying (1) given by

$$\alpha_1 = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \qquad \alpha_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \qquad \alpha_3 = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}.$$

The condition (2) ensures that we can have three and no more in such a set. Moreover,

$$\alpha_2\alpha_3 = -\alpha_1, \quad \alpha_3\alpha_1 = -\alpha_2, \quad \alpha_1\alpha_2 = -\alpha_3.$$

3.2. *Matrices of order four*

If we look for solutions of (1) in which E_r is a matrix of order four, the most systematic way is probably to follow Newman's method, to show that the canonical form of any E_r must be $\mathrm{diag}(i, i, -i, -i)$ and to take E_1 as $\mathrm{diag}(iI_2, -iI_2)$. In writing this introductory account I thought it might be interesting to build up the set of five matrices given by Jeffreys and Jeffreys in *Methods of Mathematical Physics*.†

We use capital letters to denote matrices of order four and Greek letters to denote matrices of order two. Some details of

† p. 138.

subsidiary calculations are omitted: the reader will need to fill these in for himself.

The matrices $\mathrm{diag}(\pm\alpha_r, \pm\alpha_r)$ offer themselves at once as possible members of an anti-commutative set of square roots of $-I_4$. We start with†

$$E_1 = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_1 \end{bmatrix}, \qquad E_3 = \begin{bmatrix} \alpha_3 & 0 \\ 0 & \alpha_3 \end{bmatrix},$$

leaving E_2 for the moment.

A matrix
$$E = \begin{bmatrix} 0 & \beta \\ \gamma & 0 \end{bmatrix}$$

will be anti-commutative with both E_1 and E_3 if β and γ are anti-commutative with both α_1 and α_3. To find such matrices β and γ let

$$\beta = \begin{bmatrix} 0 & b \\ d & 0 \end{bmatrix}.$$

Then (by § 3.1) β is anti-commutative with α_1; it is anti-commutative with α_3 if $b = -d$.

We see now that we cannot take $E_2 = \mathrm{diag}(\alpha_2, \alpha_2)$ as one of our set to satisfy (1): for this will be anti-commutative with E above only if β and γ are anti-commutative with α_2; and $\alpha_2\beta = -\beta\alpha_2$ requires $b = d$. We therefore take

$$E_2 = \mathrm{diag}(\alpha_2, -\alpha_2).$$

With this choice for E_2, any matrix

$$E = \begin{bmatrix} 0 & \beta \\ \gamma & 0 \end{bmatrix},$$

where
$$\beta = \begin{bmatrix} 0 & b \\ -b & 0 \end{bmatrix}, \qquad \gamma = \begin{bmatrix} 0 & c \\ -c & 0 \end{bmatrix},$$

is anti-commutative with E_1, E_2, and E_3.

The square of E is $\mathrm{diag}(\beta\gamma, \beta\gamma)$ and $\beta\gamma = \mathrm{diag}(-bc, -bc)$. Hence

$$E^2 = \mathrm{diag}(-bc, -bc, -bc, -bc)$$

† The 0's denote null sub-matrices of order two and α_1, α_2, α_3 are the matrices of the previous sub-section.

and is equal to $-I_4$ if $bc = 1$. Thus a matrix

$$E_r = \begin{bmatrix} 0 & 0 & 0 & b_r \\ 0 & 0 & -b_r & 0 \\ 0 & b_r^{-1} & 0 & 0 \\ -b_r^{-1} & 0 & 0 & 0 \end{bmatrix} \quad (r = 4, 5,...)$$

satisfies $E_r^2 = -I$ and is anti-commutative with E_1, E_2, E_3.

The condition for $E_r E_s = -E_s E_r$ is

$$(b_r/b_s)^2 = -1$$

and two matrices E_r and E_s can be included in a set satisfying (1) if the ratio b_r/b_s is $\pm i$. We take $b_4 = i$ and $b_5 = -1$ and, in the order in which they are set out in Jeffreys and Jeffreys,† write the set of five solutions of (1) as

$$E_3 = \text{diag}(\alpha_3, \alpha_3), \qquad E_1 = \text{diag}(\alpha_1, \alpha_1),$$

$$E_4 = \begin{bmatrix} 0 & 0 & 0 & i \\ 0 & 0 & -i & 0 \\ 0 & -i & 0 & 0 \\ i & 0 & 0 & 0 \end{bmatrix},$$

$$E_2 = \text{diag}(\alpha_2, -\alpha_2),$$

$$E_5 = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

3.3. *References*

A. S. Eddington, 'On sets of anticommuting matrices', *Journal London Math. Soc.* **7** (1932), 58–68.

M. H. A. Newman, ibid. 93–9.

H. Jeffreys and B. S. Jeffreys, *Methods of Mathematical Physics* (Cambridge, 1946), 136–8.

A. Hurwitz, *Math. Ann.* **88** (1923), 1–25.

The last of these deals with roots of $+I$.

† Loc. cit., p. 138.

INDEX

Numbers refer to pages. Note the entries under 'Definitions' and 'Matrix'.

Algebra of power series 113.
Anti-commutative matrices 176.

Bilinear forms 23, 50.

Characteristic:
  equation 12, 52.
  repeated roots and rank 148.
  root 25.
Classical canonical form:
  by chains 70–2.
  short proof 80.
  submatrices 59.
  theorem 68.
Conjunction 145.
Convergence 87.

Definitions:
  c-equivalence 51.
  chain 71.
  collineation 50.
  equivalent matrices 20.
  field of numbers 15.
  function of matrix 98, 132, 133.
  G changes 60.
  H changes 56.
  λ-equivalence 28, 41.
  Segre characteristic 74.
  transform 51.
  X changes 60, 61.
Diagonal form:
  under congruence 137.
  under conjunction 146.
Differentiation:
  of function of a function 126.
  of matrix 110.
  of matrix functions 123–5, 134.
  of matrix power series 111.

Eigenvalues 25.
Elementary divisors 41, 47, 75.
  linear 79.

Fractional powers of matrix 103.
Function of matrix:
  canonical form of 106.
  defined 98.

H.C.F. of minors 35, 37–9, 75–86.
Hermitian forms 147, 152.

Invariant factors 39, 46, 83, 86, 109.

Jacobi's canonical form 53.

Latent root 12, 25.
Linear:
  dependence 9.
  elementary divisors 79.
  matrix equations 166.
  scalar equations 10.

Matrix:
  adjoint 4.
  adjugate 4.
  associated with polynomial 81.
  auxiliary unit 88.
  characteristic equation 12.
  classical canonical form 59.
  diagonal 12.
  equivalent 20, 23.
  function of 5, 98, 132.
  Hermitian 14, 145.
  idem-potent 156.
  linear in λ 43.
  minimum function of 83, 153.
  nil-potent 156.
  normal form 32, 35.
  orthogonal 14, 138, 140.
  power of 5.
  rank of 8.
  rational canonical form 80.
  reciprocal 4, 120.
  rectangular 24.
  resolvent of 168.
  Segre characteristic of 73.
  skew-hermitian 14.
  skew-symmetric 14.
  square root of 160.
  submatrix 7.
  symmetric 14.
  transpose of 5.
  triangular 140.
  U (auxiliary unit) 88.
  unitary 14, 141.
Minors:
  of diagonal matrix 74.

Orthogonal congruence 142.

Phillips, H. B. 129, 165.
Power series, matrix 87.
  algebra of 114, 119, 122.
  as polynomial 93, 133.
  with zero-sum 119.
Power series, scalar 87.

Quadratic form 144.
  positive-definite 172.

Rational canonical form 80.
Roth, W. E. 166.
Rutherford, D. E. 159.

Scalar equations 164.
Schmidt's orthogonalization process 140, 150.
Segre characteristic 73, 75.

Taylor's expansion 129.
Transformations:
  elementary 15.
  elementary λ- 26.

PRINTED IN
GREAT BRITAIN
AT THE
UNIVERSITY PRESS
OXFORD
BY
CHARLES BATEY
PRINTER
TO THE
UNIVERSITY