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

\[ P(A, X) = \sum_{k=0}^{p} F_k(A) X^{p-k} = 0, \]

where \( A \) is a known \( n \times n \) matrix, \( F_k(\lambda) \) \((k = 0, 1, \ldots, p)\) are polynomials† in the scalar variable \( \lambda \), 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 \( \mu \) and \( \lambda \) 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 \( F_k(\lambda) \) \((k = 0, 1, \ldots, p-1)\); (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 \( F_k(\lambda) \) \((k = 0, 1, \ldots, p)\) 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 \( \lambda \) are permissible, provided that the characteristic values of \( A \) lie within or on the circles of convergence of each of the series representing \( F_k(\lambda) \) \((k = 0, 1, \ldots, p)\) and their first \( n-1 \) derivatives. For information on such functions of matrices the reader may consult Hensel, Über Potenzreihen 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.

689
I. Preliminary notions and lemmas

Definition. If \( A(\lambda) = (a_{ij}(\lambda)) \) \((i = 1, 2, \ldots, r; \ j = 1, 2, \ldots, s)\), where \( a_{ij}(\lambda) \) are polynomials in \( \lambda \), if
\[
a_{ij}(\lambda) = r_{ij}(\lambda), \mod (\lambda - a)^n \quad (i = 1, 2, \ldots, r; \ j = 1, 2, \ldots, s),
\]
and if \( R(\lambda) = (r_{ij}(\lambda)) \), then
\[
A(\lambda) \equiv R(\lambda), \mod (\lambda - a)^n.
\]

Definition. If \( A(\lambda) \) is an \( m \times m \) \( \lambda \)-matrix whose elements are polynomials in \( \lambda \), if
\[
A(\lambda) = R(\lambda), \mod (\lambda - a)^n,
\]
and if the \( i \)-th elementary divisor of \( R(\lambda) \) corresponding to the linear factor \( \lambda - a \) is \( (\lambda - a)^{a(i)} \) \((i = 1, 2, \ldots, \rho)\), where \( \rho \) is the rank of \( R(\lambda) \), then \( (\lambda - a)^{a(i)} \) \((i = 1, 2, \ldots, \rho)\) is the \( i \)-th elementary divisor of \( A(\lambda) \) and \( \rho \) its rank with respect to the modulus \((\lambda - a)^n\).

Definition. If \( A(\lambda) \) is an \( m \times m \) \( \lambda \)-matrix whose elements are polynomials in \( \lambda \), if the \( i \)-th elementary divisor of \( A(\lambda) \) corresponding to the linear factor \( \lambda - a \) is \( (\lambda - a)^{a(i)} \) \((i = 1, 2, \ldots, r)\), where \( r \) is the rank of \( A(\lambda) \), and if
\[
a^{(i)} < n \quad (i = 1, 2, \ldots, \sigma),
\]
\[
a^{(i)} \geq n \quad (i = \sigma + 1, \sigma + 2, \ldots, r);
\]
then \( \sigma \) is the reduced rank and \( (\lambda - a)^{a(i)} \) \((i = 1, 2, \ldots, \sigma)\) is the \( i \)-th elementary divisor of \( A(\lambda) \) with respect to the modulus \((\lambda - a)^n\).

Plainly if \( r \) is the rank \( A(\lambda) \) and if \( \rho \) and \( \sigma \) are respectively the rank and the reduced rank of \( A(\lambda) \) with respect to the modulus \((\lambda - a)^n\), then \( \sigma \leq \rho \leq r \leq m \); moreover it should be noted that in either case all minors of order \( k \leq \sigma \) are divisible by \( \prod_{i=1}^{k} (\lambda - a)^{a(i)} \) and that this \( k \)-th 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 \((\lambda - a)^{a(k)}\) the \( k \)-th elementary divisor of \( A(\lambda) \) even when \( a^{(k)} = 0 \). 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 \( a_{ij}(\lambda) \) \((i, j = 1, 2, \ldots, m)\) which are polynomials in \( \lambda \), and if the reduced rank of \( A(\lambda) \) with respect to the modulus \((\lambda - a)^n\) is \( \sigma \) and its \( i \)-th elementary divisor is \( (\lambda - a)^{a(i)} \) \((i = 1, 2, \ldots, \sigma)\), then two \( m \times m \) matrices, \( P(\lambda) \) and \( Q(\lambda) \), of degree \( n - 1 \) in \( \lambda \),
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
\[
P(\lambda)A(\lambda)Q(\lambda) = S(\lambda), \quad \text{mod } (\lambda - a)^n,
\]
where
\[
S(\lambda) = (s_{ij}(\lambda))
\]
and
\[
s_{ij}(\lambda) = 0, \quad \text{mod } (\lambda - a)^n \quad (i \neq j),
\]
\[
s_{ii}(\lambda) = (\lambda - a)^\alpha(i), \quad \text{mod } (\lambda - a)^n \quad (i = 1, 2, \cdots, r),
\]
\[
s_{ii}(\lambda) = 0, \quad \text{mod } (\lambda - a)^n \quad (i = \sigma + 1, \sigma + 2, \cdots, m),
\]
and
\[
\alpha^{(i)} \leq \alpha^{(i+1)} \quad (i = 1, 2, \cdots, \sigma - 1).
\]

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 \( \lambda \), exist such that
\[
T(\lambda)A(\lambda)U(\lambda) = \prod_{i=1}^{t} A_i(\lambda),
\]
where \( \lambda - a_i \) \((i = 1, 2, \cdots, t)\) are the distinct linear factors in \( \lambda \) 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
\[
a_{ij}^{(h)}(\lambda) = 0 \quad (i \neq j),
\]
\[
a_{ii}^{(h)}(\lambda) = (\lambda - a_h)^{\alpha^{(i)}} \quad (i \leq r),
\]
\[
a_{ii}^{(h)}(\lambda) = 0 \quad (i > r).
\]
Hence the \( k \)th composite elementary divisor of \( A(\lambda) \), as usually defined, is
\[
\prod_{h=1}^{t} (\lambda - a_h)^{\alpha^{(k)}} \quad (k = 1, 2, \cdots, r).
\]
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
\[
T(\lambda)A(\lambda)U(\lambda) = A_k(\lambda) \left[ \prod_{i=1}^{k-1} A_i(\lambda) \prod_{i=k+1}^{t} A_i(\lambda) \right].
\]
Now if the last \( n - r \) zero elements in the principal diagonal of
\[
A_i(\lambda) \quad (i = 1, 2, \cdots, k - 1, k + 1, k + 2, \cdots, t)
\]
be replaced by unit elements, the resulting matrix, \( A'_i(\lambda) \) \((i = 1, 2, \cdots, t)\),

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

$$
\prod_{i=1}^{k-1} A_i'(\lambda) \prod_{i=k+1}^{t} A_i'(\lambda) = \prod_{i=1}^{k-1} (\lambda - a_i)^{\alpha_1^{(i)}} \prod_{i=k+1}^{t} (\lambda - a_i)^{\alpha_1^{(i)}}
$$

and is a polynomial of degree $(\sum_{i=1}^{t} \alpha_1^{(i)}) - \alpha_1^{(j)}$ in $\lambda - a_k$ whose constant term is not zero. Hence the polynomial $v_j(\lambda)$ exists such that

$$
\prod_{i=1}^{k-1} (\lambda - a_i)^{\alpha_1^{(i)}} \prod_{i=k+1}^{t} (\lambda - a_i)^{\alpha_1^{(i)}} v_{jk}(\lambda) = 1 + (\lambda - a_k)^{n_k} w_{jk}(\lambda), \, j \leq r,
$$

where $w_{jk}(\lambda)$ is a polynomial in $\lambda$. 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 $v_{jk}(\lambda) = 1$ for $r < j \leq n$. Hence the diagonal matrix $V_k(\lambda)$, having as its $j$th element $v_{jk}(\lambda)$ as defined above, exists such that

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

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

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

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

$$
T(\lambda) \equiv P(\lambda), \quad \mod(\lambda - a)^n,
$$

$$
U(\lambda) V_k(\lambda) \equiv Q(\lambda), \quad \mod(\lambda - a)^n,
$$

$$
A_k(\lambda) \equiv S(\lambda), \quad \mod(\lambda - a)^n,
$$

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

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

$$
\delta_{i, i+k}(\lambda) \equiv (\lambda - a)^{m_k} d_k(\lambda) \quad (k = 0, \, 1, \ldots, \, n - 1),
$$

$$
\delta_{i, i-h}(\lambda) \equiv 0 \quad (h = 1, \, 2, \ldots, \, n - 1),
$$

where $d_k(\lambda) \quad (k = 0, \, 1, \ldots, \, n - 1)$ are polynomials in $\lambda$ and where $d_0(\alpha) \neq 0$ and $d_1(\alpha) \neq 0$, then the degree of the $n$th 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 $n$th elementary divisor of $\Delta(\lambda)$.

The minor $\delta_{i,0}(\lambda)$ has the factor $(\lambda - a)^{(n-1)\alpha_0}$, in $\lambda - a$; that of $\delta_{i,i+k}(\lambda)$ \((k = 1, 2, \ldots, n-1)\) is identically zero, whereas the minor of $\delta_{i,i-h}(\lambda)$ \((h = 1, 2, \ldots, n-1)\) 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

$$D_h(\lambda) = (\lambda - a)^{\alpha_0}d_i(\lambda)D_{h-1}(\lambda) - (\lambda - a)^{\alpha_0+\alpha_0}d_0(\lambda)d_i(\lambda) + \cdots$$

$$\pm (\lambda - a)^{(h-1)\alpha_0+\alpha_0}d_{h-1}(\lambda)d_h(\lambda) \quad (h = 1, 2, \ldots, n-1),$$

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 $\delta_{i,i-k}(\lambda)$ \((k = 1, 2, \ldots, n-1)\) 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 $n$th 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 $\delta_{i,i-k}(\lambda)$ \((k = 1, 2, \ldots, n-1)\) is divisible by $(\lambda - a)^{n-h-1+h/2}$ or $(\lambda - a)^{n-h-1-(h+1)/2}$ 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 $n$th 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 $d_{i,i-1}(\lambda)$ have this factor if $\alpha_1 = 0$. Hence in this case the $n$th elementary divisor of $\Delta(\lambda)$ corresponding to $\lambda - a$ is $(\lambda - a)^{n\alpha_2}$.

II. The unilateral solution

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

$$A_{i,j} = 0 \quad (i \neq j),$$

$$A_{i,i} = A_i \quad (i = 1, 2, \ldots, r),$$

and $A_i \,(i = 1, 2, \ldots, r)$ 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

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

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

$$X_{ij} = 0 \quad \text{ (} i \neq j \text{),}$$

$$X_{ii} = X_i = x_i I_i + D_i \quad \text{ (} j = 1, 2, \cdots, s \text{),}$$

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

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

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

$$\sum_{k=0}^{p} F_{k}(\bar{A})TX^{p-k} = 0, \tag{7}$$

where $T = Q^{-1}R$. In this equation $\bar{X}$ and $R$ (hence $T$) are the unknowns. In fact $T\bar{X}T^{-1}$ is a solution of $P(\bar{A}, X) = 0$; on the other hand if $X$ is a solution of $P(\bar{A}, X) = 0$, 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 $m_i \times n_i$ matrix; then from (7) we readily obtain the $rs$ equations

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

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 $x_i$ of $X_i$ and upon its order $n_i$.

Now from $X_i = x_i I_i + D_i$ we have

$$X_i^h = \sum_{k=0}^{h} \binom{h}{k} x_i^{h-k} D_i^k,$$

and consequently (8), for $A_i$ and $X_i$, becomes

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

where
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 $\mathcal{C}_{ij}(k)$; then

$$T_{ij} = (\mathcal{C}_{ij}(0), \mathcal{C}_{ij}(1), \ldots, \mathcal{C}_{ij}(n_j-1))$$

and

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

that is, the multiplication of $T_{ij}$ on the right by $D_j^k$ moves the first $n_i - k$ 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

$$\sum_{k=0}^{n_i-1} \frac{P_{0k}(A_i, x_j) \mathcal{C}_{ij}(k-h)}{h!} = 0$$

$(k = 0, 1, \ldots, n_j - 1)$.

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

$$(10) \quad P(A_i, \mu) \mathcal{C}_{ij}(\mu) = 0, \quad \text{mod } (\mu - x_j)^{n_j},$$

is thus obtained, where

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

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 $\mathcal{C}_{ij}(x_i) = \mathcal{C}_{ij}(0) = 0$, if $|P(A_i, x_i)| \neq 0$; hence also $\mathcal{C}_{ij}(1) = 0$ and so on, under the same hypothesis. That is, $T_{ij} = 0$, if $|P(A_i, x_i)| \neq 0$. Now not all $T_{ij}$ $(i = 1, 2, \ldots, r)$ can be zero, nor can all $T_{ij}$ $(j = 1, 2, \ldots, s)$ be zero else $T$ would have $n_i$ zero columns or $m_i$ zero rows and in either case would be a singular matrix. Hence $|P(A_i, x_i)| = 0$ for at least one pair of values of $i$ and $j$; the necessary and sufficient condition that such be the case is that $P(a_i, x_i) = 0$. We have proved the theorem.
Theorem I. If the characteristic matrix $A - \lambda I$, of $A$, have the elementary divisors $(\lambda - a_i)^{n_i}$, where $\sum_{i=1}^{r} n_i = n$, 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 $(\mu - x_j)^{m_j}$, where $\sum_{j=1}^{s} m_j = m$, then every equation $P(a_i, \mu) = 0$ $(i = 1, 2, \ldots, r)$ must be satisfied by at least one of the numbers $a_i$ $(i = 1, 2, \ldots, r)$, and every equation $P(x_j, \mu) = 0$ $(j = 1, 2, \ldots, s)$ must be satisfied by at least one of the numbers $x_j$ $(j = 1, 2, \ldots, s)$.

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 $X_i$ $(j = 1, 2, \ldots, s)$. For more definite information regarding them, we shall seek restrictions upon $n_i$, in addition to that we already know, namely that $\sum_{j=1}^{s} n_i = n$ 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 $(\mu - x_1)^{n_1}, (\mu - x_2)^{n_2}, \ldots, (\mu - x_s)^{n_s}$ corresponding to the linear factor $\mu - x$, then $(\mu - x)^{n_1} + (\mu - x)^{n_2} + \cdots + (\mu - x)^{n_s}$ is a factor of $|P(A, \mu)|$; moreover if $X - \mu I$ has the elementary divisors $(\mu - x_j)^{m_j}$ $(j = 1, 2, \ldots, s)$, then $\prod_{j=1}^{s} [P(\lambda, x_j)]^{m_j}$ 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

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

Hence the second part of the theorem is proved.

The restrictions placed upon $n_i$ by this theorem are not very severe; nevertheless the first part of the theorem places an upper bound upon $n_i$ and the second part places a lower bound upon $n_i$ $(j = 1, 2, \ldots, s)$. 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 $\rho$, with respect to the modulus $(\mu - x_i)^{n_i}$ and if the $\rho, \text{th}$ 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 \((\mu - x_i)^{\sigma_i}_n\), then the equation (1) may have the solution \(X\) whose characteristic matrix, \(X - \mu I\), has the elementary divisor \((\mu - x_i)^{\sigma_i}\) only if

\[
\alpha^{(\sigma_i)}_j \geq n_j + \rho_i - n;
\]

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

\[
\alpha^{(\sigma_i)}_j \geq n_j - \frac{n - \rho_i}{k}.
\]

If

\[
\frac{P_{h,0}(a,\mu)}{h!} = 0, \text{ mod } (\mu - x_i)^{\sigma_i}, \quad (h = 0, 1, \ldots, \sigma_i - 1),
\]

\[
\frac{P_{h,0}(a,\mu)}{h!} = \rho^{(h)}(\mu), \text{ mod } (\mu - x_i)^{\sigma_i}, \quad (h = \sigma_i, \sigma_i + 1, \ldots, m_i - 1);
\]

then \(P(A, \mu)\) is of rank \(\rho_{ij} = m_i - \sigma_{ij}\) with respect to the modulus \((\mu - x_i)^{\sigma_i}\) and equation (10) reduces to the following non-homogeneous system of \(\rho_{ij}\) equations in the \(\rho_{ij}\) unknowns \(\hat{h}_{ij}^{(r+h)}(\mu)\) \((h = 0, 1, \ldots, \rho_{ij} - 1)\):

\[
\begin{bmatrix}
(p^{(r)}(\mu) & p^{(r+1)}(\mu) & \cdots & p^{(m_i-1)}(\mu) \\
0 & p^{(r)}(\mu) & \cdots & p^{(m_i-2)}(\mu) \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & p^{(r)}(\mu)
\end{bmatrix}
\begin{bmatrix}
t^{(r)}(\mu) \\
t^{(r+1)}(\mu) \\
\vdots \\
t^{(m_i-1)}(\mu)
\end{bmatrix}
\]

\[
= (\mu - x_j)^{\sigma_i}
\begin{bmatrix}
q^{(r)}(\mu) \\
q^{(r+1)}(\mu) \\
\vdots \\
q^{(m_i-1)}(\mu)
\end{bmatrix}
\]

where \(q^{(h)}(\mu)\) \((h = r, r+1, \ldots, m_i - 1)\) are arbitrary polynomials in \(\mu\). The equation (10) imposes no restrictions upon \(t^{(r)}(\mu), t^{(r+1)}(\mu), \ldots, t^{(m_i-1)}(\mu)\), hence each of the first \(\sigma_{ij}\) 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

\[
t^{(r+h)}(\mu) = \frac{(\mu - x_j)^{\sigma_i} M_{h,\mu}}{[p^{(r)}(\mu)]^{m_i-s-h}} \quad (h = 0, 1, \ldots, \rho_{ij} - 1),
\]

\[\dagger\]In this equation and in the remainder of the proof of this lemma we suppress the subscripts \(i\) and \(j\) of \(p^{(h)}(\mu)\) of \(\sigma_{ij}\) and of \(t^{(r+h)}(\mu)\) save where ambiguity may arise.
where $M_{h}(\mu)$ is a linear combination of minors of order $\rho_{ij} - h - 1$ of $P(A, \mu)$ and consequently has the $(\rho_{ij} - h - 1)$th determinant divisor in $\mu - x_{i}$ of this matrix as its divisor. We shall now seek a lower bound to the degree of $\mu - x_{i}$ as a divisor of $t^{(k)}(\mu)$ ($k = \sigma, \sigma + 1, \cdots, m_{i} - 1$).

Let $p^{(e)}(\mu)$ have the factor $(\mu - x_{i})^{*}$ and $t^{(k)}(\mu)$ ($k = \sigma, \sigma + 1, \cdots, m_{i} - 1$) have the factor $(\mu - x_{i})^{(k)}$ and let neither have a divisor of higher degree than these in $\mu - x_{i}$. Moreover, let the $k$th elementary divisor of $P(A, \mu)$ be $(\mu - x_{i})^{(k)}$ ($k = 1, 2, \cdots, \rho_{ij}$). Then the determinant divisor of all minors of order $g$ of $P(A, \mu)$ is

$$
\prod_{k=1}^{g} (\mu - x_{i})^{(k)} , g \leq \rho_{ij},
$$

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

$$
\tau_{ij}^{(\sigma+h)} \geq \eta_{j} - \sum_{k=1}^{\rho_{ij}} \alpha_{ij}^{(k)} + h\eta_{i} + \sum_{k=1}^{\rho_{ij} - 1} \alpha_{ij}^{(k)},
$$

or

$$
(13) \quad \tau_{ij}^{(\sigma+h)} \geq n_{i} + h\eta_{i} - \sum_{k=n_{i}+k}^{\rho_{ij}} \alpha_{ij}^{(k)} (h = 0, 1, \cdots, \rho_{ij} - 1).
$$

This inequality evidently establishes a lower bound for the number of zero elements in the $(\sigma + h)$th row of $T_{ij}$, if for $\tau_{ij}^{(\sigma+h)} = k$ 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, $\tau_{ij}^{(\sigma+h)} \geq n_{i} - \alpha_{ij}^{(h)} (h = 0, 1, \cdots, \rho_{ij} - 1)$. Therefore $T_{ij}$ has only zero elements in at least the first $n_{i} - \alpha_{ij}^{(h)}$ columns of the last $\rho_{ij}$ rows, whereas the first $m_{i} - \rho_{ij} = \sigma_{ij}$ rows have arbitrary elements as was pointed out above. Since

$$
P(\overline{A}, \mu) = \begin{bmatrix}
P(A_{1}, \mu) & 0 & \cdots & 0 \\
0 & P(A_{2}, \mu) & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots \\
0 & 0 & \cdots & P(A_{r}, \mu)
\end{bmatrix},
$$

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

$$
\alpha_{ij}^{(1)} \leq \alpha_{ij}^{(2)} \leq \cdots \leq \alpha_{ij}^{(x_{ij})},
$$

then $(\mu - x_{i})^{(k)}$ is the $k$th elementary divisor of $P(\overline{A}, \mu)$ and of $P(A, \mu)$ with
respect to the modulus $(\mu - x_i)^{n_i}$ and $\alpha_j^{(p)}$ is the greatest of the numbers $\alpha_j^{(p_q)}$ $(q = 1, 2, \ldots, r)$.

According to (10)

\begin{equation}
P(A, \mu)T_j(\mu) = 0, \mod (\mu - x_i)^{n_i},
\end{equation}

where

\begin{equation}
\mathcal{G}_j(\mu) = \begin{bmatrix}
T_{1j}(\mu) \\
T_{2j}(\mu) \\
\vdots \\
T_{nj}(\mu)
\end{bmatrix} = \begin{bmatrix}
1 \\
\mu - x_j \\
\vdots \\
(\mu - x_j)^{n_j - 1}
\end{bmatrix} = T_j
\end{equation}

The matrix $T_j$ has $\sum_{i=1}^r (m_i - \rho_i) = n - \rho_j$ rows of arbitrary elements and the remaining rows have only zero elements in the first $n_j - \alpha_j^{(p)}$ columns. Hence $\alpha_j^{(p)}$ 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 $n_j > \alpha_j^{(p)}$ the reduced rank of $P(A, \mu)$ is $\rho_j$ 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 $n_j - \alpha_j^{(p)}$ zero columns in the remaining rows. Hence $k(n_j - \alpha_j^{(p)})$ 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 $\rho_j$ 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 $(\mu - x_j)^{\alpha_j^{(p)}}$ is its $\rho_j$th 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 $(n - \rho_j)/\alpha_j^{(p)}$ 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 $n_j \leq \alpha_j^{(p)} - n + r$, where $(\mu - x_j)^{n_j}$ is the $r$th elementary divisor of $P(A, \mu)$ corresponding to the linear factor $\mu - x_i$. Hence:

**Corollary II.** If $P(A, \mu)$ is of rank $r$ and if the $r$th elementary divisor of $P(A, \mu)$ corresponding to the linear factor $\mu - x_i$ is $(\mu - x_i)^{\alpha_j^{(p)}}$, 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 $\mu - x_i$ whose degree exceeds $\alpha_j^{(p)} - n + r$. 

License or copyright restrictions may apply to redistribution; see https://www.ams.org/journal-terms-of-use
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 \( k \leq n - r \).

In this case, where \( x \) is arbitrary, the \( r \)th 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\(^\dagger\), 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 \( \rho \) with respect to the modulus \((\mu - x)^\rho\), 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 \( \nu \) cannot exceed \( n - \rho \).

If \( P(A, \mu) \) has the elementary divisors \((\mu - x)^{\alpha(k)}\) \((k = 1, 2, \ldots, \rho)\), where \( \alpha^{(1)} \leq \alpha^{(2)} \leq \cdots \leq \alpha^{(\rho)} < \nu \), 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 \( \nu \), 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

\[
R(\mu)P(A, \mu)S(\mu) \equiv Q(\mu), \quad \text{mod} \ (\mu - x)^\rho,
\]

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

\[
\begin{align*}
q_{ii}(\mu) & \equiv 0, \quad \text{mod} \ (\mu - x)^\rho \\
q_{ij}(\mu) & \equiv (\mu - x)^{\alpha(i)}, \quad \text{mod} \ (\mu - x)^\rho \quad (i \leq \rho), \\
q_{ij}(\mu) & \equiv 0, \quad \text{mod} \ (\mu - x)^\rho \quad (i > \rho).
\end{align*}
\]

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

\[
R(\mu)QP(A, \mu)Q^{-1}S(\mu) = Q(\mu), \quad \text{mod} \ (\mu - x)^\rho,
\]

and (14) becomes

\[
Q(\mu)S^{-1}(\mu)Q\bar{T}'(\mu) \equiv 0, \quad \text{mod} \ (\mu - x)^\rho,
\]

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

\[
S^{-1}(x)Q\bar{T}'(x) = U
\]

\(^\dagger\) Roth, loc. cit., §3.
has \( p \) zero elements in the first \( p \) rows and arbitrary elements in the remaining \( n - p \) rows. Now if \( X - \mu I \) has the \( k \) elementary divisors \((\mu - x)^{v_i} \), \( v_i \geq v \) \((i = 1, 2, \ldots, k)\), and if \( X \) is a solution of (1), then for each \((\mu - x)^{v_i}\) we must have

\[
S^{-1}(x)Q \mathcal{G}_i(x) = U_i,
\]

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

Consequently

\[
S^{-1}(x)Q(\mathcal{G}_1(x), \mathcal{G}_2(x), \ldots, \mathcal{G}_k(x)) = (U_1, U_2, \ldots, U_k).
\]

The rank of \((S^{-1}(x)Q)\) is \( n \) and the rank of the right member is at most \( n - p \); hence in order that the \( k \) columns \( \mathcal{G}_i(x) \) \((i = 1, 2, \ldots, k)\) of \( T \) may form a matrix of rank \( k \), \( k \) cannot exceed \( n - p \). The theorem here demonstrated is more general than Corollary I, but if \( n_i - \alpha_i^{(p)} \geq 2 \), 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 \( x_i \) of multiplicity \( \beta_{ii} \), and if \( P(\lambda, x_j) = 0 \) has the root \( a_i \) of multiplicity \( \gamma_{ii} \), if \( A - \lambda I \) has the elementary divisors \((\lambda - a_i)^{m_i} \) \((i = 1, 2, \ldots, r)\) and if \( X - \mu I \) has the elementary divisors \((\mu - x_j)^{\nu_i} \) \((j = 1, 2, \ldots, s)\) where \( X \) is a solution of \( P(A, X) = 0 \), then at least one \( n_j \) \((j = 1, 2, \ldots, s) \) must equal or exceed the corresponding \( n_{ij} \) for each value of \( i \) \((i = 1, 2, \ldots, r)\), where

\[
\begin{align*}
n_{ij} &= \frac{m_i - 1}{\gamma_{ij} - 1} \quad (\beta_{ij} > 1, \gamma_{ij} > 1), \\
n_{ij} &= 2m_i - 1 \quad (\beta_{ij} > 1, \gamma_{ij} = 1), \\
n_{ij} &= \frac{m_i}{\gamma_{ij}} \quad (\beta_{ij} = 1, \gamma_{ij} \geq 1), \\
n_{ij} &= \infty \quad (\beta_{ij} = \gamma_{ij} = 0).
\end{align*}
\]

Under the hypotheses of this theorem neither \( P(a_i, \mu) \) \((i = 1, 2, \ldots, r)\) nor \( P(\lambda, x_j) \) \((j = 1, 2, \ldots, s)\) 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 \( n \)th elementary divisor of \( P(\lambda, X) \). Now at least one of the matrices \( P(\lambda, X_j) \) \((j = 1, 2, \ldots, s)\) must have \((\lambda - a_i)^{\beta_i^{(n)}} \) as its \( n \)th 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 \( n_i \)th elementary divisor of \( P(\lambda, X_i) \) \((j = 1, 2, \ldots, s)\). For if we set \( \Delta(\lambda) = P(\lambda, X_i) \), then

\[
\delta_i, i+k(\lambda) = \frac{P_{i,k}(\lambda, x_i)}{k!} \quad (k = 0, 1, \ldots, n_i - 1)
\]

and if \((\mu - x_i)^{\gamma_{ij}}, \gamma_{ij} > 1\), is a factor of \( P(a_i, \mu) \), then \( \delta_i, i+k(\lambda) = P_{i,1}(\lambda, x_i) \) will have \( \lambda - a_i \) as a factor; on the other hand if \( \gamma_{ij} = 1, P_{i,1}(\lambda, x_i) \) is prime to \( \lambda - a_i \). Let \( P(\lambda, x_i) \) 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

\[
\begin{align*}
m_{ij} &= \beta_{ij} n_i - n_j + 1 \quad (\beta_{ij} > 1, \gamma_{ij} > 1), \\
m_{ij} &= \frac{n_j + 1}{2} \quad (\beta_{ij} = 1, \gamma_{ij} > 1), \\
m_{ij} &= \beta_{ij} n_j \quad (\beta_{ij} \geq 1, \gamma_{ij} = 1), \\
m_{ij} &= 0 \quad (\beta_{ij} = 0, \gamma_{ij} = 0).
\end{align*}
\]

Hence not all \( n_i \) \((j = 1, 2, \ldots, s)\) 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 \( T_{ij} \) \((j = 1, 2, \ldots, s)\) 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 \( \beta_{ij} = 0, h \neq i, \) and \( \gamma_{ik} = 0, k \neq j \), we must have \( n_i \geq m_i \) for \( n_{ij} = m_i \) and \( n_{ik} = \infty \), \( k \neq j \). The \( m_i \)th elementary divisor of \( P(A_i, \mu) \) corresponding to the linear factor \( \mu - x_i \) is \((\mu - x_i)^{m_i}\) 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 \( \beta_{hi} = 0 \). Hence the elementary divisor of highest degree of \( P(A, \mu) \) corresponding to \( \mu - x_i \) is \((\mu - x_i)^{m_i}\). That is, by Theorem III, \( n_i \leq m_i \). Consequently under the hypotheses here laid down \( n_i = m_i \); and the equation \( P(A, \mu) \) \( \in \mathcal{C}_{ij}(\mu) \equiv 0, \) \( \mod (\mu - x_i)^{m_i} \), has a solution such that \( |T_{ij}| \neq 0 \). We have consequently proved the following corollary.

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

The numbers \( x_{ij} \) in the elementary divisor \((\mu - x_{ij})^{m_i} \) \((i = 1, 2, \ldots, r)\) can be chosen in \( p_i \) ways and all are distinct, for if \( x_{ij} = x_{ik}, i \neq k \), then \( P(a_k, x_{ij}) = P(a_k, x_{ik}) = 0 \), 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 \( B' = (b_{\alpha-i-1, \beta-j-1}) \) 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

\[
B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix}, \quad B' = \begin{pmatrix} b_{23} & b_{22} & b_{12} \\ b_{32} & b_{31} & b_{11} \end{pmatrix}.
\]

The following theorems hold:

1. The transverse of the sum of two or more matrices is equal to the sum of their transverses.
2. The transverse of the product of two or more matrices is equal to the product of their transverses taken in reverse order; \((AB)' = B'A'\).
3. If \( A = aI + D \), where \( D = (\delta_{ij}) \) and
   \[
   \delta_{i,i+1} = 1 \quad (i = 1, 2, \cdots, n - 1),
   \]
   \[
   \delta_{ij} = 0 \quad (i + 1 \neq j),
   \]
   then \( A' = A \).

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

\[ B^* = (B_{ij}'), \]

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.

1. 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.
2. If \( B = (B_{ij}) \) and \( C = (C_{jk}) \), where \( B_{ij} \ (i = 1, 2, \cdots, r; \ j = 1, 2, \cdots, s) \) are \( \alpha_i \times \beta_i \) 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 $A$ and $X$ are the matrices in the normal forms as given in §11, and if transversion of them is made with respect to their sub-matrices $A_i$ ($i = 1, 2, \ldots, r$) and $X_j$ ($j = 1, 2, \ldots, s$) respectively, then $A^* = A$ and $X^* = X$.

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 $A_{ij}$ ($i, j = 1, 2, \ldots, r$) 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 $B^{-1} = (C_{ij})$ where $B_{ij}$ are $\alpha_i \times \beta_i$ matrices and $C_{ij}$ are $\beta_i \times \alpha_i$ 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 $A = QAQ^{-1}$ and $X = RXR^{-1}$, where $\overline{A}$ and $\overline{X}$ are the normal forms of $A$ and $X$ as defined in §11, if $A - \lambda I$ and $X - \mu I$ have the elementary divisors $(\lambda - a_i)^{m_i}$ ($i = 1, 2, \ldots, r$) and $(\mu - x_j)^{n_j}$ ($j = 1, 2, \ldots, s$) 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

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

then

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

is a solution of

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

provided

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

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

If we assume that (15) has a solution on the left whose characteristic matrix $X_1 - \mu I$ has the elementary divisors $(\mu - x_j)^{n_j}$ ($j = 1, 2, \ldots, s$) then
R₁ must exist such that \( X₁ = R₁X₁R₁^{-1} \). Hence by a procedure parallel to that of §II, we obtain from (15) the following equation:

\[ \sum_{k=0}^{p} X^{p-k}UF_k(\bar{A}) = 0, \]

where \( U = R₁^{-1}Q \).

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

\[ \sum_{k=0}^{p} X^{p-k}T^*F_k(\bar{A}) = 0. \]

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

\[ U = R₁^{-1}Q = T^* = R^*(Q^*)^{-1}, \]

or \( R₁ = QQ^*(R^*)^{-1} \) 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 \( R₁ = QQ^*(R^*)^{-1} = R \); 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 \( X = RXR^{-1} \) on the right such that (7) is satisfied by \( T₁ \) and \( T₂ = (T₁)^{(2)} \), not necessarily distinct, and such that \( T₁T₂^* = I \), where \( T₂^* \) is the compound transverse of \( T₂ \) with respect to the sub-matrices \( T_{ij} \) of order \( m_i \times n_j \).

If \( X = RXR^{-1} \) is a bilateral solution of (1), then \( T₁ = Q^{-1}R \) satisfying (7) exists and \( U = R^{-1}Q \) satisfying (16) exists and \( U^* \) also satisfies (7). That is, \( T₂ \), some solution of (7), is \( U^* \) or \( (R^{-1}Q)^* \). It is sometimes possible that \( T₁T₁^* \) cannot be a unit matrix, but if in \( T₂ \) we permit the parametric elements of \( T₁ \) to take another set of values, then it is possible for \( T₁ \) to be the inverse of \( T₂^* \) where \( T₂ \) 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 $(\lambda - a_i)^{m_i}$ $(i = 1, 2, \ldots, r)$, where $a_i \neq a_j$, $i \neq j$, and

$$m_1 \geq m_2 \geq \cdots \geq m_r,$$

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

$$n_1 \geq n_2 \geq \cdots \geq n_s,$$

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

$$\sum_{k=1}^{h} n_k \leq \sum_{k=1}^{h} m_k \quad (h = 1, 2, \ldots, r)$$

and $s \geq r$.

We shall here use the notation of §II with the understanding that the submatrices $A_i (i = 1, 2, \ldots, r)$ and $X_j (j = 1, 2, \ldots, s)$ 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

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

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 $a_i (i = 1, 2, \ldots, r)$ are distinct, is $K = (K_{hk})$, where

$$K_{hh} = 0 \quad (h \neq k),$$

$$K_{hh} = c_0 D_h + c_1 D_h + \cdots + c_{m_i-1} D_h^{m_i-1} \quad (h = 1, 2, \ldots, r),$$

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

$$T_{ij}X_j = K_{ij}T_{ij} \quad (i = 1, 2, \ldots, r; j = 1, 2, \ldots, s),$$

and

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

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

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

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

space to the right, and the multiplication of \( T_{ij} \) on the left by \( D_i \) moves the rows of \( T_{ij} \) up \( k \) spaces. Because of this fact, whether \( c_i \) 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 \((h = 1, 2, \ldots, m_i)\). 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 \( m_i \) \((i = 1, 2, \ldots, s)\), 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

\[
\sum_{i=1}^{h} n_i \leq \sum_{i=1}^{h} m_i,
\]

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

\[
\begin{align*}
\sum_{i=1}^{h} n_i & \leq \sum_{i=1}^{h} m_i,
\end{align*}
\]

and let

\[
T = \begin{pmatrix}
T_h & T_{h}^{(2)} \\
T_{h}^{(1)} & T_{h}^{(3)}
\end{pmatrix},
\]

where

\[
\begin{align*}
T_h &= (T_{ij}) \quad (i, j = 1, 2, \ldots, h), \\
T_h^{(1)} &= (T_{ij}) \quad (i = h + 1, h + 2, \ldots, r; j = 1, 2, \ldots, h), \\
T_h^{(2)} &= (T_{ij}) \quad (i = 1, 2, \ldots, h; j = h + 1, h + 2, \ldots, s), \\
T_h^{(3)} &= (T_{ij}) \quad (i = h + 1, h + 2, \ldots, r; j = h + 1, h + 2, \ldots, s).
\end{align*}
\]

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

\[
\sum_{i=1}^{h+1} n_i \leq \sum_{i=1}^{h+1} m_i
\]

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 \( n_i \) \((i = 1, 2, \ldots, h+1)\) exceeds \( m_j \) for \( j \geq h+1 \) 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

\[
\sum_{i=1}^{h+1} (n_i - m_{h+1}) \leq \sum_{i=1}^{h+1} (m_i - m_{h+1}),
\]

and the theorem is proved by mathematical induction for all \( h \) less than or equal to \( r \) or \( s \). Since
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 \((\lambda - a_i)^{m_i}(i=1, 2, \ldots, r)\), where \( a_i \neq a_k, i \neq k \), and

\[
m_1 \geq m_2 \geq \cdots \geq m_r,
\]

and if \( X - \mu I \) has the elementary divisors \((\mu - x_j)^{n_j}(j=1, 2, \ldots, s)\), where \( x_j \neq x_k, j \neq k \), and

\[
n_1 \geq n_2 \geq \cdots \geq n_s,
\]

then

\[
m_i = n_i \quad (i = 1, 2, \ldots, r),
\]

and \( r = s \).

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

\[
\sum_{i=1}^{h} n_i \leq \sum_{i=1}^{h} m_i \quad \text{and} \quad r \geq s \quad (h = 1, 2, \ldots, s);
\]

and because \( x_j \neq x_k, j \neq k \), that

\[
\sum_{i=1}^{h} m_i \leq \sum_{i=1}^{h} n_i \quad \text{and} \quad r \leq s \quad (h = 1, 2, \ldots, r).
\]

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 \((\mu - x_i)^{n_i}\) 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.**