On the Canonical Form of a Rational Integral Function of a Matrix

By D. E. RUTHERFORD, St Andrews University.

(*Received* 14*th May*, 1932. *Read* 4*th June*, 1932.)

Introduction.

It is well known that the square matrix, of rank $n - k + 1$,

$$\begin{bmatrix} \cdot & \cdots & \cdot & b_{1,k} & b_{1,k+1} & \cdots & b_{1,n} \\ \cdot & \cdots & \cdot & \cdot & b_{2,k+1} & \cdots & b_{2,n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdot & \cdots & \cdots & \cdots & \cdots & \cdot & b_{n-k+1,n} \\ \cdot & \cdots & \cdots & \cdots & \cdots & \cdot & \cdot \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdot & \cdots & \cdots & \cdots & \cdots & \cdot & \cdot \end{bmatrix}$$

which we shall denote by B where any element to the left of, or below the nonzero diagonal $b_{1,k}, b_{2,k+1}, \ldots, b_{n-k+1,n}$ is zero, can be resolved into factors $Z^{-1}DZ$; where D is a square matrix of order n having the elements $d_{1,k}, d_{2,k+1}, \ldots, d_{n-k+1,n}$ all unity and all the other elements zero, and where Z is a non-singular matrix. In this paper we shall show in a particular case that this is so, and in the case in question we shall exhibit the matrix Z explicitly. Application of this is made to find the classical canonical form of a rational integral function of a square matrix A. When this has been found, it is easy to find the conditions for the existence of a solution of the matrix equation $\phi(X) = A$, where ϕ is a rational integral function of X, and then to give explicitly the canonical form of such solutions if they exist. In this last problem we shall follow the methods of R. Weitzenböck[1] who has recently discussed the matrix equation[2] $X^2 = A$. I have to thank Professor H. W. Turnbull for suggesting the problem and for discussing it with me.

§ 1. Let I_n be the unit matrix of order n, and let U_n be the auxiliary unit matrix of order n; that is to say, U_n is the square matrix of

---

[1] *Proc. Akad. Amsterdam*, 35 (1932), 157.

[2] References to the original investigation by Frobenius, and to others, are given by Turnbull and Aitken, *Canonical Matrices* (Glasgow, 1932), 81.

order n all of whose elements are zero save those on the over-diagonal which are unity. Thus

$$I_4 = \begin{bmatrix} 1 & . & . & . \\ . & 1 & . & . \\ . & . & 1 & . \\ . & . & . & 1 \end{bmatrix} \text{ and } U_4 = \begin{bmatrix} . & 1 & . & . \\ . & . & 1 & . \\ . & . & . & 1 \\ . & . & . & . \end{bmatrix}.$$

Now by the rules of matrix multiplication

$$U_4^2 = \begin{bmatrix} . & . & 1 & . \\ . & . & . & 1 \\ . & . & . & . \\ . & . & . & . \end{bmatrix}, \; U_4^3 = \begin{bmatrix} . & . & . & 1 \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \end{bmatrix} \text{ and } U_4^4 = \begin{bmatrix} . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \end{bmatrix}.$$

In general, if $m < n$, U_n^m is a matrix of order n, all of whose elements are zero with the exception of a diagonal of units beginning at the $(1, m+1)$th position. If $m \geqq n$, then U_n^m is the zero matrix of order n.

A matrix of the form $\lambda I_n + U_n$, where λ is a scalar, is called a *simple Jordan classical matrix*, or for brevity *a simple C-matrix*. We denote this matrix conveniently by $C_n(\lambda)$. A matrix which has square submatrices $M_1, M_2, \ldots, M_r$ on the leading matrix diagonal and which has zeros everywhere else, is usefully denoted by

$$\text{diag}\,(M_1, M_2, \ldots, M_r);$$

for example,

$$\text{diag}\,(U_3, 2I_2) = \left[\begin{array}{ccc:cc} . & 1 & & & \\ & . & 1 & & \\ & & . & & \\ \hdashline & & & 2 & \\ & & & & 2 \end{array}\right].$$

If $M_1, M_2, \ldots, M_r$ are all simple C-matrices, then the matrix $\text{diag}\,(M_1, M_2, \ldots, M_r)$, which we shall denote by M, is called a *compound C-matrix* or merely a *C-matrix*; and $M_1, M_2, \ldots, M_r$ are called the *latent submatrices* of the matrix M.

§ 2. *Theorem: If $a_k \neq 0$, then there will exist a square non-singular matrix Z, such that*

$$Z\,(a_k U^k + a_{k+1} U^{k+1} + \ldots + a_{n-1} U^{n-1}) = U_n^k Z. \tag{1}$$

Proof. If for convenience we write

$$0 = a_{k-1} = a_{k-2} = \ldots = a_{1-n}, \tag{2}$$

then equation (1) may be written as follows

$$\begin{bmatrix} z_{1,1} \ldots\ldots z_{1,n} \\ \ldots\ldots \\ \ldots\ldots \\ \ldots\ldots \\ z_{n,1} \ldots\ldots z_{n,n} \end{bmatrix} \begin{bmatrix} a_0 & a_1 & a_2 & \ldots\ldots a_{n-1} \\ a_{-1} & a_0 & a_1 & \ldots\ldots a_{n-2} \\ a_{-2} & a_{-1} & a_0 & \ldots\ldots a_{n-3} \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ a_{1-n} & a_{2-n} & a_{3-n} & \ldots\ldots a_0 \end{bmatrix} = \begin{bmatrix} \ldots\ldots 1 \ldots\ldots \\ \ldots\ldots\ldots 1 \ldots \\ \ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots 1 \\ \ldots\ldots\ldots\ldots \\ \vdots \qquad \vdots \\ \ldots\ldots\ldots\ldots \end{bmatrix} \begin{bmatrix} z_{1,1} \ldots\ldots z_{1,n} \\ \ldots\ldots \\ \ldots\ldots \\ \ldots\ldots \\ z_{n,1} \ldots\ldots z_{n,n} \end{bmatrix}.$$

Hence, the equations for the elements z_{ij} are

$$\begin{cases} a_{j-1} z_{i,1} + a_{j-2} z_{i,2} + \ldots + a_{j-n} z_{i,n} = z_{i+k,j}, \begin{bmatrix} i \leqq n-k \\ j = 1, \ldots, n \end{bmatrix}, & (3) \\ a_{j-1} z_{i,1} + a_{j-2} z_{i,2} + \ldots + a_{j-n} z_{i,n} = 0, \quad \begin{bmatrix} i > n-k \\ j = 1, \ldots, n \end{bmatrix}. & (4) \end{cases}$$

By examining equations (3), we see that if we give $z_{i,j}$ ($i \leqq k, j=1, \ldots, n$) any values whatever we can always find values for the other elements of Z such that equations (3) are satisfied. If, in addition, equation (4) is satisfied for all values of the elements of the matrix Z so obtained, then Z must satisfy equation (1). Let us choose therefore the following values of $z_{i,j}$. Let

$$\begin{aligned} z_{i,i} &= 1 && (i \leqq k), \\ z_{i,j} &= 0 && (i \leqq k, \ i \neq j). \end{aligned} \qquad (5)$$

We shall now show that equations (4) are satisfied identically by the values of $z_{i,j}$ given in (5). In virtue of the relations (2), (3), and (5)

$$z_{i,j} = 0 \text{ if } i > j \qquad (i = 1, \ldots, n); \qquad (6)$$

hence in view of (6), the equations (4) immediately reduce to

$$a_{j-i} z_{i,i} + \ldots + a_{j-n} z_{i,n} = 0, \quad \begin{bmatrix} i > n-k \\ j = 1, \ldots, n \end{bmatrix}.$$

In this last equation, the maximum value for

$$j - i \text{ is } n - n + k - 1 = k - 1,$$

that is to say

$$a_{j-i} = \ldots = a_{j-n} = 0,$$

https://doi.org/10.1017/S0013091500013900 Published online by Cambridge University Press

so that equations (4) are identically satisfied by the values given in (5). It has now been shown that a Z exists which satisfies equation (1). It remains to be shown that Z is non-singular. If we put $i + k = j$ in equation (3), we have, by equations (2) and (6),

$$z_{jj} = a_k \, z_{j-k,\, j-k}, \quad (j = k+1, \ldots, n); \tag{7}$$

consequently from (5), (6) and (7), the value of $|Z|$ is a power of a_k and so Z is non-singular. This concludes the proof of the theorem.

We now find it convenient to rewrite equation (1) as

$$Z\,(a_k\, U_n^k + a_{k+1}\, U_n^{k+1} + \ldots + a_{n-1}\, U_n^{n-1})\, Z^{-1} = U_n^k \tag{1a}$$

§3. We next wish to show that the form of the matrix Z can be given explicitly. The difference equation (3) can be written

$$z_{gk+p,\, j} = \sum_{q_1=1}^{n} a_{j-q_1}\, z_{(g-1)k+p,\, q_1} \qquad \begin{bmatrix} r \leqq k \\ m > 0 \end{bmatrix}.$$

By a repeated application of this formula we obtain

$$\begin{aligned} z_{gk+p,\, j} &= \sum_{q_1,\, q_2=1}^{n} a_{j-q_1}\, a_{q_1-q_2}\, z_{(g-2)k+p,\, q_2} \\ &= \ldots\ldots\ldots\ldots \\ &= \sum_{q_1, \ldots, q_g=1}^{n} a_{j-q_1}\, a_{q_1-q_2} \ldots a_{q_{g-1}-q_g}\, z_{p,\, q_g}\,; \end{aligned}$$

hence substituting the values of $z_{p,\, q_g}$ given in (5), we have

$$z_{gk+p,\, j} = \sum_{q_1, \ldots, q_{g-1}=1}^{n} a_{j-q_1}\, a_{q_1-q_2} \ldots a_{q_{g-1}-p}.$$

Thus $z_{gk+p,j}$ is a homogeneous function of degree g in the a's and the weight of each term is $j - p$, where we define the weight of any term as the sum of the suffixes of the a's. It only remains to find the numerical coefficient of a term such as $a_k^{s_k}\, a_{k+1}^{s_{k+1}} \ldots a_{n-1}^{s_{n-1}}$. A little consideration will show that the numerical coefficient is just the number of permutations of $s_k + s_{k+1} + \ldots + s_{n-1}$ things, s_k of which are alike of one kind, s_{k+1} of which are alike of a second kind, ..., and s_{n-1} of which are alike; so that the numerical coefficient is

$$\frac{(s_k + s_{k+1} + \ldots + s_{n-1})!}{s_k!\, s_{k+1}! \ldots s_{n-1}!}.$$

As an example we shall find the value of $z_{9,13}$ in the case where $k=2$. Now $z_{9,13}=z_{4.2+1,13}$; hence $z_{9,13}$ is a sum of products of the α's of degree 4 and of weight $13-1=12$. Since $\alpha_1=0$, $\alpha_2\neq 0$, the products are

$$\alpha_2^3\,\alpha_6,\quad \alpha_2^2\,\alpha_3\,\alpha_5,\quad \alpha_2^2\,\alpha_4^2,\quad \alpha_2\,\alpha_3^2\,\alpha_4,\quad \alpha_3^4.$$

Supplying in each case the appropriate numerical factor, we find that

$$z_{9,13}=4!\left(\frac{\alpha_2^3\,\alpha_6}{3!}+\frac{\alpha_2^2\,\alpha_3\,\alpha_5}{2!}+\frac{\alpha_2^2\,\alpha_4^2}{2!\,2!}+\frac{\alpha_2\,\alpha_3^2\,\alpha_4}{2!}+\frac{\alpha_3^4}{4!}\right).$$

As a second example, if $k=2$, $n=7$, then

$$Z=\begin{bmatrix} 1 & & & & & & \\ & 1 & & & & & \\ & & \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 & \alpha_6 \\ & & & \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 \\ & & & & \alpha_2^2 & 2\alpha_1\alpha_3 & 2\alpha_2\alpha_4+\alpha_3^2 \\ & & & & & \alpha_2^2 & 2\alpha_1\alpha_3 \\ & & & & & & \alpha_2^3 \end{bmatrix}.$$

§ 4. It is obvious from equation (1a) that Z satisfies the relation

$$Z\,(aI_n+\alpha_k\,U_n^k+\alpha_{k+1}\,U_n^{k+1}+\ldots+\alpha_{n-1}\,U_n^{n-1})\,Z^{-1}=aI_n+U_n^k. \tag{8}$$

Further[1] the canonical form of the matrix $aI_n+U_n^k$ is known to be

$$\operatorname{diag}\,(aI_{n_1}+U_{n_1},\ aI_{n_2}+U_{n_2},\ \ldots,\ aI_{n_k}+U_{n_k}),$$

where, if $n\equiv pk+q$, $q<k$, then $n_1=n_2=\ldots=n_q=p+1$ and $n_{q+1}=\ldots=n_k=p$. This result is obtained, in fact, merely through the interchange of suitable rows and columns in the matrix $aI_n+U_n{}^k$. Let us denote a compound C-matrix of this sort by $C_n\,(a)_k$. Thus, for example, the canonical form of $aI_5+U_5{}^3$ is $C_5\,(a)_3$ where

$$aI_5+U_5{}^3=\begin{bmatrix} a & & & 1 & \\ & a & & & 1 \\ & & a & & \\ & & & a & \\ & & & & a \end{bmatrix}\quad\text{and}\quad C_5(a)_3=\begin{bmatrix} a & 1 & & & \\ & a & & & \\ & & a & 1 & \\ & & & a & \\ & & & & a \end{bmatrix}$$

[1] See *Canonical Matrices*, 67.

https://doi.org/10.1017/S0013091500013900 Published online by Cambridge University Press

The interchanges required are indicated by the small dotted lines. The nonzero elements of a simple C-matrix in $C_5(\alpha)_3$ are connected together by a dotted line in $\alpha I_5 + U_5^3$. It follows from the above that the canonical form of $\alpha I_n + \alpha_k U_n^k + \ldots + \alpha_{n-1} U_n^{n-1}$ is $C_n(\alpha)_k$.

§ 5. If $KAK^{-1} = \Lambda$, where K is a non-singular matrix, then $K\phi(A)K^{-1} = \phi(\Lambda)$, where ϕ is a rational integral function of its argument. Let Λ be the canonical form of A, that is to say, Λ is a C-matrix which can be represented as

$$\text{diag}\,(\Lambda_1, \Lambda_2, \ldots, \Lambda_r)$$

where each sub-matrix Λ_h is a simple C-matrix of order t_h; it follows that $\phi(\Lambda)$ is the matrix

$$\text{diag}\,(\phi(\Lambda_1), \phi(\Lambda_2), \ldots, \phi(\Lambda_r)).$$

Suppose, then, that

$$\Lambda_h = \lambda_h I_{t_h} + U_{t_h};$$

therefore, on expanding by Taylor's Theorem, we have

$$\phi(\Lambda_h) = \phi(\lambda_h I_{t_h} + U_{t_h}) = \phi(\lambda_h) I_{t_h} + \frac{\phi'(\lambda_h)}{1!} U_{t_h} + \ldots + \frac{\phi^{(t_h-1)}(\lambda_h)}{(t_h-1)!} U_{t_h}^{t_h-1}.$$

Now, let $\phi^{(k_h)}(\lambda_h)$ be the first of the derivatives $\phi'(\lambda_h), \phi''(\lambda_h), \ldots$ which does not vanish: then, if we put

$$\alpha_c = \phi^{(c)}(\lambda_h) / c!$$

in equation (8), we see that there exists a non-singular matrix Z_h, such that

$$Z_h \cdot \phi(\Lambda_h) \cdot Z_h^{-1} = \phi(\lambda_h) I_{t_h} + U_{t_h}^{k_h}.$$

The right hand side can, in turn, be reduced to the canonical form $C_{t_h}(\phi(\lambda_h))_{k_h}$; there exists, then, a non-singular matrix T_h, such that

$$T_h \cdot \phi(\Lambda_h) \cdot T_h^{-1} = C_{t_h}(\phi(\lambda_h))_{k_h}.$$

It follows that

$$\text{diag}\,(T_1, T_2, \ldots, T_r) \cdot \phi(\Lambda) \cdot \text{diag}\,(T_1^{-1}, T_2^{-1}, \ldots, T_r^{-1})$$
$$= \text{diag}\,(C_{t_1}(\phi(\lambda_1))_{k_1}, \ldots, C_{t_r}(\phi(\lambda_r))_{k_r}).$$

Hence if $H = K \cdot \text{diag}\,(T_1, T_2, \ldots, T_r)$, then

$$H \cdot \phi(A) \cdot H^{-1} = \text{diag}\,(C_{t_1}(\phi(\lambda_1))_{k_1}, \ldots, C_{t_r}(\phi(\lambda_r))_{k_r});$$

further H is a non-singular matrix and hence we have found the canonical form of $\phi(A)$ where ϕ is a rational integral function of the matrix A.

§ 6. Let Λ be the canonical form of the matrix A. Then, as before, $A = K^{-1}\Lambda K$, where K is non-singular. Now Λ is a C-matrix. Suppose that

$$\Lambda = \text{diag}\,(\Lambda_1, \Lambda_2, \ldots, \Lambda_r)$$

where each sub-matrix Λ_h is a simple C-matrix of order t_h with latent root λ_h. All the latent matrices do not necessarily have different latent roots.

Suppose that $\Lambda_{h_1}, \Lambda_{h_2}, \ldots, \Lambda_{h_r}$ are the only latent matrices with latent root λ_h and consider the matrix

$$V_h = \text{diag}\,(\Lambda_{h_1}, \Lambda_{h_2}, \ldots, \Lambda_{h_r}).$$

We can write this alternatively as

$$V_h = \text{diag}\,(C_{t_{h_1}}(\lambda_h), \ldots, C_{t_{h_r}}(\lambda_h)).$$

It is frequently possible to group several of these simple C-matrices together in the following manner

$$V_h = \text{diag}\,(C_{\tau_1}(\lambda_h)_{\sigma_1}, \ldots, C_{\tau_\mu}(\lambda_h)_{\sigma_\mu}).$$

This can usually be accomplished in a number of ways. Thus, for example,

$$\begin{aligned} \text{diag}\,(C_4(\lambda), C_4(\lambda), C_3(\lambda), C_2(\lambda)) &= \text{diag}\,(C_8(\lambda)_2, C_5(\lambda)_2) \\ &= \text{diag}\,(C_8(\lambda)_2, C_3(\lambda), C_2(\lambda)) \\ &= \text{diag}\,(C_4(\lambda), C_4(\lambda), C_5(\lambda)_2) \\ &= \text{diag}\,(C_{11}(\lambda)_3, C_2(\lambda)) \\ &= \text{diag}\,(C_4(\lambda), C_7(\lambda)_2, C_2(\lambda)). \end{aligned}$$

It is thus possible, in general, by pursuing this method to arrange the whole matrix Λ in a number of different ways in the form

$$\Lambda = \text{diag}\,(N_1, \ldots, N_\rho)$$

where each N_h is of the form $C_{\xi_h}(\nu_h)_{\theta_h}$.

Now if there exist a θ_h-fold repeated root β_h of the equation

$$\phi(x) - \nu_h = 0, \tag{9}$$

https://doi.org/10.1017/S0013091500013900 Published online by Cambridge University Press

then

$$\phi(\beta_h) = \nu_h,\ \phi'(\beta_h) = 0, \dots,\ \phi^{(\theta_h - 1)}(\beta_h) = 0,\ \phi^{(\theta_h)}(\beta_h) \neq 0,$$

and the canonical form of $\phi(\beta_h I_{\xi_h} + U_{\xi_h})$ is $C_{\xi_h}(\nu_h)_{\theta_h}$. Hence, if, for any arrangement

$$\Lambda = \text{diag}\,(N_1, \dots, N_\rho), \tag{10}$$

there exist for every value of h, a θ_h-fold repeated root β_h of the equation (9), then a solution of the matrix equation $\phi(X) = A$ will exist. For let

$$Y = \text{diag}\,(C_{\xi_1}(\beta_1), \dots, C_{\xi_\rho}(\beta_\rho));$$

then Y is a solution of the equation

$$\phi(Y) = F\Lambda F^{-1},$$

where F is a non-singular matrix; hence $K^{-1} F^{-1} Y F K$ is a solution of $\phi(X) = A$.

Since several arrangements (10) are generally possible and since equation (9) may have several θ_h-fold repeated roots, we find in general that there are several solutions of the matrix equation

$$\phi(X) = A.$$

A little consideration will show that the above conditions are both necessary and sufficient.

§ 7. We shall conclude this paper by the solution of an example: to find the canonical form Y of a matrix X which satisfies the equation

$$\phi(X) = X^3 - X^2 - X - I = A,$$

where the canonical form Λ of A is given by

$$\Lambda = \begin{bmatrix} -2 & 1 & & & & & \\ & -2 & 1 & & & & \\ & & -2 & & & & \\ & & & -2 & 1 & & \\ & & & & -2 & & \\ & & & & & -1 & 1 \\ & & & & & & -1 \end{bmatrix}.$$

Here

$$\Lambda = \text{diag}\,(C_3(-2),\ C_2(-2),\ C_2(-1))$$

or

$$\Lambda = \text{diag}\,(C_5(-2)_2,\ C_2(-1)).$$

https://doi.org/10.1017/S0013091500013900 Published online by Cambridge University Press

Now $\phi(x) - (-2) = x^3 - x^2 - x + 1 = (x-1)^2(x+1)$, hence both $\phi(C_5(1))$ and $\phi(\text{diag}(C_3(-1),\ C_2(-1)))$ are equivalent to $C_5(-2)_2$. Further,

$$\phi(x) - (-1) = x^3 - x^2 - x = x\left(x - \frac{1+\sqrt{5}}{2}\right)\left(x - \frac{1-\sqrt{5}}{2}\right),$$

hence $\phi(C_2(0))$, $\phi\left(C_2\left(\frac{1+\sqrt{5}}{2}\right)\right)$, $\phi\left(C_2\left(\frac{1-\sqrt{5}}{2}\right)\right)$ are all equivalent to $C_2(-1)$. We thus obtain the following six values for Y

$$Y_1 = \text{diag}(C_5(1),\ C_2(0)),$$

$$Y_2 = \text{diag}\left(C_5(1),\ C_2\left(\frac{1+\sqrt{5}}{2}\right)\right),$$

$$Y_3 = \text{diag}\left(C_5(1),\ C_2\left(\frac{1-\sqrt{5}}{2}\right)\right),$$

$$Y_4 = \text{diag}(C_3(-1),\ C_2(-1),\ C_2(0)),$$

$$Y_5 = \text{diag}\left(C_3(-1),\ C_2(-1),\ C_2\left(\frac{1+\sqrt{5}}{2}\right)\right),$$

$$Y_6 = \text{diag}\left(C_3(-1),\ C_2(-1),\ C_2\left(\frac{1-\sqrt{5}}{2}\right)\right).$$