ON THE EQUATION $P(A, X)=0$ IN MATRICES*

BY

WILLIAM E. ROTH

In the present discussion we shall consider the solution of the equation

$$P(A, X) \equiv \sum_{k=0}^{p} F_k(A)X^{p-k} = 0, \tag{1}$$

where A is a known $n\times n$ matrix, $F_k(\lambda)$ $(k=0, 1, \cdots, p)$ are polynomials† in the scalar variable λ, and X is the unknown $n\times n$ matrix. The equation is a special case under that of an earlier paper by the author in which the coefficient matrices are not polynomials in a given matrix, but are known $m\times n$ matrices.‡ With the restrictions upon the coefficients which we now impose, it is possible to establish inequalities limiting the degree and the number of the elementary divisors of $X-\mu I$, where X is a solution of (1). These inequalities depend upon a knowledge of the elementary divisors of $P(A, \mu)$ and of $A-\lambda I$, where μ and λ are scalar variables. Certain theorems below, particularly Theorems III and IV with appropriate changes, are valid for the more general equation of the type studied by the author and others.§

Solutions of (1) are taken up under the following hypotheses: (a) that X be a unilateral solution on the right (or left) of the polynomials $F_k(\lambda)$ $(k=0, 1, \cdots, p-1)$; (b) that X be a bilateral solution; and (c) that X be commutative with A. By means of the idea of transversion of matrices as defined in §III, we show the fundamental relationship which exists between solutions on the right and those on the left of (1), and between these and the bilateral solutions if such exist.

---

* Presented to the Society, November 30, 1928, and September 9, 1931; received by the editors October 7, 1932.

† Considerations which follow do not require the functions $F_k(\lambda)$ $(k=0, 1, \cdots, p)$ to be polynomials. In fact any functions which, together with at most their first $n-1$ derivatives, may be expanded into series of non-negative powers of λ are permissible, provided that the characteristic values of A lie within or on the circles of convergence of each of the series representing $F_k(\lambda)$ $(k=0, 1, \cdots, p)$ and their first $n-1$ derivatives. For information on such functions of matrices the reader may consult Hensel, *Über Potenzenreihen von Matrizen*, Journal für die reine und angewandte Mathematik, vol. 155, pp. 107–110; Sheffer, *A note on matrix power series*, American Mathematical Monthly, vol. 36 (1930), pp. 228–231.

‡ Roth, *On the unilateral equation in matrices*, these Transactions, vol. 32 (1930), pp. 61–80. This paper cites several articles on algebraic equations in matrices.

§ Roth, loc. cit.

I. Preliminary notions and lemmas

Definition. *If* $A(\lambda) = (a_{ij}(\lambda))$ $(i=1, 2, \cdots, r; j=1, 2, \cdots, s)$, *where* $a_{ij}(\lambda)$ *are polynomials in* λ, *if*

$$a_{ij}(\lambda) \equiv r_{ij}(\lambda), \bmod (\lambda - a)^n \qquad (i = 1, 2, \cdots, r; j = 1, 2, \cdots, s),$$

and if $R(\lambda) = (r_{ij}(\lambda))$, *then*

$$A(\lambda) \equiv R(\lambda), \qquad \bmod (\lambda - a)^n.$$

Definition. *If* $A(\lambda)$ *is an* $m\times m$ λ-*matrix whose elements are polynomials in* λ, *if*

$$A(\lambda) \equiv R(\lambda), \qquad \bmod (\lambda - a)^n,$$

and if the ith elementary divisor of $R(\lambda)$ *corresponding to the linear factor* $\lambda - a$ *is* $(\lambda-a)^{\alpha^{(i)}}$ $(i=1, 2, \cdots, \rho)$, *where* ρ *is the rank of* $R(\lambda)$, *then* $(\lambda-a)^{\alpha^{(i)}}$ $(i=1, 2, \cdots, \rho)$ *is the ith elementary divisor of* $A(\lambda)$ *and* ρ *its rank with respect to the modulus* $(\lambda-a)^n$.

Definition. *If* $A(\lambda)$ *is an* $m\times m$ λ-*matrix whose elements are polynomials in* λ, *if the ith elementary divisor of* $A(\lambda)$ *corresponding to the linear factor* $\lambda-a$ *is* $(\lambda-a)^{\alpha^{(i)}}$ $(i=1, 2, \cdots, r)$ *where r is the rank of* $A(\lambda)$, *and if*

$$\alpha^{(i)} < n \qquad (i = 1, 2, \cdots, \sigma),$$
$$\alpha^{(i)} \geqq n \qquad (i = \sigma + 1, \sigma + 2, \cdots, r);$$

then σ *is the reduced rank and* $(\lambda-a)^{\alpha^{(i)}}$ $(i=1, 2, \cdots, \sigma)$ *is the ith elementary divisor of* $A(\lambda)$ *with respect to the modulus* $(\lambda-a)^n$.

Plainly if r is the rank $A(\lambda)$ and if ρ and σ are respectively the rank and the reduced rank of $A(\lambda)$ with respect to the modulus $(\lambda-a)^n$, then $\sigma\leqq\rho\leqq r\leqq m$; moreover it should be noted that in either case all minors of order $k\leqq\sigma$ are divisible by $\prod_{i=1}^{k} (\lambda-a)^{\alpha^{(i)}}$ and that this kth determinant divisor may consequently be congruent to zero modulo $(\lambda-a)^n$, while the elementary divisors with respect to this modulus are not. In speaking of the elementary divisors of $A(\lambda)$ with respect to the modulus $(\lambda-a)^n$, it is not necessary to designate the linear factor, for it is always that occurring in the modulus. As a matter of convenience we shall still call $(\lambda-a)^{\alpha^{(k)}}$ the kth elementary divisor of $A(\lambda)$ even when $\alpha^{(k)}=0$. Thus if $|A(\lambda)|$ is prime to $\lambda-a$, then each of the m elementary divisors of $A(\lambda)$ with respect to the modulus $(\lambda-a)^n$ is unity.

Lemma I. *If* $A(\lambda)$ *is an* $m\times m$ *matrix having elements* $a_{ij}(\lambda)$ $(i, j=1, 2, \cdots, m)$ *which are polynomials in* λ, *and if the reduced rank of* $A(\lambda)$ *with respect to the modulus* $(\lambda-a)^n$ *is* σ *and its ith elementary divisor is* $(\lambda-a)^{\alpha^{(i)}}$ $(i=1, 2, \cdots, \sigma)$, *then two* $m\times m$ *matrices,* $P(\lambda)$ *and* $Q(\lambda)$, *of degree* $n-1$ *in* λ, *exist such that $P(a)$ and $Q(a)$ are non-singular matrices whose elements do not depend upon n but do depend upon a, and that*

$$P(\lambda)A(\lambda)Q(\lambda) \equiv S(\lambda), \qquad \bmod (\lambda - a)^n,$$

where

$$S(\lambda) \equiv (s_{ij}(\lambda))$$

and

$$\begin{aligned} s_{ij}(\lambda) &\equiv 0, & &\bmod (\lambda - a)^n & &(i \neq j),\\ s_{ii}(\lambda) &\equiv (\lambda - a)^{\alpha^{(i)}}, & &\bmod (\lambda - a)^n & &(i = 1, 2, \cdots, \sigma),\\ s_{ii}(\lambda) &\equiv 0, & &\bmod (\lambda - a)^n & &(i = \sigma + 1, \sigma + 2, \cdots, m),\end{aligned}$$

and

$$\alpha^{(i)} \leqq \alpha^{(i+1)} \qquad (i = 1, 2, \cdots, \sigma - 1).$$

According to a well known theorem* two non-singular $m\times m$ matrices, $T(\lambda)$ and $U(\lambda)$, $|T(\lambda)|$ and $|U(\lambda)|$ independent of λ, exist such that

$$T(\lambda)A(\lambda)U(\lambda) = \prod_{i=1}^{t} A_i(\lambda), \tag{2}$$

where $\lambda - a_i$ $(i=1, 2, \cdots, t)$ are the distinct linear factors in λ common to all r-rowed minors of $A(\lambda)$, where r is the rank of $A(\lambda)$, and where $A_h(\lambda) = (a_{ij}^{(h)}(\lambda))$ $(h=1, 2, \cdots, t)$ are such that

$$\begin{aligned} a_{ij}^{(h)}(\lambda) &= 0 & &(i \neq j),\\ a_{ii}^{(h)}(\lambda) &= (\lambda - a_h)^{\alpha_h^{(i)}} & &(i \leqq r),\\ a_{ii}^{(h)}(\lambda) &= 0 & &(i > r).\end{aligned}$$

Hence the kth composite elementary divisor of $A(\lambda)$, as usually defined, is

$$\prod_{h=1}^{t}(\lambda - a_h)^{\alpha_h^{(k)}} \qquad (k = 1, 2, \cdots, r).$$

Since $A_i(\lambda)$ $(i=1, 2, \cdots, t)$ are diagonal matrices, they are commutative one with another and (2) can be written in the form

$$T(\lambda)A(\lambda)U(\lambda) = A_k(\lambda)\left[\prod_{i=1}^{k-1} A_i(\lambda)\prod_{i=k+1}^{t} A_i(\lambda)\right]. \tag{3}$$

Now if the last $n-r$ zero elements in the principal diagonal of

$$A_i(\lambda) \qquad (i = 1, 2, \cdots, k-1, k+1, k+2, \cdots, t)$$

be replaced by unit elements, the resulting matrix, $A_i'(\lambda)$ $(i=1, 2, \cdots,$

* Bôcher, *Introduction to Higher Algebra*, 1922, p. 91, and Theorem I, p. 94.

$k-1$, $k+1$, $k+2, \cdots, t$) may replace the corresponding matrix $A_i(\lambda)$ in the right member of (3), and this substitution will not affect the validity of this equation in that the product by $A_k(\lambda)$ leaves this member unchanged. The jth element $j \leqq r$ of the diagonal matrix

$$\prod_{i=1}^{k-1} A_i'(\lambda)\prod_{i=k+1}^{t} A_i'(\lambda) \text{ is } \prod_{i=1}^{k-1}(\lambda - a_i)^{\alpha_i^{(j)}}\prod_{i=k+1}^{t}(\lambda - a_i)^{\alpha_i^{(j)}}$$

and is a polynomial of degree $(\sum_{i=1}^{t} \alpha_i^{(j)}) - \alpha_k^{(j)}$ in $\lambda - a_k$ whose constant term is not zero. Hence the polynomial $v_{jk}(\lambda)$ exists such that

$$\prod_{i=1}^{k-1}(\lambda - a_i)^{\alpha_i^{(j)}}\prod_{i=k+1}^{t}(\lambda - a_i)^{\alpha_i^{(j)}} v_{jk}(\lambda) = 1 + (\lambda - a_k)^{n_k} w_{jk}(\lambda),\ j \leqq r,$$

where $w_{jk}(\lambda)$ is a polynomial in λ. The remaining $n-r$ elements of $\prod_{i=1}^{k-1} A_i'(\lambda) \prod_{i=k+1}^{t} A_i'(\lambda)$ are unit elements, hence we may let $v_{jk}(\lambda) \equiv 1$ for $r < j \leqq n$. Hence the diagonal matrix $V_k(\lambda)$, having as its jth element $v_{jk}(\lambda)$ as defined above, exists such that

$$\left[\prod_{i=1}^{k-1} A_i'(\lambda)\prod_{i=k+1}^{t} A_i'(\lambda)\right] V_k(\lambda) = I + (\lambda - a_k)^{n_k} W_k(\lambda),$$

where $|V_k(a_k)| \neq 0$ and $V_k(a_k)$ is again independent of n_k. Hence (3) becomes

$$T(\lambda)A(\lambda)U(\lambda)V_k(\lambda) = A_k(\lambda)\left[I + (\lambda - a_k)^{n_k} W_k(\lambda)\right],$$

where $U(a_k)V_k(a_k)$ is a non-singular matrix independent of n_k. Now if we let $a_k = a$ and let

$$\begin{aligned} T(\lambda) &\equiv P(\lambda), &\quad \bmod (\lambda - a)^n, \\ U(\lambda)V_k(\lambda) &\equiv Q(\lambda), &\quad \bmod (\lambda - a)^n, \\ A_k(\lambda) &\equiv S(\lambda), &\quad \bmod (\lambda - a)^n, \end{aligned}$$

we have demonstrated that $P(\lambda)$ and $Q(\lambda)$ satisfying the lemma exist.

LEMMA II. *If* $\Delta(\lambda) = (\delta_{ij}(\lambda))$ $(i, j = 1, 2, \cdots, n)$ *where*

$$\begin{aligned} \delta_{i,i+k}(\lambda) &\equiv (\lambda - a)^{\alpha_k} d_k(\lambda) &\quad (k &= 0, 1, \cdots, n-1), \\ \delta_{i,i-h}(\lambda) &\equiv 0 &\quad (h &= 1, 2, \cdots, n-1), \end{aligned}$$

where $d_k(\lambda)$ $(k=0, 1, \cdots, n-1)$ *are polynomials in* λ *and where* $d_0(a) \neq 0$ *and* $d_1(a) \neq 0$, *then the degree of the nth elementary divisor of* $\Delta(\lambda)$ *corresponding to the linear factor* $\lambda - a$

(1) *does not exceed* $n\alpha_0 - n + 1$ *if* $\alpha_0 > 1$ *and* $\alpha_1 > 0$,

(2) *does not exceed* $(n+1)/2$ *if* $\alpha_0 = 1$ *and* $\alpha_1 > 0$,

(3) *is equal to* $n\alpha_0$ *if* $\alpha_1 = 0$.

The proof of this lemma consists in seeking a lower bound to the degree of $\lambda - a$ as a divisor of all $(n-1)$st-order minors of $\Delta(\lambda)$. The determinant

$|\Delta(\lambda)|$ has the divisor $(\lambda-a)^{n\alpha_0}$, and none of higher degree in $\lambda-a$ if the lemma is satisfied; hence the difference between the lower bound so sought and $n\alpha_0$ is an upper bound for the degree of the nth elementary divisor of $\Delta(\lambda)$.

The minor $\delta_{ii}(\lambda)$ has the factor $(\lambda-a)^{(n-1)\alpha_0}$, in $\lambda-a$; that of $\delta_{i,i+k}(\lambda)$ $(k=1, 2, \cdots, n-1)$ is identically zero, whereas the minor of $\delta_{i,i-h}(\lambda)$ $(h=1, 2, \cdots, n-1)$ is $[(\lambda-a)^{\alpha_0}d_0(\lambda)]^{n-h-1} D_h(\lambda)$, where $D_h(\lambda)$ is the minor of order h obtained by dropping the first column and last $n-h-1$ columns and the last $n-h$ rows of $\Delta(\lambda)$.

Now

$$
\begin{aligned}
D_h(\lambda) &= (\lambda - a)^{\alpha_1}d_1(\lambda)D_{h-1}(\lambda) - (\lambda - a)^{\alpha_0+\alpha_2}d_0(\lambda)d_2(\lambda) + \cdots \\
&\quad \pm (\lambda - a)^{(h-1)\alpha_0+\alpha_h}d_0^{\,h-1}(\lambda)d_h(\lambda) \qquad (h = 1, 2, \cdots, n - 1),
\end{aligned} \tag{4}
$$

and $D_0(\lambda)=1$. If $\alpha_0>1$ and $\alpha_1>0$, we can show by mathematical induction on the basis of the recurrence relation (4), that $D_h(\lambda)$ has at least the factor $(\lambda-a)^h$, hence the minor of any element $\delta_{i,i-h}(\lambda)$ $(h=1, 2, \cdots, n-1)$ has at least the divisor $(\lambda-a)^{(n-h-1)\alpha_0+h}$ and that of lowest degree among them occurs for $h=n-1$; hence all $(n-1)$st-order minors of $\Delta(\lambda)$ have at least the factor $(\lambda-a)^{n-1}$ in common and for this case the degree of the nth elementary divisor of $\Delta(\lambda)$ corresponding to the linear factor $\lambda-a$ is at most $n\alpha_0-n+1$.

If $\alpha_0=1$ and $\alpha_1>0$, $D_1(\lambda)$ and $D_2(\lambda)$ have at least the factor $\lambda-a$. Then from (4) it readily follows that $(\lambda-a)^{h/2}$ or $(\lambda-a)^{(h+1)/2}$ are factors of $D_h(\lambda)$ according as h is an even or an odd integer. Hence we can infer that the minor of $\delta_{i,i-h}(\lambda)$ $(h=1, 2, \cdots, n-1)$ is divisible by $(\lambda-a)^{n-h-1+h/2}$ or by $(\lambda-a)^{n-h-1+(h+1)/2}$ according as h is even or odd. The divisor of lowest degree occurs for $h=n-1$, and is $(\lambda-a)^{n/2}$ or $(\lambda-a)^{(n-1)/2}$ according as n is an even integer or an odd integer. That is, if $\alpha_0=1$ and $\alpha_1>0$ the degree of the nth elementary divisor of $\Delta(\lambda)$ does not exceed $n/2$ or $(n+1)/2$ according as n is an even or an odd integer.

The third part of the lemma is evident, for the minor of $\delta_{n,1}(\lambda)$ is prime to $\lambda-a$, since all terms of its expansion save $d_1^{n-1}(\lambda)$ have this factor if $\alpha_1=0$. Hence in this case the nth elementary divisor of $\Delta(\lambda)$ corresponding to $\lambda-a$ is $(\lambda-a)^{n\alpha_0}$.

II. The unilateral solution

Let the normal form of A be given by $\bar{A}=(A_{ij})$, where

$$
\begin{aligned}
A_{ij} &= 0 &\qquad (i \neq j),\\
A_{ii} &= A_i &\qquad (i = 1, 2, \cdots, r),
\end{aligned}
$$

and A_i $(i=1, 2, \cdots, r)$ is an $m_i\times m_i$ matrix, $\sum_{i=1}^{r} m_i=n$, the elements of whose principal diagonal are a_i and those in the diagonal directly above are

m_i-1 unit elements and the remaining $(m_i-1)^2$ elements of A_i are zeros. Hence $A-\lambda I$ has the simple elementary divisors $(\lambda-a_i)^{m_i}$ $(i=1, 2, \cdots, r)$ and the non-singular matrix Q exists such that

$$A = Q\overline{A}Q^{-1}. \tag{5}$$

Moreover let (1) have the solution X on the right whose normal form is $\overline{X}=(X_{ij})$ $(i, j=1, 2, \cdots, s)$, where

$$X_{ij} = 0 \qquad (i \neq j),$$
$$X_{jj} = X_j = x_jI_j + D_j \qquad (j = 1, 2, \cdots, s),$$

where I_j is the $n_j\times n_j$ unit matrix, x_j is a scalar constant and D_j is the $n_j\times n_j$ matrix, whose elements are all zeros save those in the kth row and $(k+1)$st column $(k=1, 2, \cdots, n_j-1)$ which are unities. Thus we may write $D_j^0=I_j$ and $D_j^k=0$, $k\geqq n_j$. The matrix $X-\mu I$ has the elementary divisors $(\mu-x_j)^{n_j}$ $(j=1, 2, \cdots, s)$ and the non-singular matrix R exists such that

$$X = R\overline{X}R^{-1}. \tag{6}$$

On substituting for A and X in (1) by means of (5) and (6), and noting that Q and R are non-singular matrices, we obtain

$$\sum_{k=0}^{p} F_k(\overline{A})T\overline{X}^{p-k} = 0, \tag{7}$$

where $T=Q^{-1}R$. In this equation $\overline{X}$ and R (hence T) are the unknowns. In fact $T\overline{X}T^{-1}$ is a solution of $P(\overline{A}, X)=0$; on the other hand if X is a solution of $P(\overline{A}, X)=0$, then $Q^{-1}XQ$ is a solution of (1).

Let $T=(T_{ij})$, where T_{ij} $(i=1, 2, \cdots, r; j=1, 2, \cdots, s)$ is an $m_i\times n_j$ matrix; then from (7) we readily obtain the rs equations

$$\sum_{k=0}^{p} F_k(A_i)T_{ij}X_j^{p-k} = 0 \quad (i = 1, 2, \cdots, r; j = 1, 2, \cdots, s), \tag{8}$$

which must be satisfied by the rs independent matrices T_{ij}. Each of these equations provides a means of computing the corresponding T_{ij}, and consequently T, provided the matrices $X_j (j=1, 2, \cdots, s)$ were known. We shall seek restrictions upon x_j of X_j and upon its order n_j.

Now from $X_j=x_jI_j+D_j$ we have

$$X_j^h = \sum_{k=0}^{h} \binom{h}{k} x_j^{h-k} D_j^k,$$

and consequently (8), for A_i and X_j, becomes

$$\sum_{k=0}^{h} \frac{P_{0k}(A_i, x_j)}{k!} T_{ij}D_j^k = 0, \tag{9}$$

where

$$P_{hk}(\lambda, \mu) = \frac{\partial^{h+k}}{\partial\lambda^h\partial\mu^k}P(\lambda, \mu).$$

This equation must be satisfied by the sub-matrix T_{ij} of T, in order that (1) have a solution whose characteristic matrix, $X-\mu I$, has the elementary divisor $(\mu-x_j)^{n_j}$, where $(\lambda-a_i)^{m_i}$ is an elementary divisor of $A-\lambda I$.

Indicate the $m_i\times 1$ matrix formed by the $(k+1)$st column of T_{ij} by the *script* letter $\mathcal{T}_{ij}^{(k)}$; then

$$T_{ij} = (\mathcal{T}_{ij}^{(0)}, \mathcal{T}_{ij}^{(1)}, \cdots, \mathcal{T}_{ij}^{(n_j-1)})$$

and

$$T_{ij}D_j^k = (0, \cdots, 0, \mathcal{T}_{ij}^{(0)}, \mathcal{T}_{ij}^{(1)}, \cdots, \mathcal{T}_{ij}^{(n_j-k-1)});$$

that is, the multiplication of T_{ij} on the right by D_j^k moves the first n_j-k columns of T_{ij} k spaces to the right and replaces the evacuated spaces by k zero columns. Hence from (9) we readily obtain the equations

$$\sum_{h=0}^{k}\frac{P_{0h}(A_i, x_j)}{h!}\mathcal{T}_{ij}^{(k-h)} = 0 \qquad (k = 0, 1, \cdots, n_j - 1).$$

Multiply these for $k=0, 1, \cdots, n_j-1$ respectively by $1, \mu-x_j, \cdots, (\mu-x_j)^{n_j-1}$ and add the results; the single equation

$$P(A_i, \mu)\mathcal{T}_{ij}(\mu) \equiv 0, \qquad \text{mod } (\mu - x_j)^{n_j}, \tag{10}$$

is thus obtained, where

$$\mathcal{T}_{ij}(\mu) = \sum_{h=0}^{n_j-1}(\mu - x_j)^h\mathcal{T}_{ij}^{(h)} = T_{ij}\begin{pmatrix} 1 \\ \mu - x_j \\ \vdots \\ (\mu - x_j)^{n_j-1}\end{pmatrix} \tag{11}$$

is consequently an $m_i\times 1$ matrix whose elements are polynomials of degree n_j-1 in $\mu-x_j$. We shall henceforth concentrate upon equation (10) instead of (8).

From (10) and (11), we see that $\mathcal{T}_{ij}(x_j)=\mathcal{T}_{ij}^{(0)}=0$, if $|P(A_i, x_j)|\neq 0$; hence also $\mathcal{T}_{ij}^{(1)}=0$ and so on, under the same hypothesis. That is, $T_{ij}=0$, if $|P(A_i, x_j)|\neq 0$. Now not all T_{ij} $(i=1, 2, \cdots, r)$ can be zero, nor can all T_{ij} $(j=1, 2, \cdots, s)$ be zero else T would have n_j zero columns or m_i zero rows and in either case would be a singular matrix. Hence $|P(A_i, x_j)|=0$ for at least one pair of values of i and j; the necessary and sufficient condition that such be the case is that $P(a_i, x_j) \doteq 0$. We have proved the theorem.

THEOREM I. *If the characteristic matrix $A-\lambda I$, of A, have the elementary divisors $(\lambda-a_i)^{m_i}$ $(i=1, 2, \cdots, r)$, where $\sum_{i=1}^{r} m_i=n$, and if $P(A, X)=0$ have a solution, X, on the right (or left) whose characteristic matrix, $X-\mu I$, has the elementary divisors $(\mu-x_j)^{n_j}$, where $\sum_{j=1}^{s} n_j=n$, then every equation $P(a_i, \mu)=0$ $(i=1, 2, \cdots, r)$ must be satisfied by at least one of the numbers x_j $(j=1, 2, \cdots, s)$ and every equation $P(\lambda, x_j)=0$ $(j=1, 2, \cdots, s)$ must be satisfied by at least one of the numbers a_i $(i=1, 2, \cdots, r)$.*†

The above theorem shows where and how the characteristic values x_j of a solution of (1) must be sought and consequently gives us some knowledge of the sub-matrices X_j $(j=1, 2, \cdots, s)$. For more definite information regarding them, we shall seek restrictions upon n_j, in addition to that we already know, namely that $\sum_{j=1}^{s} n_j=n$ in order that the non-singular matrix $T=(T_{ij})$ may exist. Such is given by the following theorems.

THEOREM II. *If X is a solution of the polynomial equation $P(A, X)=0$, and if $X-\mu I$ has the elementary divisors $(\mu-x)^{\nu_1}, (\mu-x)^{\nu_2}, \cdots, (\mu-x)^{\nu_k}$ corresponding to the linear factor $\mu-x$, then $(\mu-x)^{\nu_1+\nu_2+\cdots+\nu_k}$ is a factor of $|P(A, \mu)|$; moreover if $X-\mu I$ has the elementary divisors $(\mu-x_j)^{n_j}$ $(j=1, 2, \cdots, s)$, then $\prod_{j=1}^{s} \{P(\lambda, x_j)\}^{n_j}$ must be an exact multiple of $|A-\lambda I|$.*

It is known that if $P(A, X)=0$, then $|P(A, \mu)|$ is divisible by $|X-\mu I|$‡, and this determinant in turn is the product of all its elementary divisors; hence the first part of the theorem is proved. Similarly if X is a solution of (1), then A satisfies the same equation, where we regard X as the known matrix, and consequently $|P(\lambda, X)|$ must be divisible by $|A-\lambda I|$. But we can readily show that

$$|P(\lambda, X)| = \prod_{j=1}^{s} |P(\lambda, X_j)| = \prod_{j=1}^{s} \{P(\lambda, x_j)\}^{n_j}.$$

Hence the second part of the theorem is proved.

The restrictions placed upon n_j by this theorem are not very severe; nevertheless the first part of the theorem places an upper bound upon n_j and the second part places a lower bound upon n_j $(j=1, 2, \cdots, s)$. The following results are far more restrictive and quite as easily applied in particular examples as are the above.

THEOREM III. *If $P(A, \mu)$ is of rank ρ_j with respect to the modulus $(\mu-x_j)^{n_j}$ and if the ρ_jth elementary divisor of $P(A, \mu)$ with respect to the same modulus*

† This theorem is in part a special case of one proved elsewhere, Roth, loc. cit., Theorem I, p. 65.
‡ Roth, loc. cit., Corollary I, p. 66.

is $(\mu-x_j)^{\alpha_j^{(\rho_j)}}$, *then the equation* (1) *may have the solution X whose characteristic matrix,* $X-\mu I$, *has the elementary divisor* $(\mu-x_j)^{n_j}$ *only if*

$$\alpha_j^{(\rho_j)} \geqq n_j + \rho_j - n;$$

and if $\alpha_j^{(\rho_j)} < n_j$, *that is, if the reduced rank and the rank of* $P(A,\mu)$ *with respect to the modulus* $(\mu-x_j)^{n_j}$ *are the same, then* $X-\mu I$ *has the elementary divisor* $(\mu-x_j)^{n_j}$ *at most k times only if*

$$\alpha_j^{(\rho_j)} \geqq n_j - \frac{n-\rho_j}{k}.$$

If

$$\frac{P_{h,0}(a_i,\mu)}{h!} \equiv 0, \bmod (\mu - x_j)^{n_j} \qquad (h = 0, 1, \cdots, \sigma_{ij} - 1),$$

$$\frac{P_{h,0}(a_i,\mu)}{h!} \equiv p_{ij}^{(h)}(\mu), \bmod (\mu - x_j)^{n_j} \quad (h = \sigma_{ij}, \sigma_{ij} + 1, \cdots, m_i - 1);$$

then $P(A_i, \mu)$ is of rank $\rho_{ij}=m_i-\sigma_{ij}$ with respect to the modulus $(\mu-x_j)^{n_j}$ and equation (10) reduces to the following non-homogeneous system of ρ_{ij} equations in the ρ_{ij} unknowns $t_{ij}^{(\sigma+h)}(\mu)$ $(h=0, 1, \cdots, \rho_{ij}-1)$†:

$$\begin{pmatrix} p^{(\sigma)}(\mu) & p^{(\sigma+1)}(\mu) & \cdots & p^{(m_i-1)}(\mu) \\ 0 & p^{(\sigma)}(\mu) & \cdots & p^{(m_i-2)}(\mu) \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & p^{(\sigma)}(\mu) \end{pmatrix} \begin{pmatrix} t^{(\sigma)}(\mu) \\ t^{(\sigma+1)}(\mu) \\ \vdots \\ t^{(m_i-1)}(\mu) \end{pmatrix} = (\mu - x_j)^{n_j} \begin{pmatrix} q^{(\sigma)}(\mu) \\ q^{(\sigma+1)}(\mu) \\ \vdots \\ q^{(m_i-1)}(\mu) \end{pmatrix}, \tag{12}$$

where $q^{(k)}(\mu)$ $(k=\sigma, \sigma+1, \cdots, m_i-1)$ are arbitrary polynomials in μ. The equation (10) imposes no restrictions upon $t^{(0)}(\mu), t^{(1)}(\mu), \cdots, t^{(\sigma-1)}(\mu)$, hence each of the first σ_{ij} rows of T_{ij} has n_j arbitrary elements and its remaining rows must be such that (11) and (12) are satisfied. From (12) we see at once that

$$t^{(\sigma+h)}(\mu) = \frac{(\mu - x_j)^{n_j} M_h(\mu)}{[p^{(\sigma)}(\mu)]^{m_i-\sigma-h}} \quad (h = 0, 1, \cdots, \rho_{ij} - 1),$$

† In this equation and in the remainder of the proof of this lemma we suppress the subscripts i and j of $p_{ij}^{(k)}(\mu)$ of σ_{ij} and of $t_{ij}^{(\sigma+h)}(\mu)$ save where ambiguity may arise.

where $M_h(\mu)$ is a linear combination of minors of order $\rho_{ij}-h-1$ of $P(A_i, \mu)$ and consequently has the $(\rho_{ij}-h-1)$th determinant divisor in $\mu-x_j$ of this matrix as its divisor. We shall now seek a lower bound to the degree of $\mu-x_j$ as a divisor of $t^{(k)}(\mu)$ $(k=\sigma, \sigma+1, \cdots, m_i-1)$.

Let $p^{(\sigma)}(\mu)$ have the factor $(\mu-x_j)^{\epsilon_{ij}}$ and $t^{(k)}(\mu)$ $(k=\sigma, \sigma+1, \cdots, m_i-1)$ have the factor $(\mu-x_j)^{\tau_{ij}^{(k)}}$ and let neither have a divisor of higher degree than these in $\mu-x_j$. Moreover, let the kth elementary divisor of $P(A_i, \mu)$ be $(\mu-x_j)^{\alpha_{ij}^{(k)}}$ $(k=1, 2, \cdots, \rho_{ij})$. Then the determinant divisor of all minors of order g of $P(A_i, \mu)$ is $\prod_{k=1}^{g}(\mu-x_j)^{\alpha_{ij}^{(k)}}$, $g \leqq \rho_{ij}$, and

$$(m_i-\sigma_{ij})\epsilon_{ij} = \rho_{ij}\epsilon_{ij} = \sum_{k=1}^{\rho_{ij}} \alpha_{ij}^{(k)}.$$

$M_h(\mu)$ has the factor $\mu-x_j$ at least $\sum_{k=1}^{\rho_{ij}-h-1}\alpha_{ij}^{(k)}$ times, and $\tau_{ij}^{(\sigma+h)}$ must satisfy the inequality

$$\tau_{ij}^{(\sigma+h)} \geqq \eta_j - \sum_{k=1}^{\rho_{ij}} \alpha_{ij}{}^{k} + h\epsilon_{ij} + \sum_{k=1}^{\rho_{ij}-h-1} \alpha_{ij}{}^{k},$$

or

$$\tau_{ij}^{(\sigma+h)} \geqq n_j + h\epsilon_{ij} - \sum_{k=\rho_{ij}-h}^{\rho_{ij}} \alpha_{ij}^{(k)} \qquad (h=0, 1, \cdots, \rho_{ij}-1). \tag{13}$$

This inequality evidently establishes a lower bound for the number of zero elements in the $(\sigma+h)$th row of T_{ij}, for if $\tau_{ij}^{(\sigma+h)}=k$ then the elements in the first k columns and the $(\sigma+h)$th row of T_{ij} are zero. The least value the right member of the inequality may have for any h occurs for $h=0$, that is, $\tau_{ij}^{(\sigma+h)} \geqq n_j-\alpha_{ij}^{(\rho_{ij})}$ $(h=0, 1, \cdots, \rho_{ij}-1)$. Therefore T_{ij} has only zero elements in at least the first $n_j-\alpha_{ij}^{(\rho_{ij})}$ columns of the last ρ_{ij} rows, whereas the first $m_i-\rho_{ij}=\sigma_{ij}$ rows have arbitrary elements as was pointed out above. Since

$$P(\overline{A}, \mu) = \begin{Vmatrix} P(A_1, \mu) & 0 & \cdots & 0 \\ 0 & P(A_2, \mu) & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & P(A_r, \mu) \end{Vmatrix},$$

the rank of $P(\overline{A}, \mu)$, hence of $P(A, \mu)$, with respect to the modulus $(\mu-x_j)^{n_j}$ is ρ_j, where $\rho_j=\sum_{i=1}^{r}\rho_{ij}$ and ρ_{ij} is the rank of $P(A_i, \mu)$ with respect to the same modulus. Moreover, if the ρ_j numbers $\alpha_{ij}^{(k)}$ $(k=1, 2, \cdots, \rho_{ij}; i=1, 2, \cdots, r)$ be rearranged in an ascending sequence

$$\alpha_j^{(1)} \leqq \alpha_j^{(2)} \leqq \cdots \leqq \alpha_j^{(\rho_j)},$$

then $(\mu-x_j)^{\alpha_j^{(k)}}$ is the kth elementary divisor of $P(\overline{A}, \mu)$ and of $P(A, \mu)$ with respect to the modulus $(\mu - x_j)^{n_j}$ and $\alpha_j^{\rho_j}$ is the greatest of the numbers $\alpha_{ij}^{(\rho_{ij})}$ $(i=1, 2, \cdots, r)$.

According to (10)

$$P(A, \mu)\mathfrak{T}_j(\mu) \equiv 0, \bmod (\mu - x_j)^{n_j}, \tag{14}$$

where

$$\mathfrak{T}_j(\mu) = \begin{bmatrix} \mathfrak{T}_{1j}(\mu) \\ \mathfrak{T}_{2j}(\mu) \\ \vdots \\ \mathfrak{T}_{rj}(\mu) \end{bmatrix} = \begin{bmatrix} T_{1j} \\ T_{2j} \\ \vdots \\ T_{rj} \end{bmatrix} \begin{bmatrix} 1 \\ \mu - x_j \\ \vdots \\ (\mu - x_j)^{n_j - 1} \end{bmatrix} = T_j \begin{bmatrix} 1 \\ \mu - x_j \\ \vdots \\ (\mu - x_j)^{n_j - 1} \end{bmatrix}.$$

The matrix T_j has $\sum_{i=1}^{r} (m_i - \rho_{ij}) = n - \rho_j$ rows of arbitrary elements and the remaining rows have only zero elements in the first $n_j - \alpha_j^{(\rho_j)}$ columns. Hence $\alpha_j^{(\rho_j)}$ must equal or exceed $n_j - n + \rho_j$ else T_j is of rank less than n_j and T would be singular. This proves the first part of the theorem. Now if $n_j > \alpha_j^{(\rho_j)}$ the reduced rank of $P(A, \mu)$ is ρ_j and if $X - \mu I$ have the elementary divisor $(\mu - x_j)^{n_j}$ repeated k times then T has k matrices T_j all having the same $n - \rho_j$ rows of arbitrary elements. Each T_j has at least $n_j - \alpha_j^{(\rho_j)}$ zero columns in the remaining rows. Hence $k(n_j - \alpha_j^{(\rho_j)})$ cannot exceed $n - \rho_j$ else the corresponding kn_j columns of T are of rank less than kn_j and T would be singular. This proves the final part of the theorem.

The second part of the theorem may be stated as follows:

COROLLARY I. *If the rank ρ_j of $P(A, \mu)$ with respect to the modulus $(\mu - x_j)^{n_j}$ is equal to the reduced rank with respect to the same modulus and if $(\mu - x_j)^{\alpha_j^{(\rho_j)}}$ is its ρ_jth elementary divisor, then the characteristic matrix, $X - \mu I$, of a solution of* (1) *cannot have the elementary divisor $(\mu - x_j)^{n_j}$ more than $(n - \rho_j)/(n_j - \alpha_j^{(\rho_j)})$ times.*

Plainly if n_j be taken sufficiently large the rank of $P(A, \mu)$ with respect to the modulus $(\mu - x_j)^{n_j}$ is equal to the rank, r, of $P(A, \mu)$ in the usual sense; that is, all minors of order $r+1$ and above are identically zero whereas those of order r are not all identically zero and in this case by Theorem III we have $n_j \leqq \alpha_j^{(r)} - n + r$, where $(\mu - x_j)^{\alpha_j^{(r)}}$ is the rth elementary divisor of $P(A, \mu)$ corresponding to the linear factor $\mu - x_j$. Hence:

COROLLARY II. *If $P(A, \mu)$ is of rank r and if the rth elementary divisor of $P(A, \mu)$ corresponding to the linear factor $\mu - x_j$ is $(\mu - x_j)^{\alpha_j^{(r)}}$, then no solution X of $P(A, X) = 0$ exists whose characteristic matrix, $X - \mu I$, has an elementary divisor corresponding to the linear factor $\mu - x_j$ whose degree exceeds*

$$\alpha_j^{(r)} - n + r.$$

The following corollary is at once evident from the foregoing.

COROLLARY III. *If $P(A, \mu)$ is of rank $r<n$, then $X-\mu I$, where X is a solution of* (1), *may have the elementary divisor $(\mu-x)^k$, where x is an arbitrary parameter, only if $k\leqq n-r$.*

In this case, where x is arbitrary, the rth elementary divisor of $P(A, \mu)$ corresponding to $\mu-x$ is unity, and $\alpha^{(r)}=0$. A more complete discussion of this case is given in the paper cited above†, where the method of computing the matrix corresponding to T is covered in some detail.

THEOREM IV. *If $P(A, \mu)$ has the reduced rank ρ with respect to the modulus $(\mu-x)^\nu$, and if $P(A, X)=0$, then the number of elementary divisors of $X-\mu I$ corresponding to the same linear factor $\mu-x$ and whose degree equals or exceeds ν cannot exceed $n-\rho$.*

If $P(A, \mu)$ has the elementary divisors $(\mu-x)^{\alpha^{(k)}}$ $(k=1, 2, \cdots, \rho)$, where $\alpha^{(1)}\leqq\alpha^{(2)}\leqq \cdots \leqq\alpha^{(\rho)}<\nu$, and if the remaining elementary divisors of $P(A, \mu)$ corresponding to the same linear factor $\mu-x$ are all of degree equal to or greater than ν, then according to Lemma I there exist matrices $R(\mu)$ and $S(\mu)$ such that $|R(x)|\neq 0$ and $|S(x)|\neq 0$ and that

$$R(\mu)P(A, \mu)S(\mu) \equiv Q(\mu), \qquad \text{mod } (\mu - x)^\nu,$$

where $Q(\mu)=(q_{ij}(\mu))$ $(i, j=1, 2, \cdots, n)$ is given by

$$\begin{aligned} q_{ij}(\mu) &\equiv 0, && \text{mod } (\mu - x)^\nu && (i \neq j),\\ q_{ii}(\mu) &\equiv (\mu - x)^{\alpha^{(i)}}, && \text{mod } (\mu - x)^\nu && (i \leqq \rho),\\ q_{ii}(\mu) &\equiv 0, && \text{mod } (\mu - x)^\nu && (i > \rho). \end{aligned}$$

By (5) $P(A, \mu)=QP(\overline{A}, \mu)Q^{-1}$, hence

$$R(\mu)QP(\overline{A}, \mu)Q^{-1}S(\mu) = Q(\mu), \text{ mod } (\mu - x)^\nu,$$

and (14) becomes

$$Q(\mu)S^{-1}(\mu)Q\mathfrak{T}'(\mu) \equiv 0, \text{ mod } (\mu - x)^\nu,$$

where T' is an $n\times\nu$ matrix formed of ν adjacent columns of T and $\mathfrak{T}'(\mu)$ is the $n\times 1$ matrix, whose elements are polynomials of degree $\nu-1$ in $\mu-x$, and $\mathfrak{T}'(x)$ is the first of these ν columns of T'. From this equation we see that the element of the kth row of the $n\times 1$ matrix $S^{-1}(\mu)$ $Q\mathfrak{T}'(\mu)$ is divisible by $(\mu-x)^{\nu-\alpha^{(k)}}$, if $k\leqq\rho$, and is prime to $\mu-x$, if $k>\rho$. Consequently the $n\times 1$ matrix

$$S^{-1}(x)Q\mathfrak{T}'(x) = U$$

† Roth, loc. cit., §3.

has ρ zero elements in the first ρ rows and arbitrary elements in the remaining $n-\rho$ rows. Now if $X-\mu I$ has the k elementary divisors $(\mu-x)^{\nu_i}$, $\nu_i \geqq \nu$ $(i=1, 2, \cdots, k)$, and if X is a solution of (1), then for each $(\mu-x)^{\nu_i}$ we must have

$$S^{-1}(x)Q\mathfrak{T}_i'(x) = U_i,$$

where U_i has zero elements in at least the first ρ rows. The reduced rank of $P(A, \mu)$ with respect to the modulus $(\mu-x)^{\nu_i}$, $\nu_i \geqq \nu$, cannot be less than ρ, and $S^{-1}(x)$ is not dependent upon the degree of the modulus $(\mu-x)^{\nu}$. Consequently

$$S^{-1}(x)Q(\mathfrak{T}_1'(x), \mathfrak{T}_2'(x), \cdots, \mathfrak{T}_k'(x)) = (U_1, U_2, \cdots, U_k).$$

The rank of $S^{-1}(x)Q$ is n and the rank of the right member is at most $n-\rho$; hence in order that the k columns $\mathfrak{T}_i'(x)$ $(i=1, 2, \cdots, k)$ of T may form a matrix of rank k, k cannot exceed $n-\rho$. The theorem here demonstrated is more general than Corollary I, but if $n_j-\alpha_j^{(\rho j)} \geqq 2$, the latter offers the more restrictive bound upon the number of equal elementary divisors that $X-\mu I$ may have.

THEOREM V. *If $P(a_i, \mu)=0$ has the root x_j of multiplicity β_{ij}, and if $P(\lambda, x_j)=0$ has the root a_i of multiplicity γ_{ij}, if $A-\lambda I$ has the elementary divisors $(\lambda-a_i)^{m_i}$ $(i=1, 2, \cdots, r)$ and if $X-\mu I$ has the elementary divisors $(\mu-x_j)^{n_j}$ $(j=1, 2, \cdots, s)$ where X is a solution of $P(A, X)=0$, then at least one $n_j (j=1, 2, \cdots, s)$ must equal or exceed the corresponding n_{ij} for each value of i $(i=1, 2, \cdots, r)$, where*

$$n_{ij} = \frac{m_i - 1}{\gamma_{ij} - 1} \qquad (\beta_{ij} > 1, \gamma_{ij} > 1),$$

$$n_{ij} = 2m_i - 1 \qquad (\beta_{ij} > 1, \gamma_{ij} = 1),$$

$$n_{ij} = \frac{m_i}{\gamma_{ij}} \qquad (\beta_{ij} = 1, \gamma_{ij} \geqq 1),$$

$$n_{ij} = \infty \qquad (\beta_{ij} = \gamma_{ij} = 0).$$

Under the hypotheses of this theorem neither $P(a_i, \mu)$ $(i=1, 2, \cdots, r)$ nor $P(\lambda, x_j)$ $(j=1, 2, \cdots, s)$ is identically zero. Hence the rank of $P(\lambda, X)$, where X is a solution of $P(A, X)=0$, and of $P(A, \mu)$ is n. Now if we regard A as a solution of (1), where X is the known matrix, then according to Corollary II, m_i is less than or equal to $\beta_i^{(n)}$, where $(\lambda-a_i)^{\beta_i^{(n)}}$ is the nth elementary divisor of $P(\lambda, X)$. Now at least one of the matrices $P(\lambda, X_j)$ $(j=1, 2, \cdots, s)$ must have $(\lambda-\alpha_i)^{\beta_i^{(n)}}$ as its n_jth elementary divisor corresponding to the linear factor $\lambda-a_i$. Lemma II gives us a means of computing an upper bound to the degree of the n_jth elementary divisor of $P(\lambda, X_j)$ $(j=1, 2, \cdots, s)$. For if we set $\Delta(\lambda)=P(\lambda, X_j)$, then

$$\delta_{i,i+k}(\lambda) = \frac{P_{0,k}(\lambda, x_j)}{k!} \qquad (k = 0, 1, \cdots, n_j - 1)$$

and if $(\mu-x_j)^{\gamma_{ij}}$, $\gamma_{ij}>1$, is a factor of $P(a_i, \mu)$, then $\delta_{i,i+1}(\lambda)=P_{0,1}(\lambda, x_j)$ will have $\lambda-a_i$ as a factor; on the other hand if $\gamma_{ij}=1$, $P_{0,1}(\lambda, x_j)$ is prime to $\lambda-a_i$. Let $P(\lambda, x_j)$ have the factor $(\lambda-a_i)_{\beta}{}^{ij}$ but not one of higher degree in $\lambda-a_i$; then according to Lemma II and Theorem III, m_i cannot exceed every m_{ij}, where

$$m_{ij} = \beta_{ij}n_j - n_j + 1 \qquad (\beta_{ij} > 1, \gamma_{ij} > 1),$$

$$m_{ij} = \frac{n_j + 1}{2} \qquad (\beta_{ij} = 1, \gamma_{ij} > 1),$$

$$m_{ij} = \beta_{ij}n_j \qquad (\beta_{ij} \geqq 1, \gamma_{ij} = 1),$$

$$m_{ij} = 0 \qquad (\beta_{ij} = 0, \gamma_{ij} = 0).$$

Hence not all n_j $(j=1, 2, \cdots, s)$ can be less than the numbers n_{ij} defined in the theorem. When $\beta_{ij}=\gamma_{ij}=0$, then $|P(A_i, x_j)|\neq 0$ and the corresponding $T_{ij}=0$, and we must take $m_{ij}=0$ since not all T_{ij} $(j=1, 2, \cdots, s)$ can be zero. Similarly n_{ij} must be taken sufficiently large in case $\beta_{ij}=\gamma_{ij}=0$.

According to the theorem above if $\beta_{ij}=\gamma_{ij}=1$, and if $\beta_{hj}=0$, $h\neq i$, and $\gamma_{ik}=0$, $k\neq j$, we must have $n_j\geqq m_i$, for $n_{ij}=m_i$ and $n_{ik}=\infty$, $k\neq j$. The m_ith elementary divisor of $P(A_i, \mu)$ corresponding to the linear factor $\mu-x_j$ is $(\mu-x_j)^{m_i}$ because of Lemma I, and $P(A_h, \mu)$, $h\neq i$, has only the elementary divisors unity corresponding to the same linear factor if $\beta_{hj}=0$. Hence the elementary divisor of highest degree of $P(A, \mu)$ corresponding to $\mu-x_j$ is $(\mu-x_j)^{m_i}$. That is, by Theorem III, $n_j\leqq m_i$. Consequently under the hypotheses here laid down $n_j=m_i$; and the equation $P(A_i, \mu)\ \mathfrak{T}_{ij}(\mu)\equiv 0$, mod $(\mu-x_j)^{m_i}$, has a solution such that $|T_{ij}|\neq 0$. We have consequently proved the following corollary.

Corollary IV. *If $A-\lambda I$ has the elementary divisors $(\lambda-a_i)^{m_i}$ $(i=1, 2, \cdots, r)$ such that $a_k\neq a_i$, $k\neq i$, and if the equations $P(a_i, \mu)=0$ $(i=1, 2, \cdots, r)$ have the distinct simple roots x_{ij} $(j=1, 2, \cdots, p_i)$ such that a_i is a simple root of each of the equations $P(\lambda, x_{ij})=0$ $(j=1, 2, \cdots, p_i)$ and that $P(a_k, x_{ij})\neq 0$, $k\neq i$ $(j=1, 2, \cdots, p_i)$, then $P(A, X)=0$ has $\sum_{i=1}^{r} p_i$ solutions, X, such that $X-\mu I$ has the elementary divisors $(\mu-x_{ij})^{m_i}$ $(i=1, 2, \cdots, r)$.*

The numbers x_{ij} in the elementary divisor $(\mu-x_{ij})^{m_i}$ $(i=1, 2, \cdots, r)$ can be chosen in p_i ways and all are distinct, for if $x_{ij}=x_{hk}$, $i\neq h$, then $P(a_h, x_{ij})=P(a_h, x_{hk})=0$, which is contrary to hypothesis.

III. The bilateral solution

Definition. *If* $B=(b_{ij})$ $(i=1, 2, \cdots, \alpha; j=1, 2, \cdots, \beta)$, *then* $B' = (b_{\beta-j-1,\alpha-i+1})$ *is the transverse of* B.

The transverse of a matrix is obtained by reflecting its elements with respect to a line at right angles to that with respect to which the transpose of the matrix is obtained. For example, if

$$B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix}, \quad B' = \begin{pmatrix} b_{32} & b_{22} & b_{12} \\ b_{31} & b_{21} & b_{11} \end{pmatrix}.$$

The following theorems hold:

The transverse of the sum of two or more matrices is equal to the sum of their transverses.

The transverse of the product of two or more matrices is equal to the product of their transverses taken in reverse order; $(AB)' = B'A'$.

If $A = aI + D$, *where* $D = (\delta_{ij})$ *and*

$$\delta_{i,i+1} = 1 \qquad (i = 1, 2, \cdots, n-1),$$
$$\delta_{ij} = 0 \qquad (i + 1 \neq j),$$

then $A' = A$.

Definition. *If* $B = (B_{ij})$, *where* B_{ij} $(i=1, 2, \cdots, r; j=1, 2, \cdots, s)$ *are* $\alpha_i \times \beta_j$ *matrices such that* $\sum_{i=1}^{r}\alpha_i = \alpha$, $\sum_{j=1}^{s}\beta_j = \beta$, *then the* $\beta \times \alpha$ *matrix*

$$B^* = (B_{ji}'),$$

where B_{ij}' *is the transverse of* B_{ij}, *is the compound transverse of* B *with respect to the sub-matrices* B_{ij}.

The compound transverse of a matrix depends upon the way it is divided into sub-matrices. The following theorems hold.

The compound transverse of the sum of two or more matrices is equal to the sum of the transverses of the addend matrices, provided all addend matrices are divided into sub-matrices in the same way.

If $B = (B_{ij})$ *and* $C = (C_{jk})$, *where* B_{ij} $(i=1, 2, \cdots, r; j=1, 2, \cdots, s)$. *are* $\alpha_i \times \beta_j$ *matrices and* C_{jk} $(j=1, 2, \cdots, s; k=1, 2, \cdots, t)$ *are* $\beta_j \times \gamma_k$ *matrices, such that* $\sum_{i=1}^{r}\alpha_i = \alpha$, $\sum_{j=1}^{s}\beta_j = \beta$, *and* $\sum_{k=1}^{t}\gamma_k = \gamma$, *then the compound transverse of* AB *is the* $\gamma \times \alpha$ *matrix obtained by multiplying the compound transverse of* B *on the right by the compound transverse of* A; *that is,*

$$(AB)^* = B^*A^*.$$

If $\overline{A}$ and $\overline{X}$ are the matrices in the normal forms as given in §II, *and if transversion of them is made with respect to their sub-matrices A_i $(i=1, 2, \cdots, r)$ and $X_j (j=1, 2, \cdots, s)$ respectively, then $\overline{A}^*=\overline{A}$ and $\overline{X}^*=\overline{X}$.*

If A is an $n\times n$ matrix, then the elementary divisors of $A-\lambda I$ are identical with those of $(A-\lambda I)^*$ for any division of $A-\lambda I$ into sub-matrices, and identical with those of $A^*-\lambda I$ provided the transversion of A is made with respect to its sub-matrices A_{ij} $(i, j=1, 2, \cdots, r)$ such that A_{ii} are all square matrices of order n_i and $\sum_{i=1}^{r} n_i=n$.

If B is a non-singular $n\times n$ matrix and if B^{-1} is its inverse, then for every division of B into sub-matrices there exists a corresponding division of B^{-1} into sub-matrices such that for B and B^{-1} so divided

$$(B^*)^{-1} = (B^{-1})^*.$$

If $B=(B_{ij})$ and $B^{-1}=(C_{ji})$ where B_{ij} are $\alpha_i\times\beta_j$ matrices and C_{ji} are $\beta_j\times\alpha_i$ matrices, the theorem is satisfied provided $\sum_{i=1}^{r}\alpha_i=\sum_{j=1}^{s}\beta_j=n$.

The idea of transversion and compound transversion of matrices as defined above enables us to determine the relationship of a solution on the right of (1) to one having the same normal form on the left, and their relation to the bilateral solution of the same equation.

THEOREM VI. *If $A=Q\overline{A}Q^{-1}$ and $X=R\overline{X}R^{-1}$, where $\overline{A}$ and $\overline{X}$ are the normal forms of A and X as defined in* §II, *if $A-\lambda I$ and $X-\mu I$ have the elementary divisors $(\lambda-a_i)^{m_i}$ $(i=1, 2, \cdots, r)$ and $(\mu-x_j)^{n_j}$ $(j=1, 2, \cdots, s)$ respectively, where $\sum_{i=1}^{r} m_i=\sum_{j=1}^{s} n_j=n$, and if X is a solution on the right of*

$$P(A, X) = \sum_{k=0}^{p} F_k(A)X^{p-k} = 0, \tag{1}$$

then

$$X_1 = R_1\overline{X}R_1^{-1}$$

is a solution of

$$\sum_{k=0}^{p} X^{p-k}F_k(A) = 0, \tag{15}$$

provided

$$R_1 = QQ^*(R^*)^{-1},$$

where Q and R are divided into sub-matrices of order $m_i\times n$ and $n_j\times n$ respectively.

If we assume that (15) has a solution on the left whose characteristic matrix $X_1-\mu I$ has the elementary divisors $(\mu-x_j)^{n_j}$ $(j=1, 2, \cdots, s)$ then

R_1 must exist such that $X_1 = R_1\overline{X}R_1^{-1}$. Hence by a procedure parallel to that of §II, we obtain from (15) the following equation:

$$\sum_{k=0}^{p} \overline{X}^{p-k} U F_k(\overline{A}) = 0, \tag{16}$$

where $U = R_1^{-1}Q$.

Now if we take the compound transverses of the members of (7) with respect to the sub-matrices A_i $(i=1, 2, \cdots, r)$, X_j $(j=1, 2, \cdots, s)$ and T_{ij} of $\overline{A}$, $\overline{X}$, and T respectively we have

$$\sum_{k=0}^{p} \overline{X}^{p-k} T^* F_k(\overline{A}) = 0.$$

That is $U = T^*$ satisfies (16) provided T satisfies (7), and similarly any U satisfying (16) is such that U^* satisfies (7). Hence

$$U = R_1^{-1}Q = T^* = R^*(Q^*)^{-1},$$

or $R_1 = QQ^*(R^*)^{-1}$ according to the theorems on the transversion of matrices. This proves the theorem.

In order that X be a bilateral solution of (1) it suffices but is not necessary that $R_1 = QQ^*(R^*)^{-1} = R$; in other words, that $RR^* = QQ^*$. The following theorem holds.

THEOREM VII. *In order that X be a bilateral solution of* (1) *it is necessary and sufficient that this equation have a solution $X = R\overline{X}R^{-1}$ on the right such that* (7) *is satisfied by T_1 and $T_2 = (T_{ij}^{(2)})$, not necessarily distinct, and such that $T_1T_2^* = I$, where T_2^* is the compound transverse of T_2 with respect to the sub-matrices $T_{ij}^{(2)}$ of order $m_i \times n_j$.*

If $X = R\overline{X}R^{-1}$ is a bilateral solution of (1), then $T_1 = Q^{-1}R$ satisfying (7) exists and $U = R^{-1}Q$ satisfying (16) exists and U^* also satisfies (7). That is, T_2, some solution of (7), is U^* or $(R^{-1}Q)^*$. It is sometimes possible that $T_1T_1^*$ cannot be a unit matrix, but if in T_2 we permit the parametric elements of T_1 to take another set of values, then it is possible for T_1 to be the inverse of T_2^* where T_2 is so taken.

IV. SOLUTIONS COMMUTATIVE WITH A

Little that is general can be said regarding the solutions of (1) which are commutative with A besides that already demonstrated to hold for the unilateral and bilateral solutions of the same equation. But with certain restrictions upon either A or X or on both we can derive such results on commutative matrices as are set forth in the following theorems.

THEOREM VIII. *If $AX=XA$, if $A-\lambda I$ has the elementary divisors $(\lambda-a_i)^{m_i}$ $(i=1, 2, \cdots, r)$, where $a_i\neq a_j$, $i\neq j$, and*

$$m_1 \geqq m_2 \geqq \cdots \geqq m_r,$$

and if $X-\mu I$ has the elementary divisors $(p-x_j)^{n_j}$ $(j=1, 2, \cdots, s)$, where

$$n_1 \geqq n_2 \geqq \cdots \geqq n_s,$$

but where $x_j (j=1, 2, \cdots, s)$ are not necessarily distinct; then

$$\sum_{k=1}^{h} n_k \leqq \sum_{k=1}^{h} m_k \qquad (h = 1, 2, \cdots, r)$$

and $s \geqq r$.

We shall here use the notation of §II with the understanding that the submatrices A_i $(i=1, 2, \cdots, r)$ and $X_j (j=1, 2, \cdots, s)$ of $\overline{A}$ and $\overline{X}$ are so ordered that their orders m_i and n_j respectively form non-increasing sequences of numbers. This is in no sense a restriction upon A nor upon X.

Since $AX=XA$, we have from (5) and (6) that

$$\overline{A}T\overline{X}T^{-1} = T\overline{X}T^{-1}\overline{A},$$

where $T=R^{-1}Q$. Therefore $T\overline{X}T^{-1}$ is commutative with $\overline{A}$, but the most general matrix commutative with $\overline{A}$, where all a_i $(i=1, 2, \cdots, r)$ are distinct, is $K=(K_{hk})$, where

$$K_{hk} = 0 \qquad (h \neq k),$$
$$K_{hh} = c_0 I_h + c_1 D_h + \cdots + c_{m_h-1} D_h^{m_h-1} \quad (h = 1, 2, \cdots, r),$$

and where D_h is an $m_h \times m_h$ matrix having only unit elements in the diagonal immediately above the principal diagonal and having zero elements in the remaining $m_h^2-m_h+1$ places.† Hence

$$T_{ij}X_j = K_{ii}T_{ij} \qquad (i = 1, 2, \cdots, r; j = 1, 2, \cdots, s),$$

and

$$T_{ij}x_j + T_{ij}D_j = c_0 T_{ij} + \left\{c_1 D_i + c_2 D_i^2 + \cdots + c_{m_i-1} D_i^{m_i-1}\right\} T_{ij}.$$

If $c_0 \neq x_j$, then $T_{ij}=0$, but not all T_{ij} $(i=1, 2, \cdots, r)$ may be zero else T would have n_j zero columns and would be singular. Hence we can assume that $c_0=x_j$ and the equation above reduces to

$$T_{ij}D_j = (c_1 D_i + c_2 D_i^2 + \cdots + c_{m_i-1} D_i^{m_i-1}) T_{ij}.$$

The multiplication of T_{ij} by D_j on the right moves the columns of T_{ij} one

† Kreis, *Contributions à la Théorie des Systèmes Linéaires*, Zurich Thesis, 1906. Hilton, *Linear Substitutions*, 1914, pp. 112–118.

space to the right, and the multiplication of T_{ij} on the left by D_i^k moves the rows of T_{ij} up k spaces. Because of this fact, whether c_1 is zero or not, T_{ij} has at least n_j-h zero elements in the first n_j-h places of the (m_i-h+1)st row $(h=1, 2, \cdots, m_i)$. If $n_j>m_i$, then at least the first n_j-m_i columns of T_{ij} have only zero elements. If any n_j exceeds every m_i $(i=1, 2, \cdots, s)$, then T will have at least one column of zero elements and this is impossible, consequently the largest n_j cannot exceed the largest m_i or

$$n_1 \leqq m_1.$$

This completes the first step of a mathematical induction proof. Now suppose we have shown that

$$\sum_{i=1}^{h} n_i \leqq \sum_{i=1}^{h} m_i, \tag{17}$$

and let

$$T = \begin{pmatrix} T_h & T_h^{(2)} \\ T_h^{(1)} & T_h^{(3)} \end{pmatrix},$$

where

$$\begin{aligned} T_h &= (T_{ij}) && (i, j = 1, 2, \cdots, h), \\ T_h^{(1)} &= (T_{ij}) && (i = h+1, h+2, \cdots, r; j = 1, 2, \cdots, h), \\ T^{(2)} &= (T_{ij}) && (i = 1, 2, \cdots, h; j = h+1, h+2, \cdots, s), \\ T^{(3)} &= (T_{ij}) && (i = h+1, h+2, \cdots, r; j = h+1, h+2, \cdots, s). \end{aligned}$$

If $n_{h+1} \leqq m_{h+1}$, then from (17) we have at once

$$\sum_{i=1}^{h+1} n_i \leqq \sum_{i=1}^{h+1} m_i$$

and the theorem holds; however if $n_{h+1}>m_{h+1}$, further proof is required. The number of zero columns in $T_{h+1}^{(1)}$ is at least equal to $\sum_{i=1}^{h+1}(n_i-m_{h+1})$, for if $n_{h+1}>m_{h+1}$, then every $n_i(i=1, 2, \cdots, h+1)$ exceeds m_j for $j \geqq h+1$ and all T_{ij} in $T_{h+1}^{(1)}$ have at least the first m_i-m_{h+1} columns of zero elements. The number of non-zero rows in T_{h+1} in the same columns where $T_{h+1}^{(1)}$ has only zero elements is equal to at most $\sum_{i=1}^{h+1}(m_i-m_{h+1})$. The rank of the first $\sum_{i=1}^{h+1} n_i$ columns of T will be less than $\sum_{i=1}^{h+1} n_i$ unless

$$\sum_{i=1}^{h+1}(n_i - m_{h+1}) \leqq \sum_{i=1}^{h+1}(m_i - m_{h+1}),$$

and the theorem is proved by mathematical induction for all h less than or equal to r or s. Since

$$\sum_{i=1}^{r} m_i = \sum_{i=1}^{s} n_j = n,$$

and if the inequality (17) holds, then it is not possible for r to be less than s. The theorem is proved.

COROLLARY V. *If $AX=XA$, if $A-\lambda I$ has the elementary divisors $(\lambda-a_i)^{m_i}$ $(i=1, 2, \cdots, r)$, where $a_i \neq a_k$, $i \neq k$, and*

$$m_1 \geqq m_2 \geqq \cdots \geqq m_r,$$

and if $X-\mu I$ has the elementary divisors $(\mu-x_j)^{n_j}$ $(j=1, 2, \cdots, s)$, where $x_j \neq x_h$, $j \neq h$, and

$$n_1 \geqq n_2 \geqq \cdots \geqq n_s,$$

then

$$m_i = n_i \qquad (i = 1, 2, \cdots, r),$$

and $r=s$.

This corollary is a direct consequence of the theorem above, for from it we find that

$$\sum_{i=1}^{h} n_i \leqq \sum_{i=1}^{h} m_i \text{ and } r \geqq s \qquad (h = 1, 2, \cdots, s);$$

and because $x_j \neq x_k$, $j \neq k$, that

$$\sum_{i=1}^{h} m_i \leqq \sum_{i=1}^{h} n_i \text{ and } r \leqq s \qquad (h = 1, 2, \cdots, r).$$

In order that these inequalities hold simultaneously, n_i must equal m_i and r must equal s.

The above theorem and corollary have obvious application to the solution of the equation $P(A, X)=0$ for X commutative with A. However, matrices A and X such that $A-\lambda I$ and $X-\mu I$ have the elementary divisors $(\lambda-a_i)^{m_i}$ and $(\mu-x_i)^{m}$ respectively are not necessarily commutative, so that even when Corollary IV of the preceding section is satisfied it is not a simple matter to show that the solutions, whose existence is there established, are also commutative with A.

EXTENSION DIVISION,
UNIVERSITY OF WISCONSIN,
MILWAUKEE, WIS.