Diagonal matrices. By Mr H. W. TURNBULL, Trinity College, Professor of Mathematics, University of St Andrews.

[*Received* 31 March, *read* 15 May 1933.]

Introduction.

In the following pages I have developed the theory of matrices by resolving them into parallel components arranged diagonally, rather than into the usual rows and columns. This treatment is natural in view of the fundamental fact that the resolution is undestroyed when matrices are formed into products (Theorem 2). It is closely related to the theory of continuants and of continued fractions. Certain features stand out in such a presentation—the distinction between the *length* and *range* of a diagonal (§ 4), that between *regular* and *irregular* diagonals (§ 6), and the use of *equable partition* (§ 7). The exact conditions for the existence of an rth root of a given singular matrix are examined in § 9 and summarized under the title, *the condition of equability.*

The index law of Theorem 1 (§ 1) is capable of extension by the introduction of *secondary diagonal* matrices E_r, a secondary being defined as a diagonal perpendicular to a primary D_r. The law would then assume the form

$$D_r D_s = D_{r+s}, \quad D_r E_s = E_{r+s}, \quad E_r D_s = E_{r-s}, \quad E_r E_s = D_{r-s},$$

where $-\nu \leqslant r \pm s \leqslant \nu$. The further development of this secondary theory would be useful with a view to its application to symmetric and Hermitian matrices.

1. *Matrices resolved into diagonal matrices.*

In the usual notation $[x_{ij}]$ for a square array of elements x_{ij} the suffix i refers to the *row* and the suffix j to the *column* of the typical element x_{ij}. Useful as this is for purposes of addition it rather obscures the issue when matrices are multiplied; and an alternative notation, determined not by row and column but by gnomon and diagonal, has certain advantages.

An example will make this notation clear, and will also serve to indicate what is meant by *gnomon* and *diagonal.* Let

$$\begin{bmatrix} x_{00} & x_{01} & x_{02} \\ x_{10} & x_{11} & x_{12} \\ x_{20} & x_{21} & x_{22} \end{bmatrix} = \begin{bmatrix} x_{(0,0)} & x_{(0,1)} & x_{(0,2)} \\ x_{(0,-1)} & x_{(1,0)} & x_{(1,1)} \\ x_{(0,-2)} & x_{(1,-1)} & x_{(2,0)} \end{bmatrix},$$

and in brief let $[x_{ij}] = [x_{(\gamma\delta)}]$.

I shall suppose that, unless the contrary is stated, the suffixes i, j, γ each take the $\nu+1$ values

$$0, 1, 2, \ldots, \nu,$$

and that δ takes the $2\nu+1$ values

$$\delta = 0, \pm 1, \pm 2, \ldots, \pm \nu.$$

When $\delta = 0$ the element occupies a position on the *leading* or *principal diagonal*, all other diagonals being numbered in an obvious way from this zero position. Diagonals which are indicated by a negative suffix lie below and to the left of the leading diagonal: those with a positive suffix lie above and to the right.

In *infinite matrices* (when $\nu \to \infty$) each diagonal has an infinite *length*: in a *finite matrix* the leading diagonal is said to have a length $\nu+1$; it possesses $\nu+1$ elements. The extreme diagonals ($\delta = \pm \nu$) have length unity (they each possess a single element). The length of the diagonal in general is given by

$$l = \nu + 1 - |\delta|.$$

When $\gamma = 0$ the element occupies a place in the *leading gnomon* or Γ-shaped border of the array, which encloses the next gnomon ($\gamma = 1$), which in turn encloses the next, and so on, until a final single element ($\gamma = \nu$) gives the final (and νth) gnomon.

Evidently one pair of values i, j defines one pair γ, δ by the relations

$$x_{(\gamma\delta)} = x_{ij}, \quad \gamma = \min(i, j), \quad \delta = j - i,$$

where γ is the smaller of i and j when they differ. Conversely the values γ, δ define values i, j uniquely.

Next let D_δ denote the square array obtained by replacing all the elements of X by zeros except for those which belong to the δth diagonal, which are unaltered. If this process is applied for each value of δ, a set of $2\nu+1$ matrices is obtained whose sum (according to the addition law of matrices) is X. Thus

$$X = \sum_{\delta=-\nu}^{\nu} D_\delta.$$

For example, in the usual ij notation, if $\nu = 2$, then

$$X = D_{-2} + D_{-1} + D_0 + D_1 + D_2$$

$$= \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ x_{20} & \cdot & \cdot \end{bmatrix} + \begin{bmatrix} \cdot & \cdot & \cdot \\ x_{10} & \cdot & \cdot \\ \cdot & x_{21} & \cdot \end{bmatrix} + \begin{bmatrix} x_{00} & \cdot & \cdot \\ \cdot & x_{11} & \cdot \\ \cdot & \cdot & x_{22} \end{bmatrix}$$

$$+ \begin{bmatrix} \cdot & x_{01} & \cdot \\ \cdot & \cdot & x_{12} \\ \cdot & \cdot & \cdot \end{bmatrix} + \begin{bmatrix} \cdot & \cdot & x_{02} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix},$$

and the matrix is said to be resolved into its *diagonal components*.

https://doi.org/10.1017/S0305004100016406 Published online by Cambridge University Press

The following notation is very useful for denoting one or other such diagonal component:

$$D_r = \mathrm{diag}_r (x_{i,i+r}) = \mathrm{diag}_r (x_\gamma),$$

where the double suffixes indicate row and column, whereas the single suffix γ indicates the gnomon. In particular, if the suffix r is suppressed, the *principal diagonal* is indicated:

$$D_0 = \mathrm{diag}_0 (x_{ii}) = \mathrm{diag}\,(x_{ii})$$
$$= \mathrm{diag}\,(x_{00}, x_{11}, \ldots, x_{\nu\nu}),$$

of which the *unit matrix* is a special example,

$$I = \mathrm{diag}_0 (1) = \mathrm{diag}_0 (1, 1, \ldots, 1), \qquad \text{to } \nu + 1 \text{ terms.}$$

The first *over-diagonal* is

$$\mathrm{diag}_1 (x_{i,i+1}) = \mathrm{diag}_1 (x_{01}, x_{12}, \ldots, x_{\nu-1,\nu}).$$

In particular the matrix

$$U = \mathrm{diag}_1 (1), \qquad (x_{i,i+1} = 1),$$

with its ν non-zero elements, is called the *auxiliary unit matrix*. Again the first *under-diagonal* is

$$\mathrm{diag}_{-1} (x_{i,i-1}) = \mathrm{diag}_{-1} (x_{10}, x_{21}, \ldots, x_{\nu,\nu-1});$$

and in particular the matrix

$$U' = \mathrm{diag}_{-1} (1), \qquad (x_{i,i-1} = 1)$$

is called the *transposed auxiliary unit matrix*.

Transposition is usually indicated by an accent, so that, if $X = [x_{ij}]$, then $X' = [x_{ji}]$.

The Kronecker delta,

$$\delta_{ii} = 1, \quad \delta_{ij} = 0 \quad (i \neq j),$$

provides a still more compact notation for the diagonal components of a matrix. Evidently it is legitimate to write $\delta_{i,j-r}$ if it is understood that i and j still refer to the row and column respectively in which the element stands. These various diagonal matrices can now be written

$$I = [\delta_{ij}], \quad U = [\delta_{i,j-1}], \quad U' = [\delta_{i-1,j}],$$
$$D_r = [x_{ij}\delta_{i,j-r}], \quad D_{-r} = [x_{ij}\delta_{i-r,j}],$$

where $0 \leqslant i - r \leqslant \nu$, $0 \leqslant j - r \leqslant \nu$, $r \geqslant 0$.

It is evident that, if two matrices $X = \Sigma D$ and $Y = \Sigma E$ are resolved into their diagonal components D_r and E_r, then the sum $X + Y$ is given by the sum of diagonal components of which the rth is $D_r + E_r$. There is a corresponding property arising from the product of two matrices, as the following two theorems shew.

https://doi.org/10.1017/S0305004100016406 Published online by Cambridge University Press

THEOREM 1. *The product of two arbitrary diagonal matrices D_r and D_s is a diagonal matrix D_{r+s} or else is zero, according as $r+s$ is numerically less than or not less than ν.*

Proof. From the product law of matrices, $[x_{ij}][y_{ij}]=[z_{ij}]$, where $z_{ij}=\sum_{k=0}^{\nu} x_{ik}y_{kj}$, it follows that, if

$$[x_{ij}]=\mathrm{diag}_r(x_{i,r+i})$$

and

$$[y_{ij}]=\mathrm{diag}_s(y_{i,s+i}),$$

then z_{ij} has only one possible non-zero term

$$z_{ij}=x_{ik}y_{kj},$$

where $k=i+r$, $j=k+s=i+r+s$. Hence

$$j-i=r+s;$$

so that z_{ij} is zero unless it is an element upon the $(r+s)$th diagonal, in which case

$$x_{i,i+r}\,y_{i+r,i+r+s}=z_{i,i+r+s}.$$

We may therefore write the result

$$\mathrm{diag}_r(x_{i,i+r})\,\mathrm{diag}_s(y_{i,i+s})=\mathrm{diag}_{r+s}(z_{i,i+r+s}),$$

subject to the obvious conditions

$$0\leqslant i\leqslant \nu,\quad 0\leqslant i+r\leqslant \nu,\quad 0\leqslant i+r+s\leqslant \nu.$$

This first theorem asserts the *index law* obeyed by products of diagonal matrices, and it holds for negative as well as for positive and zero values of r and s.

THEOREM 2. *The product of any two matrices X and Y may be expressed in terms of diagonal components each of which is linear in those of both X and Y.*

Proof. For if $X=\sum_{r=-\nu}^{\nu} D_r$, $Y=\sum_{r=-\nu}^{\nu} E_r$ are two such matrices, then their product can be written

$$XY=Z=\sum_{r=-\nu}^{\nu} F_r,$$

where, by Theorem 1, the expression

$$F_r=D_{-\nu+r}E_\nu+\ldots+D_iE_{r-i}+\ldots+D_\nu E_{r-\nu}$$

is the sum of terms all of which belong to the type D_r. Hence F_r is a diagonal component of the product XY, for each value of r.

https://doi.org/10.1017/S0305004100016406 Published online by Cambridge University Press

2. *Continuants.*

Two particular types of matrix which commonly occur are the *continuant*,

$$C = D_{-1} + D_0 + D_1, \tag{1}$$

and the *classical canonical form*,

$$HXH^{-1} = D_0 + D_1.$$

In the first of these examples it will be seen that there are three consecutive diagonal components placed on and adjacent to the principal diagonal. Thus

$$C = \begin{bmatrix} b_0 & c_1 & \cdot & \cdot & \cdots \\ d_1 & b_1 & c_2 & \cdot & \cdots \\ \cdot & d_2 & b_2 & c_3 & \cdots \\ \cdot & \cdot & d_3 & b_3 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}. \tag{2}$$

This is the matrix which underlies the theory of continued fractions, as indeed is implied by a remark of Sylvester(1), relative to the leading element in the reciprocal matrix of C. If this element is denoted by f, then it is a fact that

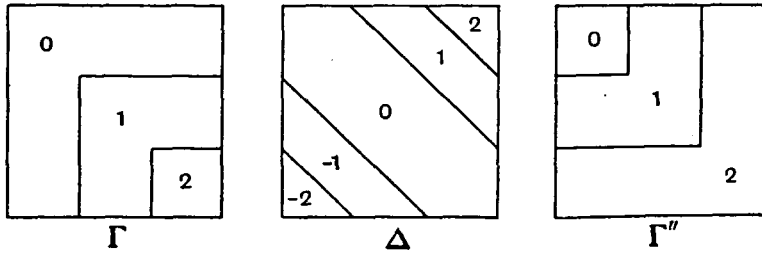
$$\frac{1}{C} = \begin{bmatrix} f, & \cdots \\ \cdots & \cdots \end{bmatrix}, \tag{3}$$

where

$$f = \frac{1}{b_0} - \frac{c_1 d_1}{b_1} - \frac{c_2 d_2}{b_2} - \cdots - \frac{c_\nu d_\nu}{b_\nu}.$$

It is interesting to notice that the suffixes of the letters appear in a gnomon pattern in the matrix C, whereas they follow the natural order in the continued fraction. The letters themselves indicate the diagonals of C.

This gnomon pattern is, however, the reverse of that which was earlier introduced, and provides an equally useful way of analysing matrices. If the present is called the Γ'' pattern, then the fundamental analysis by diagonal or by gnomon may be visualized as follows:



https://doi.org/10.1017/S0305004100016406 Published online by Cambridge University Press

Matrices of this continuant type have also arisen in quantum algebra, and for this, apart from the earlier, reason they invite further examination. Evidently the sum or difference of two continuant matrices is itself a continuant matrix; but this property does not necessarily hold for the product or quotient. It is interesting, therefore, to find out the requisite conditions, an enquiry which in fact has been suggested to me by Professor E. T. Whittaker, namely:

To find necessary and sufficient conditions for the product of two continuants to be a continuant.

These follow readily enough by forming the product PQ of two continuants

$$P = D_{-1} + D_0 + D_1, \quad Q = E_{-1} + E_0 + E_1.$$

Since $PQ = \Sigma D_r E_s = \Sigma F_{r+s}$, where F_{r+s} is a diagonal component, it follows that necessary and sufficient conditions are

$$D_1 E_1 = 0 = D_{-1} E_{-1},$$

these being the only terms in the product PQ which contribute anything beyond the requisite three diagonals. Now each of these four matrices contains exactly ν possible non-zero elements. Let them be written as

$$D_1 = \mathrm{diag}_1(d_0, d_1, d_2, \ldots, d_{\nu-1}), \quad D_{-1} = \mathrm{diag}_{-1}(d_i'),$$
$$E_1 = \mathrm{diag}_1(e_0, e_1, e_2, \ldots, e_{\nu-1}), \quad E_{-1} = \mathrm{diag}_{-1}(e_i').$$

Then the condition $D_1 E_1 = 0$ is equivalent to the $\nu - 1$ equations

$$d_0 e_1 = 0, \quad d_1 e_2 = 0, \quad \ldots, \quad d_{\nu-2} e_{\nu-1} = 0,$$

while a similar set of $\nu - 1$ equations holds for $D_{-1} E_{-1}$. It follows that at least $\nu - 1$ elements, one from each of the pairs (d_i, e_{i+1}), must be zero; and that at least $\nu - 1$ elements, one from each of the pairs (d_i', e'_{i-1}), must also be zero. In other respects the elements may be arbitrary: in particular the first of E_1 and the last element of D_1 are arbitrary.

Evidently there are exactly $2^{\nu-1}$ different alternative ways of stating the minimum requirements for satisfying each of these sets of conditions; among which the most elegant is that wherein all elements with odd suffixes are zero. In this case both P and Q are of the type

$$R \equiv \begin{bmatrix} \alpha_0 & \beta_0 & & & & & \\ \beta_0' & \alpha_1 & \cdot & & & & \\ & \cdot & \alpha_2 & \beta_2 & & & \\ & & \beta_2' & \alpha_3 & \cdot & & \\ & & & \cdot & \alpha_4 & \beta_4 & \\ & & & & \beta_4' & \alpha_5 & \\ & & & & & & \ddots \end{bmatrix},$$

https://doi.org/10.1017/S0305004100016406 Published online by Cambridge University Press

that is, a matrix in which a set of arbitrary binary matrices (each consisting of four elements with a possible residual single element $x_{\nu\nu}$ if ν is even) is disposed along the diagonal. More succinctly one may write

$$R = \text{diag}\left(\begin{bmatrix} \alpha_0 & \beta_0 \\ \beta_0' & \alpha_1 \end{bmatrix}, \begin{bmatrix} \alpha_2 & \beta_2 \\ \beta_2' & \alpha_3 \end{bmatrix}, \ldots\right).$$

Evidently, too, when each of P and Q is of type R, so is the alternative product QP. Thus by satisfying the $2(\nu-1)$ conditions

$$d_{2i+1} = d'_{2i+1} = e_{2i+1} = e'_{2i+1} = 0$$

both products PQ and QP are continuants.

Conversely, if both PQ and QP are continuants, then all the $4(\nu-1)$ equations

$$d_i e_{i+1} = 0, \quad e_i d_{i+1} = 0, \quad d_i' e'_{i-1} = 0, \quad e_i' d'_{i-1} = 0$$

must be satisfied. These conditions can be stated in graphical form by means of two rectangular matrices

$$S = \begin{bmatrix} d_0, & d_1, & \ldots, & d_{\nu-1} \\ e_0, & e_1, & \ldots, & e_{\nu-1} \end{bmatrix}, \quad T = \begin{bmatrix} d_0', & d_1', & \ldots, & d'_{\nu-1} \\ e_0', & e_1', & \ldots, & e'_{\nu-1} \end{bmatrix},$$

which must be *such that each diagonal of two letters must include a zero element*, the diagonals being either *primary* (parallel to that of $d_0 e_1$) or *secondary* (parallel to that of $e_0 d_1$). The minimum requirements are now attained by inserting the fewest possible zeros according to these cited rules, and this leads to an interesting theorem.

THEOREM 3. *When each of the continuants P and Q has an even number of rows and columns, the minimum requirements for ensuring that both PQ and QP are also continuants are satisfied only if both P and Q are of type R.*

Proof. In such matrices ν is odd, and a possible scheme for both S and T is the following,

$$S_1 = \begin{bmatrix} \times & \bullet & \times & \bullet & \times & \ldots & \bullet & \times \\ \times & \bullet & \times & \bullet & \times & \ldots & \bullet & \times \end{bmatrix},$$

which evidently belongs to type R, where the dots occur in alternate columns and denote zeros, in total $\nu-1$, there being $\nu+1$ arbitrary elements ($\times$) in the residual columns.

Furthermore, any other possible scheme must be of one of the following types,

$$S_2 = \begin{bmatrix} \times & \times & \ldots & \times \\ \bullet & \bullet & \ldots & \bullet \end{bmatrix}, \quad S_3 = \begin{bmatrix} \bullet & \bullet & \ldots & \bullet \\ \times & \times & \ldots & \times \end{bmatrix},$$

$$S_4 = \begin{bmatrix} \times & \bullet & \times & \bullet & \ldots \\ \times & \bullet & \times & \bullet & \ldots \end{bmatrix}, \quad S_5 = \begin{bmatrix} \bullet & \times & \bullet & \times & \ldots \\ \bullet & \times & \bullet & \times & \ldots \end{bmatrix},$$

or a combination of these types such as $[S_2, :, S_3]$, $[S_2, S_5]$, etc. Here S_2 contains consecutive crosses throughout the top row, and S_3 throughout the bottom row. In $[S_2, S_5]$ and all such combinations the total number of columns is understood to be ν. But in all these there are at least ν zeros. Hence S_1 alone satisfies the minimum condition; which proves the theorem.

In this case it is further verifiable that all the matrices $P \pm Q$, PQ, QP, PQ^{-1}, $Q^{-1}P$ are of type R, provided that Q is non-singular.

The case of continuants of an odd order (when ν is even) is more complicated. It is impossible to have fewer than ν zeros; and inspection shews that the only cases when there are exactly ν zeros are as follows,

$$S_2,\ S_3,\ S_4,\ S_5,\ [S_2, S_5],\ [S_4, S_2],\ [S_3, S_5],\ [S_4, S_3],$$

where each S_i has an *even* number of columns, and each of these eight schemes has exactly ν columns. For example, when $\nu = 10$ one of the arrangements of S_2, S_5 is

$$[S_2, S_5] = \begin{bmatrix} \times & \times & \times & \times & \bullet & \times & \bullet & \times & \bullet & \times \\ \bullet & \bullet & \bullet & \bullet & \bullet & \times & \bullet & \times & \bullet & \times \end{bmatrix},$$

where the number of arbitrary elements belonging to S_2 is any even number less than ν. Similar remarks refer to the matrices T. By combining these in all possible ways the various types to which P and Q may belong are ascertained. Of these the most interesting are those when S and T are of type S_4 or S_5. For then P and Q are each of the same type, chosen from among the following possibilities:

$$R_1 = \begin{bmatrix} \times & \times & & & & & \\ \times & \times & \bullet & & & & \\ & \bullet & \times & \times & & & \\ & & \times & \times & \bullet & & \\ & & & & \ddots & & \\ & & & & & & \times \end{bmatrix}, \quad R_2 = \begin{bmatrix} \times & \bullet & & & & & \\ \bullet & \times & \times & & & & \\ & \times & \times & \bullet & & & \\ & & \bullet & \times & \times & & \\ & & & \times & \times & & \\ & & & & & \ddots & \end{bmatrix},$$

$$R_3 = \begin{bmatrix} \times & \bullet & & & & & \\ \times & \times & \times & & & & \\ & \bullet & \times & \bullet & & & \\ & & \times & \times & \times & & \\ & & & \bullet & \times & \bullet & \\ & & & & & \ddots & \end{bmatrix}, \quad R_4 = \begin{bmatrix} \times & \times & & & & & \\ \bullet & \times & \bullet & & & & \\ & \times & \times & \times & & & \\ & & \bullet & \times & \bullet & & \\ & & & \times & \times & & \\ & & & & & \ddots & \end{bmatrix}.$$

Here the first two types are like R but with an extra diagonal element $\times$ added at one or other end of the diagonal, to provide the necessary odd total number of rows and columns. In each case

https://doi.org/10.1017/S0305004100016406 Published online by Cambridge University Press

$P \pm Q$, PQ, QP, PQ^{-1}, $Q^{-1}P$ are all of the same type: so that we have now obtained, in all, five types R, R_1, R_2, R_3, R_4 of continuant, each of which defines a *field*: namely that *all matrices belonging to any one such type may be combined by addition, subtraction, multiplication and division into further matrices still of the same type.*

This however is not the case with matrices P and Q, which are derived from the alternative schemes S_2, S_3 or $[S_i, S_j]$. It is however interesting to note that the schemes S_2 and S_3 lead to products PQ which have actually arisen in the matrix treatment of continued fractions. Apart from the trivial case when P or Q is merely a diagonal matrix, the characters of P and Q derived from taking S and T to be of types S_2 and S_3 are

$$P = \begin{bmatrix} \times & \times & & \\ \bullet & \times & \times & \\ & \bullet & \times & \times \\ & & & \ddots \end{bmatrix}, \quad Q = \begin{bmatrix} \times & \bullet & & \\ \times & \times & \bullet & \\ & \times & \times & \\ & & \times & \ddots \end{bmatrix}.$$

But these are exactly the types of factors whose product PQ has been used by E. T. Whittaker(2) in his exposition of the work of Stieltjes upon the expansion of a continued fraction in the form of a power series.

3. *Classical canonical forms.*

The type P which has just been stated can be written as $D_0 + D_1$, which also is the type assumed by reducing any given square matrix X to its classical canonical form (Ref. (3), 58). Thus if the *characteristic equation*

$$|X - \lambda I| \equiv |x_{ij} - \lambda \delta_{ij}| = 0 \qquad (1)$$

is solved, then the resulting $\nu + 1$ values of λ are called the *latent roots* of the matrix X, and equally of its canonical form C. These latent roots, grouped according to possible repetitions, constitute the $\nu + 1$ elements of the diagonal D_0: let us say

$$D_0 = \operatorname{diag}(\lambda_i) = \operatorname{diag}(\alpha, \alpha, \ldots, \alpha, \beta, \beta, \ldots, \gamma, \ldots). \qquad (2)$$

The over-diagonal D_1 is then not arbitrary but is determined as a sequence of ν elements, each equal to unity or to zero, according to the *Segre characteristic* of the original matrix X (or of its canonical form). An actual example makes this clearer. A possible canonical form of a five-rowed matrix with latent roots $\alpha, \alpha, \alpha, \beta, \beta$ is

$$C = HXH^{-1} = \begin{bmatrix} \alpha & 1 & \bullet & \bullet & \bullet \\ \bullet & \alpha & \bullet & \bullet & \bullet \\ \bullet & \bullet & \alpha & \bullet & \bullet \\ \bullet & \bullet & \bullet & \beta & 1 \\ \bullet & \bullet & \bullet & \bullet & \beta \end{bmatrix} = \operatorname{diag}\left(\begin{bmatrix} \alpha & 1 \\ & \alpha \end{bmatrix}, [\alpha], \begin{bmatrix} \beta & 1 \\ & \beta \end{bmatrix} \right),$$

where $\quad D_0 = \operatorname{diag}_0(\alpha, \alpha, \alpha, \beta, \beta), \quad D_1 = \operatorname{diag}_1(1, 0, 0, 1). \qquad (3)$

This canonical matrix consists of three *latent matrices* placed diagonally and of orders 2, 1, 2 repectively. Each such latent matrix is either a single element (a latent root α) when its order μ is unity, or else is of order $\mu > 1$ and of type

$$C_\mu(\alpha) \equiv \alpha I + U, \tag{4}$$

where naturally αI contains a diagonal of μ equal elements α, while U contains $\mu - 1$ units upon the over-diagonal. If $\alpha \neq \beta$ the Segre characteristic is written $[\{21\}\, 2]$, but if $\alpha = \beta$ it is $\{212\}$. The numbers enclosed in brackets $\{\}$, or else standing separately, are called the *exponents of the elementary divisors* of X with regard to the latent root involved: or briefly they are the exponents of the latent root α. Evidently they are the orders of the respective latent matrices involving α.

The matrix X may of course possess but one latent matrix, in which case its canonical form is $C_{\nu+1}(\alpha)$, and D_1 simply becomes the auxiliary unit matrix U. Or again X may possess $\nu + 1$ latent matrices, in which case D_1 is entirely zero. Between these extremes come the more typical cases when D_1 is a diagonal possessing runs of consecutive non-zeros interspersed with zero gaps.

I have introduced this term *latent matrix* deliberately, for it seems to be the most natural term to meet a want which certainly exists in the accepted nomenclature.

4. *Length and range.*

The length of a diagonal has already been defined in the obvious way, but it will be useful to add a further definition, that of the *range*: namely, *the range is the number of elements in a sequence, counted from the first non-zero to the last non-zero element inclusive.* Length and range may apply not only to diagonal but also to row or column or sequence of any such kind.

For example, the sequence 0, 1, 5, 0, 2, 0 has a length six and a range four. Or again, in a continuant the sequence of diagonals D_{-1}, D_0, D_1 has a range three, when at least one non-zero element occurs in D_{-1} and at least one in D_1.

Manifestly the range ρ cannot exceed the length l of a sequence. When $\rho = l$ let the sequence be called *complete*, when $\rho < l$, *incomplete*. Further, when no zero gap occurs within the range let it be called *unbroken* or *close*, and when a gap or gaps occur, *broken* or *open*.

These ideas enable us to formulate the varieties of behaviour when the product of any two diagonal matrices D_r and D_s is formed according to Theorem 1, where, as before, r and s may be positive or negative.

THEOREM 4. *If $rs \geqslant 0$, the product D_{r+s} of two complete diagonals D_r, D_s, when non-zero, is complete, the length l and range ρ of the product being given by*

$$l = \nu + 1 - |r+s| = \rho.$$

Proof. In the notation of Theorem 1 let the single-term element z_{ij} at the ijth position in the product be now denoted by $\binom{i\ j}{k}$, where $k = i + r$, $j = k + s$. If $rs > 0$, then k must lie between i and j. If $rs = 0$, then k must equal i or j. In either case the resultant $(r+s)$th diagonal is completely filled, provided that the original diagonals D_r, D_s are complete and that $-\nu \leqslant r+s \leqslant \nu$.

For example, if $\nu = 9$, $r = 2$, $s = 5$, then

$$\binom{i\ j}{k} = \binom{0\ 7}{2}, \quad \binom{1\ 8}{3}, \quad \binom{2\ 9}{4}.$$

The number of terms is l or ρ, and is determined by the steps from the initial position (0, 7) to the final (2, 9). The above formula will be seen to include all cases whenever $rs \geqslant 0$. Hence the theorem is proved.

THEOREM 5. *If $rs < 0$, the product of the two complete diagonals is incomplete but unbroken, and such that its length and range satisfy the relation*

$$l = \nu + 1 - |r+s| > \rho.$$

Proof. If $rs < 0$, either $r = -r' < 0$ and $s > 0$ or else $r > 0$ and $s = -s' < 0$. The typical term of the product diagonal is then given by

$$\binom{i\ j}{k} = \binom{k+r'\ \ k+s}{k} \text{ or } \binom{k-r\ \ k-s'}{k},$$

where in each case k does not lie between i and j. In the former of these cases k takes successive values from zero upwards until the larger of i and j attains the value ν. The range is then unbroken but incomplete, since a complete range necessarily would have started with a zero value of the smaller of i and j. For example, $\nu = 9$, $r = -5$, $s = 2$,

$$\binom{i\ j}{k} = \binom{5\ 2}{0}, \quad \binom{6\ 3}{1}, \quad \binom{7\ 4}{2}, \quad \binom{8\ 5}{3}, \quad \binom{9\ 6}{4},$$

and the positions where $(i, j) = (3, 0), (4, 1)$ are necessarily filled with zeros. Such a diagonal is indicated by

$$\bullet\ \bullet\ \times\ \times\ \times\ \times\ \times \qquad (k < i,\ k < j,\ \rho < l).$$

https://doi.org/10.1017/S0305004100016406 Published online by Cambridge University Press

Again, in the latter of these cases k may take successive ascending values but concluding with the value ν; and the argument is analogous. For example, $\nu = 9$, $r = 2$, $s = -5$,

$$\begin{pmatrix} i & j \\ k & \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 5 & \end{pmatrix}, \quad \begin{pmatrix} 4 & 1 \\ 6 & \end{pmatrix}, \quad \begin{pmatrix} 5 & 2 \\ 7 & \end{pmatrix}, \quad \begin{pmatrix} 6 & 3 \\ 8 & \end{pmatrix}, \quad \begin{pmatrix} 7 & 4 \\ 9 & \end{pmatrix},$$

and the two final positions ($(i, j) = (8, 5), (9, 6)$) upon this third negative diagonal ($r + s = -3$) must contain zeros. Thus

$$\times \times \times \times \times \bullet \bullet \qquad (k > i,\ k > j,\ \rho < l).$$

It will be seen that the deficiency $l - \rho$ is equal to the smaller of $|r|$ and $|s|$.

Corollaries. If ρ_x and l_x denote the range and length of $D_r = \mathrm{diag}_r(x)$, while ρ_y and l_y denote those of $D_s = \mathrm{diag}_s(y)$, and if

$$\rho_x = l_x, \quad \rho_y = l_y,$$

then

(i) $\rho_x > \rho$, $\rho_y > \rho$, when $rs > 0$;

(ii) $\rho_x > \rho = \rho_y$, when $r = 0$, $s \neq 0$;

(iii) $\rho_x = \rho < \rho_y$, when $r \neq 0$, $s = 0$;

(iv) $\rho_x > \rho$, $\rho_y > \rho$, when $rs < 0$.

These are easily verified.

It is also to be noted that a principle of duality underlies this theory of ranges, as may be seen by attaching to each suffix group i, j, k a *dual* group i', j', k', such that

$$i + i' = j + j' = k + k' = \nu + 1.$$

Similar rules can also be formulated for continued products such as

$$\mathrm{diag}_r(x) . \mathrm{diag}_s(y) . \mathrm{diag}_t(z) \ldots = \mathrm{diag}_{r+s+t+\ldots}(w).$$

5. *Intensity.*

By the *intensity* of a diagonal will be meant the actual sequence of values of its elements. Such elements form a vector, let us say

$$u = (a, b, c, d, \ldots). \qquad (1)$$

consisting of p elements. If $D_r = \mathrm{diag}_r(u)$, $D_s = \mathrm{diag}_s(u)$, then D_r and D_s will have the same intensity. A sequence of p consecutive elements of a diagonal whose length exceeds p is called a *subdiagonal of length p*. Evidently two such subdiagonals lying in the same or in parallel lines may have equal intensities.

From a principal diagonal matrix $D_0 = \mathrm{diag}(x_i)$ further diagonal matrices may be derived by forming its powers. They will usually

https://doi.org/10.1017/S0305004100016406 Published online by Cambridge University Press

differ in intensity but will not change position. For this reason it is useful to call this the *static* diagonal. For example,

$$D_0^q = \mathrm{diag}_0 (x_0^q, x_1^q, \ldots, x_\nu^q), \tag{2}$$

and, if $f(x_i)$ is a scalar function of a single argument x_i,

$$f(D_0) = \mathrm{diag}_0 (f(x_0), f(x_1), \ldots, f(x_\nu)), \tag{3}$$

as may easily be verified.

From an adjacent diagonal $D_{\pm 1}$ further and parallel diagonal matrices may be obtained by forming its powers. The diagonal changes position as well as intensity in the following characteristic way,

$$\left.\begin{aligned} X_1 &= \mathrm{diag}_{\pm 1}(x_0, x_1, x_2, \ldots) \\ X_1^2 &= \mathrm{diag}_{\pm 2}(x_0x_1, x_1x_2, x_2x_3, \ldots) \\ X_1^3 &= \mathrm{diag}_{\pm 3}(x_0x_1x_2, x_1x_2x_3, x_2x_3x_4, \ldots) \end{aligned}\right\}, \tag{4}$$

etc., where the same sign is taken throughout: X_1^q is then a diagonal whose typical element contains exactly q *consecutive* elements x_i as factors, until $q \geqslant \nu + 2$, when $X_1^q = 0$. The intensities, formed in this way, all consist of *close-range* products of the x_i.

Once more, from further diagonals $D_{\pm r}$ $(r > 1$, or $r < -1)$, the various powers can be evaluated thus:

$$\left.\begin{aligned} X_r &= \mathrm{diag}_{\pm r}(x_0, x_1, x_2, \ldots) \\ X_r^2 &= \mathrm{diag}_{\pm 2r}(x_0x_r, x_1x_{r+1}, \ldots) \\ X_r^3 &= \mathrm{diag}_{\pm 3r}(x_0x_rx_{2r}, x_1x_{r+1}x_{2r+1}, \ldots) \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ X_r^q &= \mathrm{diag}_{\pm qr}(x_0x_rx_{2r}\ldots x_{(q-1)r}, x_1x_{r+1}x_{2r+1}\ldots x_{(q-1)r+1}, \ldots) \end{aligned}\right\}. \tag{5}$$

Also if q is so large that $(q-1)r$ exceeds each suffix of x, then $X_r^q = 0$. The suffixes of a typical element in a power of X are in arithmetical progression with common difference r exceeding unity; and these intensities are said to be formed from *open-range* products of the x_i.

Now let there be two arbitrary diagonal matrices

$$D = D_r = \mathrm{diag}_r (x_0, x_1, \ldots, x_h) \qquad (r = \pm r'),$$
$$D_s = \mathrm{diag}_s (y_0, y_1, \ldots, y_k) \qquad (s = \pm s'),$$

where $r' \geqslant 0$, $s' \geqslant 0$ and where $\nu - r' = h$, $\nu - s' = k$. Certain properties which are useful can now easily be verified.

(i) *The product of two close-range diagonals is either close-range or else zero.*

This means that all the non-zeros among the x_i are consecutive, as also among the y_i.

https://doi.org/10.1017/S0305004100016406 Published online by Cambridge University Press

(ii) *The matrices D_r, D_s commute identically for all x_i and y_i if and only if $r = s = 0$.* That is,

$$\operatorname{diag}_0(x)\operatorname{diag}_0(y) \equiv \operatorname{diag}_0(x_i y_i) \equiv \operatorname{diag}_0(y)\operatorname{diag}_0(x);$$
$$\operatorname{diag}_r(x)\operatorname{diag}_s(y) \not\equiv \operatorname{diag}_s(y)\operatorname{diag}_r(x) \qquad (r' + s' > 0).$$

Next let r' be positive and not zero, and let p be an integer such that

$$(p-2)r' \leqslant h < (p-1)r'.$$

Then it follows that

(iii) $$D_r^{p-1} \not\equiv 0, \quad D_r^p \equiv 0 \qquad (\pm r = r' > 0).$$

This follows from the relations (5) above. Again

(iv) *The latent roots of D_r ($r \neq 0$) are all zero: and D_r^p is called the reduced characteristic function of D_r.*

It is well known (Ref. (3), 48), and is here easily verified, that D_r can satisfy no polynomial relation of degree lower than p. On the other hand the latent roots of a principal diagonal D_0 are the elements themselves.

6. *Subdiagonals of an open-range diagonal.*

The following notation is convenient for the purpose of representing the most general type of non-zero diagonal matrix:

$$D_r = \operatorname{diag}_r\left(O_{\omega_0}, \Delta_{\rho_1}, O_{\omega_1}, \Delta_{\rho_2}, O_{\omega_2}, \ldots, \Delta_{\rho_\mu}, O_{\omega_\mu}\right), \qquad (1)$$

where, for each value of i, O_{ω_i} denotes a sequence of ω_i zeros, and Δ_{ρ_i} a sequence of ρ_i non-zeros. The diagonal is now said to be resolved into μ close-range subdiagonals, where $\mu \geqslant 1$. The suffixes are such as to satisfy the relations

$$\omega_0 \geqslant 0, \quad \rho_1 > 0, \quad \omega_\mu \geqslant 0, \quad \mu \geqslant 1,$$
$$\omega_0 + \rho_1 + \omega_1 + \rho_2 + \ldots + \rho_\mu + \omega_\mu = l = \omega_0 + \rho + \omega_\mu, \qquad (2)$$

which are in keeping with the existing definitions of the length l and the range ρ. As before it is useful to write

$$\pm r = r' \geqslant 0. \qquad (3)$$

It now appears that there are two distinct types of diagonal matrix, the *regular* and the *irregular*, which are defined as follows. In the regular case either $r = 0$ or, when $\mu > 1$,

$$r' \leqslant \omega_i \qquad (i = 1, 2, \ldots, (\mu - 1)), \quad (4)$$

and in the irregular case r' exceeds at least one of the suffixes ω_i belonging to intermediate runs of zeros. The case when $\mu = 1$ has no such intermediate run of zeros, and is classed as regular.

THEOREM 6. *The powers of a regular diagonal are found by powering each of its subdiagonals independently.*

Proof. The case when $r=0$ is obvious. Next assume that $0<r'\leqslant\rho_i$ for each subdiagonal and that $r'\leqslant\omega_i$ as in (4). Then evidently

$$D_r^2=\mathrm{diag}_{2r}\left(O_{\omega_0},\ \Delta_{\rho_1-r'},\ O_{\omega_1+r'},\ \Delta_{\rho_2-r'},\ \ldots,\ \Delta_{\rho_\mu-r'},\ O_{\omega_\mu}\right),$$

where $\Delta_{\rho_i-r'}$ denotes the second power of Δ_{ρ_i}, for each of the μ values of ρ_i, and where each of the intermediate null ranges is swelled by an addition of r' zero elements. Powers of this open diagonal D_r accordingly proceed by powering each separate subdiagonal—shortening its length by r' elements at a time, and increasing the intermediate null elements a like amount. If $r>0$, the leading element of *each* close-range Δ_{ρ_i} remains in its own row of the matrix (and, if $r<0$, in its own column) until a power is reached when the particular subdiagonal is annihilated. Put another way, *if $r>0$, each leading element of a close-range moves horizontally to the right when D_r is powered, while the final element of the same close-range moves at a constant inclination upwards to the right.*

Next, the provisional condition, that $r'\leqslant\rho_i$, may be removed. For suppose that at the $(q+1)$th power Δ_{ρ_2} is annihilated whereas the adjacent Δ_{ρ_1} and Δ_{ρ_3} are not. Then the gaps O_{ω_1} and O_{ω_2} are merged in a larger gap

$$O_{\omega_1+\rho_2+\omega_2+qr'}\equiv O_{\omega'}.$$

In such a case

$$D_r^{q+1}=\mathrm{diag}_{(q+1)r}\left(O_{\omega_0},\ \Delta_{\rho_1-qr'},\ O_{\omega'},\ \Delta_{\rho_3-qr'},\ \ldots\right),$$

and the original properties of the still existing non-zero subdiagonals are preserved. Hence condition (4) alone is essential to the case: and this proves the theorem.

Thus it will be seen that all regular diagonals (excepting the case when $r=0$) tend to lose range when powered. Irregular diagonals may however increase range under this same process; but the rule is exceedingly complicated in general. Consider, for example, the matrix

$$D_r=\mathrm{diag}_r\left(O_3,\ \Delta_4,\ O_2,\ \Delta_3,\ O_1\right),$$

where there are two subdiagonals, Δ_4 and Δ_3, in an open-range 9 within a length 13. This diagonal is regular if $r\leqslant 2$, but irregular when r exceeds the suffix of O_2. Now let D_r^2 be formed. According to the rule for the regular case Δ_3 should disappear when $r=3$, yet here it becomes Δ_1. For $r=4, 5, 6, 7, 8$ there is but one subdiagonal in D_r^2, and its range is 2, 3, 3, 2, 1 respectively.

https://doi.org/10.1017/S0305004100016406 Published online by Cambridge University Press

7. *Invariant factors of a close-range diagonal.*

By an *equable partition* of a positive integer N into r parts is meant the *unique* partition

$$((n+1), (n+1), \ldots, n, n, \ldots, n), \tag{1}$$

each part being equal either to n or to $n+1$. Here

$$N = nr + s \qquad (s < r), \tag{2}$$

so that s is the remainder when N is divided by r. There are consequently s parts equal to $n+1$ and $r-s$ parts equal to n. For example (3, 2, 2), (2, 2, 2) are the equable partitions of 7 and 6 respectively into three parts. This definition is very useful as a way of specifying the canonical form of a given close-range diagonal matrix, all of whose elements are non-zero:

$$D_r = \mathrm{diag}_r (u_0, u_1, \ldots, u_{\rho-1}), \tag{3}$$

where $r > 1$, all $u_i \neq 0$, $r + \rho = \nu + 1$.

Such a matrix has all its $\nu + 1$ latent roots equal to zero, and, in such a case, a certain process of chain formation which is due to a remark of A. C. Aitken (Ref. (3), 66, 76) leads at once to the following theorem.

THEOREM 7. *The Segre characteristic of a complete non-principal diagonal $D_{\pm r}$ is identical with the equable partition of the integer $\nu + 1$ into r parts.*

Proof. Let (i, j) denote the element of D_r which occupies row i and column j. Then the ρ non-zero elements are given by

$$(i, j) = (0, r), (1, r+1), \ldots, (\rho-1, r+\rho-1). \tag{4}$$

Now let every rth element, counted from $(k, r+k)$, be selected and set down as a chain called θ_k. Thus

$$\theta_k = ((k, r+k), (r+k, 2r+k), (2r+k, 3r+k), \ldots). \tag{5}$$

In this way r chains, $\theta_0, \theta_1, \ldots, \theta_{r-1}$, can be formed, including all the non-zeros of D_r without repetitions, each chain having $\sigma + 1$ or else σ members, where

$$(\sigma+1, \sigma+1, \ldots, \sigma, \ldots)$$

is the equable partition of ρ into r parts. Take therefore

$$\rho = r\sigma + s \qquad (0 \leqslant s < r). \tag{6}$$

It then follows from the remark of A. C. Aitken that an identity exists of the following form,

$$H_r^{-1} D_r H_r = \mathrm{diag}_1 (u_0, u_r, u_{2r}, \ldots, u_{\sigma r}, 0, u_1, \ldots, 0, u_2, \ldots, \ldots), \tag{7}$$

which may be written still more simply as

$$D_1 = \mathrm{diag}_1 (\theta_0, 0, \theta_1, 0, \ldots, 0, \theta_{r-1}), \tag{8}$$

https://doi.org/10.1017/S0305004100016406 Published online by Cambridge University Press

where H_r is a suitably chosen non-singular matrix which operates by bringing this rth diagonal to the over-diagonal position. But the Segre characteristic of D_r is given by that of D_1, which in turn is obtained by adding unity to the number of each set θ_k of non-zero elements:

$$(\sigma + 2, \sigma + 2, \ldots, \sigma + 1, \ldots).$$

This proves the theorem.

It will be seen that the length of the diagonal D_1 is $\rho + r - 1 = \nu$, as it should be, there being ρ non-zeros equably partitioned, and $r - 1$ intermediate zeros. The length of D_r is ρ, and that of the principal diagonal $\nu + 1$.

The same result may be obtained directly by constructing the matrix H_r which produces this reduction. In fact, let H_r be formed by inserting in a blank matrix exactly $\nu + 1$ units in the positions

$$(0, 0), (r, 1), (2r, 2), \ldots, (\nu r, \nu), \mod(\nu + 1), \qquad (9)$$

by which is meant that each index i is always taken to be less than $\nu + 1$ by subtracting, if necessary, a suitable multiple of $\nu + 1$. Such a process may naturally be called the equable distribution of units at a gradient r within the matrix, and it is said to form *a unit matrix of gradient r*. For example, if $\nu = 6$, $r = 2$, then

$$H_2 = \begin{bmatrix} 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot & \cdot \end{bmatrix}. \qquad (10)$$

The units lie upon r parallel oblique lines whose common falling gradient is r. Such a matrix is obviously non-singular and orthogonal; hence its transposed is equal to its reciprocal matrix. Thus

$$H_2^{-1} = H_2' = H_{\frac{1}{2}}, \qquad (11)$$

where the last suffix indicates a gradient of $\frac{1}{2}$, which is actually the case. It is easy to verify the identity by calculating the product matrices in the form

$$D_r H_r = H_r D_1.$$

Alternatively the matter may be treated by the methods of linear transformations, which are well illustrated by the case when $\nu = 6$, $r = 2$. Beginning with the substitution

$$\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 4 & 1 & 5 & 2 & 6 & 3 \end{pmatrix}, \qquad (12)$$

as suggested by the column succession in the matrix H_2, let us form the following relations:

$$\left.\begin{aligned} &\eta_0 = y_0 = x_2 = \xi_1, && \\ &\eta_1 = y_2 = x_4 = \xi_2, && \\ &\eta_2 = y_4 = x_6 = \xi_3, && \xi_0 = x_0, \\ &\eta_3 = y_6 = 0, && \xi_4 = x_1, \\ &\eta_4 = y_1 = x_3 = \xi_5, && \\ &\eta_5 = y_3 = x_5 = \xi_6, && \\ &\eta_6 = y_5 = 0. && \end{aligned}\right\} \qquad (13)$$

These are merely the expanded statement of certain matrix equations,

$$y = D_r x, \quad x = H_r \xi, \quad y = H_r \eta, \qquad (14)$$

holding between four sets of variables x, y, ξ, η, each with $\nu + 1$ components, and each treated as a column vector. These equations between x and y clearly assert the matrix equation $y = D_r x$. The further equations between x and ξ are derived directly from the above substitutional scheme (as inspection will shew). In this way H_r is derived, and in turn the relations between y and η.

But the equations (13) shew at once that ξ and η are connected by a relation of the form

$$\eta = D_1 \xi,$$

while the matrix equations (14) shew that

$$\eta = H_r^{-1} D_r H_r \xi.$$

This yields the desired identity, which reduces D_r to D_1. In this illustration each non-zero element u_i of D_r has been taken to be unity, but there is no difficulty in writing down the corresponding results when the non-zeros are arbitrary.

THEOREM 8. *A regular diagonal matrix $D_{\pm r}$ $(r > 0)$, whose subdiagonals have ranges ρ_i, may be brought to a canonical form D_1 whose subdiagonals are derived by equable partition of each ρ_i into r parts.*

The Segre characteristic of $D_{\pm r}$ is the set of positive integers obtained from the equable partition of $\rho_i + r$ into r parts, applied to each separate subdiagonal of $D_{\pm r}$, and supplemented if necessary by units.

Proof. The second part of the enunciation is merely an alternative version of the first. Each separate non-zero range ρ' of D_1 contributes an exponent $\rho' + 1$ to the Segre characteristic. When the sum $\Sigma(\rho' + 1)$ falls short of $\nu + 1$ the supplementary unit exponents are necessary.

https://doi.org/10.1017/S0305004100016406 Published online by Cambridge University Press

The actual values of the ρ' follow at once from the preceding investigation by the process of chain formation. For each subdiagonal gives rise to separate chains, it being impossible for a chain to bridge the gap between adjacent non-zero subdiagonals when the whole diagonal is regular. The actual reduction is effected as follows:

If $$D_r = \mathrm{diag}_r (O_{\omega_0}, \Delta_{\rho_1}, O_{\omega_1}, \Delta_{\rho_2}, \ldots, O_{\omega_\mu}),$$
where $r > 1$, $r < \omega_i$, then take H to be of the form
$$H = \mathrm{diag}_0 (I_{\omega_0}, H_{(\rho_1+r)}, I_{\omega_1-r}, H_{(\rho_2+r)}, \ldots, I_{\omega_\mu}),$$
where $H_{(i)}$ is a unit matrix of gradient r and having i rows and columns, and where I_i is a unit matrix having i rows and columns. The suffixes of these subdiagonals I are necessarily non-negative in the regular case: if they are zero the corresponding letters I are merely suppressed.

This matrix H leads, by what has already been proved, to a formula of the type
$$H^{-1} D_r H = D_1,$$
where D_1 possesses the properties stated in the enunciation of the theorem.

Kinematically, this matrix H may be regarded as the operation which draws each leading non-zero of the subdiagonals Δ_{ρ_i} horizontally leftwards on to the first over-diagonal. The chains inherent in each Δ_{ρ_i} are thereby disentangled equably and become non-zero sets separated by single zeros (along the over-diagonal) as in (8). Between the last such set in Δ_{ρ_1} and the first in Δ_{ρ_2} the actual number of zeros will be $\omega_1 - r + 1$.

Example. When $X = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} = \mathrm{diag}_2 (0, 1, 1, 0),$

let $H = \mathrm{diag}_0 (1, H_{(4)}, 1)$. Then
$$H^{-1} X H = \mathrm{diag}_1 (0, 1, 0, 1, 0),$$
a matrix whose Segre characteristic is evidently $\{2, 2, 1, 1\}$.

In full we have
$$H = \begin{bmatrix} 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix}.$$

https://doi.org/10.1017/S0305004100016406 Published online by Cambridge University Press

So far the reductions have applied to an upper diagonal D_r for which $r > 0$. The negative case which reduces D_{-r} to D_{-1} is analogous, it being only necessary to transpose every matrix in the above work. The further reduction in the negative case from D_{-1} to D_1 is effected by the identity

$$JD_{-1}J = D_1,$$

where $J\ (= J^{-1})$ is the *secondary unit matrix*. For example,

$$J = \begin{bmatrix} \cdot & \cdot & 1 \\ \cdot & 1 & \cdot \\ 1 & \cdot & \cdot \end{bmatrix}.$$

The cases when $r = -1, 0, 1$ need no consideration.

8. *Canonical form of an irregular diagonal.*

The process of chain formation yields the exact canonical form of any given diagonal matrix, but it is only possible in the regular case (when zero gaps cannot be bridged) to state a simple general rule. To illustrate the irregular case let us consider a range

$$\times \times \times \cdot \cdot \times \times \cdot \cdot \times \qquad (1)$$

of length ten and possessing two zero gaps each of two elements.

The method used in forming θ_k of § 7 (5) is available; namely, chains of non-zeros selected at equal intervals r must be formed. When this is actually done—and all the six non-zeros of the given range (1) are exhausted—the result may be tabulated as follows:

$r = \pm 1$	3, 2, 1,
$r = \pm 2$	2, 1, 1, 1, 1,
$r = \pm 3$	2, 2, 1, 1,
$r = \pm 4$	3, 2, 1,
$r = \pm 5$	2, 2, 1, 1,
$r = \pm 6$	2, 1, 1, 1, 1,
$r = \pm 7$	2, 1, 1, 1, 1,
$r = \pm 8$	2, 1, 1, 1, 1,
$r = \pm 9$	2, 1, 1, 1, 1,
$r = \pm 10, \pm 11$, etc.	1, 1, 1, 1, 1, 1.

In explanation of this table it will be noted that the non-zeros of the range occur at positions 1, 2, 3, 6, 7, 10. When, for instance, $r = \pm 4$, the subchains can only be (2, 6, 10), (3, 7), 1, obtained by arranging the position numbers in arithmetical progression with common difference 4 (which is equivalent to the rule of Dr Aitken).

https://doi.org/10.1017/S0305004100016406 Published online by Cambridge University Press

The number of elements in each chain is noted: here such numbers are 3, 2, 1. This means that the Segre characteristic of the matrix $D_{\pm 4}$ whose range is given by (1) would be $\{4, 3, 2\}$, which is obtained as before (cf. Ref. (3), 67) by adding unity to each of the integers just found.

The *regular* cases (when $r > 2$ or < -2) are here tabulated. They illustrate the equable partition of the three given non-zero subdiagonals each into one, or each into two parts. The *irregular* case is illustrated in the table by the rise in value of the integers on the third line.

It is obvious that, if the specification (1) is given, then for all values of $r' (= \pm r)$, equal to or greater than the length l of the range (1), each of the m non-zero elements in (1) is isolated and implies the existence of an elementary divisor whose exponent is equal to 2. Also the rank of such a matrix $D_{\pm r}$ is $m \leqslant l$. This establishes the curious theorem:

THEOREM 9. *If $r' \geqslant l$, the canonical form of any open or close diagonal matrix D_r whose length is l and whose rank is $m \leqslant l$ consists of exactly m latent matrices of the type* $\begin{bmatrix} \cdot & 1 \\ \cdot & \cdot \end{bmatrix}$.

9. *The rth root of a matrix.*

As an illustration of the foregoing principles I shall now consider the problem of finding the necessary and sufficient conditions for a matrix to possess an rth root. That such a problem exists may easily be shewn by attempting to find the square root of the matrix $\begin{bmatrix} \cdot & 1 \\ \cdot & \cdot \end{bmatrix}$, which has no square root *of its own type.*

By this is meant that it is impossible to find four real or complex numbers x_1, x_2, x_3, x_4 such that

$$\begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}^2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

In fact, on expanding the left-hand expression, the resulting equations

$$x_1^2 + x_2 x_3 = 0, \quad x_1 x_2 + x_2 x_4 = 1, \quad x_3 x_1 + x_4 x_3 = 0, \quad x_3 x_2 + x_4^2 = 0$$

are incompatible. On the other hand it is easily verified that

$$\begin{bmatrix} \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot \end{bmatrix}^2 = \begin{bmatrix} \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix},$$

so that the matrix on the right has a square root.

https://doi.org/10.1017/S0305004100016406 Published online by Cambridge University Press

The problem before us is to determine when it is possible to solve the equation

$$X^r = A, \tag{1}$$

where A is a given matrix(4). In the first place, if A is non-singular, so that all its latent roots are non-zero, then an rth root exists—as Frobenius(5) has shewn. Next, if A is singular, let its classical canonical form be C, where

$$H^{-1}AH = C = \text{diag}\,(C_0, C_1), \tag{2}$$

C_0 involving zero latent roots entirely, and C_1 non-zero latent roots (if any exist). Let C_0 contain σ, and C_1 μ latent matrices.

Now suppose that X exists and therefore possesses a classical canonical form Y, given, let us say, by the relation

$$K^{-1}XK = Y. \tag{3}$$

Then, by iteration, $K^{-1}X^rK = Y^r$; that is $K^{-1}AK = Y^r$, so that

$$PCP^{-1} = Y^r, \tag{4}$$

where $P = K^{-1}H$, all of P, K, H being non-singular. Hence the four matrices A, C, X^r, Y^r are equivalent, and therefore possess exactly the same sets of latent roots and elementary divisors. Also these latent roots are rth powers of those of X: consequently X must have as many zero latent roots as C_0 has, and as many non-zero latent roots as C_1 has. It will therefore be convenient to write the canonical form Y in analogous fashion as

$$Y = \text{diag}\,(Y_0, Y_1), \tag{5}$$

so that $Y^r = \text{diag}\,(Y_0{}^r, Y_1{}^r)$, where the latent roots of Y_0 are all zero, and of Y_1 all non-zero. Furthermore $Y_0{}^r$ is now equivalent to C_0 and $Y_1{}^r$ to C_1.

Next suppose that Y_1 possesses exactly μ *latent matrices* of which any one is selected for consideration: let this have a latent root α and an exponent e. Then, by the general theory (Ref. (3), 75). $Y_1{}^r$ will have exactly μ latent matrices, of which the corresponding one has a latent root α^r and an exponent e (since $\alpha \neq 0$). But $Y_1{}^r$ is equivalent to C_1: it follows immediately that the μ exponents e must coincide with those of C_1 and that the μ expressions α^r must coincide with the latent roots of C_1. Thus, from the given submatrix C_1, which is conveniently expressed as

$$C_1 = S(\alpha_1, e_1;\ \alpha_2, e_2;\ \ldots;\ \alpha_\mu, e_\mu), \tag{6}$$

we deduce the matrix

$$Y_1 = S(\sqrt[r]{\alpha_1}, e_1;\ \sqrt[r]{\alpha_2}, e_2;\ \ldots;\ \sqrt[r]{\alpha_\mu}, e_\mu), \tag{7}$$

where each latent root is indicated along with its appropriate exponent. Among these μ non-zero latent roots α_i there may of

course be repetitions: if they are all distinct there will then be μ^r distinct values of Y_1 none of which are equivalent, since all the rth roots of all the α_i will differ. The same will happen if, when $\alpha_r = \alpha_s$, $e_r \neq e_s$.

In general, if there are exactly μ_1 equal roots α_1 of which exactly μ_{11} possess the same exponent e_1, then this distribution gives rise to m_{11} different values of Y_1, where evidently m_{11} is the number of homogeneous products of r things taken μ_{11} at a time. For this will enumerate all the essentially distinct values of Y_1, as far as these cited repetitions are concerned.

If, further, exactly μ_{12} of the α_1 possess a new exponent e_2, they give rise to a similar number m_{12} of values—in all to $m_{11}m_{12}$ distinct values. And so on, until all such subclasses among the repetitions of the roots α and of their exponents e are dealt with.

In this way the exact number of distinct (non-equivalent) matrices Y_1 is obtained, as rth roots of a given non-singular matrix C_1.

We turn now to the purely singular matrix C_0, which may conveniently be specified in terms of its Segre characteristic by

$$C_0 = S_0(e_1, e_2, \ldots, e_\sigma) \qquad (e_1 \geqslant e_2 \geqslant \ldots \geqslant e_\sigma > 0), \qquad (8)$$

where σ is the number of its latent matrices. Since all the latent roots of C_0 are zero, this canonical form can alternatively be denoted by

$$C_0 = \mathrm{diag}_1(I_{e_1-1}, 0, I_{e_2-1}, 0, \ldots, I_{e_\mu-1}, O_\omega), \qquad (9)$$

that is, by a range of $e_1 - 1$ units followed by a zero, then by $e_2 - 1$ units, then a zero, and terminating with possibly ω zeros. (This O_ω will not occur if $e_\sigma > 1$, but is needed if one or more of the e_σ are equal to 1.)

It will now be shewn that C_0 possesses an rth root when, and only when, the integers e_i satisfy certain properties depending upon the equable partition of an integer into r parts.

To shew this let

$$Y_0 = S_0(e_1', e_2', \ldots, e_q') \qquad (e_1' \geqslant e_2' \geqslant \ldots \geqslant e_q' > 0) \quad (10)$$

be the canonical form of the supposed rth root of C_0, so that each latent root of Y_0 is necessarily zero. Then this matrix can evidently be written in the alternative form

$$Y_0 = \mathrm{diag}_1(I_{e_1'-1}, 0, I_{e_2'-1}, 0, \ldots, 0, I_{e_q'-1}), \qquad (11)$$

where each $I_{e_k'-1}$ denotes a range of $e_k' - 1$ units if $e_k' > 1$, but is entirely omitted if $e_k' = 1$. The $q - 1$ zeros here indicated are all present whether or not an I is omitted. Now let the rth power of Y_0 be formed. If none of the e_i' exceeds r, this power is zero. We infer that, *if C_0 is the null matrix of s rows and columns, then its*

most general rth root is equivalent to the above matrix Y_0, where the positive integers e_i' are chosen subject only to the conditions

$$e_1' + e_2' + \ldots + e_q' = s,$$
$$r \geqslant e_1' \geqslant e_2' \geqslant \ldots \geqslant e_q' > 0. \qquad (12)$$

In all other cases some, let us say $p\,(\leqslant q)$, of the exponents e_i' will exceed r. The rth power of Y_0 will now take the form

$$Y_0^r = \mathrm{diag}_r\,(I_{e_1'-r},\ O_r,\ I_{e_2'-r},\ \ldots,\ O_r,\ I_{e_p'-r},\ O_\omega), \qquad (13)$$

where there are $p-1$ sets O_r each consisting of r consecutive zeros, and a final set of ω zeros. But this is evidently a *regular* diagonal matrix, and, according to Theorem 8, its canonical form will have exponents e_1'', e_2'', ..., obtained by adding r to each of the p suffixes of the I's, and then partitioning each of the p resulting integers e_i' equably into r parts, and finally attaching just enough unit exponents ($e_k'' = 1$) as shall make the total sum of all exponents equal to s, the order of the matrix Y_0. But Y_0 is a perfectly general matrix, all of whose latent roots are zero. Hence every such matrix (other than the null matrix) which is a perfect rth power must have for its exponents numbers of this particular type e'': and this applies to the exponents of C_0. In other words, *the σ exponents e_i of C_0 must be capable of being distributed into sets each of which is an equable partition of the sum of the exponents within the set, and such that each set which does not possess exactly r members must consist entirely of unit exponents.*

This, which may be referred to as *the condition of equability*, is evidently both necessary and sufficient for C_0 to possess an rth root. It includes the case when $C_0 = 0$.

The significance of these results can best be seen by actual examples:

(i) $C_0 = \begin{bmatrix} \cdot & 1 \\ \cdot & \cdot \end{bmatrix}$. Here $C_0 = S_0(2)$, which has a single exponent other than unity. It is incapable of equable distribution, and therefore possesses neither square, cube nor higher root.

(ii) $C_0 = \begin{bmatrix} \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}$ possesses a square root, $\begin{bmatrix} \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot \end{bmatrix}$. Here

$$C_0 = S_0(2, 1);$$

and (2, 1) is an equable distribution of an integer into two parts. There is neither cube nor higher root.

(iii) $C_0 = S_0(4, 3, 3, 3, 3, 2, 2, 2, 1, 1, 1, 1, 1, 1)$. This singular matrix, containing twenty-eight zero latent roots and fourteen exponents, has square, cube, fourth and fifth but neither sixth nor higher roots. The possible fifth roots are governed by the equable

https://doi.org/10.1017/S0305004100016406 Published online by Cambridge University Press

groupings (4, 3, 3, 3, 3) and (2, 2, 2, 1, 1), followed by all possible partitions of the number $(1+1+1+1)$. There are exactly five essentially non-equivalent fifth roots,

$$S_0(16, 8, 4),$$
$$S_0(16, 8, 3, 1),$$
$$S_0(16, 8, 2, 2),$$
$$S_0(16, 8, 2, 1, 1),$$
$$S_0(16, 8, 1, 1, 1, 1).$$

The sixth root is excluded since the six highest exponents do not form an equable grouping. There are eleven distinct non-equivalent square roots, four arising from the equable distribution (4, 3), (3, 3), (3, 2), (2, 2), and four more from the distribution (4, 3), (3, 2), (3, 2), (3, 2), and again three more from (4, 3), (3, 3), (3, 2), (2, 1), (2, 1). For example, these last three are

$$S_0(7, 6, 5, 3, 3, 2, 2),$$
$$S_0(7, 6, 5, 3, 3, 2, 1, 1),$$
$$S_0(7, 6, 5, 3, 3, 1, 1, 1, 1).$$

When C_0 satisfies the condition of equability let the number of its distinct (non-equivalent) rth roots be N. Evidently this number can be evaluated, as above, in any definite example. Further, let the number of distinct rth roots of C_1 be M, as already found. Then the number of distinct (non-equivalent) rth roots of C (and equally well of A) is evidently MN.

In such a case it is possible to write down the most general matrix X which satisfies the equation

$$X^r = A.$$

For X is given by $X = KYK^{-1}$, where $K = HP^{-1}$ (by (4)). Also $PC = Y^rP$ may be regarded as an equation to determine P, which can be solved in its most general form. If A is a given matrix, then H may also be regarded as given, so that the matrix X is determined by

$$X = HP^{-1}YPH^{-1}. \qquad (14)$$

Here Y can take any of its MN distinct values; and P contains a definite and known number (let us say n) of arbitrary constants: P^{-1} will involve the same arbitrary constants, while H and Y have none. At first sight it might be supposed that X consequently has n arbitrary constants. Unfortunately this is not the case, since some or all of them may cancel out of this product expression. As far as I am aware, this matter has never been completely settled, and the problem of finding the exact number of arbitrary constants, even in the solution of the equation $X^2 = A$, still awaits treatment.

10. *Complex matrices.*

It is possible by a very simple device to arrange that every matrix should possess an rth root. All that is necessary is to prescribe that *each element of the matrix should be regarded as a scalar matrix of r rows and columns.* For example, the square root of $\begin{bmatrix} \cdot & 1 \\ \cdot & \cdot \end{bmatrix}$ may be found by interpreting this non-zero unit element as $\begin{bmatrix} 1 & \cdot \\ \cdot & 1 \end{bmatrix}$ and the matrix itself as

$$\left[\begin{array}{cc:cc} \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & 1 \\ \hdashline \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{array}\right].$$

This device automatically places the non-zero elements upon the *second* over-diagonal (in general the rth); consequently a square root exists, namely

$$\left[\begin{array}{cc:cc} \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot \\ \hdashline \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot \end{array}\right].$$

This may be re-written as $\begin{bmatrix} e_1 & e_2 \\ \cdot & e_1 \end{bmatrix}$, where

$$e_1 = \begin{bmatrix} \cdot & 1 \\ \cdot & \cdot \end{bmatrix}, \qquad e_2 = \begin{bmatrix} \cdot & \cdot \\ 1 & \cdot \end{bmatrix} = e_1'.$$

Thus the square root of $\begin{bmatrix} \cdot & 1 \\ \cdot & \cdot \end{bmatrix}$ is $\begin{bmatrix} e_1 & e_2 \\ \cdot & e_1 \end{bmatrix}$, where the elements e_1, e_2 are no longer scalar. In this way the condition of equability may be evaded, without modifying any of the other conditions.

REFERENCES.

(1) Sylvester, *Phil. Mag.* (4) **5** (1853), 446; **6** (1853), 297; *Math. Papers*, **1**, 609, 641.

(2) E. T. Whittaker, *Proc. Roy. Soc. Edinburgh*, **36** (1915–16), 243–255.

(3) Turnbull and Aitken, *Canonical Matrices* (Glasgow, 1932).

On p. 75 the enunciation of the theorem needs a slight modification. If $f'(\lambda) \neq 0$ for each latent root λ of a matrix A, then the matrices A and $f(A)$ possess the same exponents, but the arrangement of these exponents within the Segre characteristic is not always the same, in the case when $f(\lambda)$ is a many-valued function of λ.

(4) D. E. Rutherford, *Proc. Edinburgh Math. Soc.* (2) **3** (1932), 135–143. Rutherford discusses the equation $\phi(X) = A$, where $\phi(x)$ is a polynomial in its argument x, and where the principle of equable partitions is implicitly used. Cf., also, S. Krishnamurthy Rao, "Invariant factors of a certain class of linear substitutions", *Jour. Indian Math. Soc.* **19** (1932), 233–240.

(5) Frobenius, *Berliner Sitzungsberichte* (1896), 7–16. Cf. H. F. Baker, *Proc. Cambridge Phil. Soc.* **23** (1925), 22–27, who gives several references, dealing more particularly with the case of symmetric matrices.