ON THE MATRIC EQUATIONS $P(X)=A$ AND $P(A, X)=0$

BY E. T. BROWNE

1. *Introduction.* The matric equation

$$P(X) = A, \tag{1}$$

where $P(\lambda)$ is a polynomial with scalar coefficients and A is a given square matrix of order n, has received a good deal of attention within the past few years. The problem is to find square matrices X satisfying (1). This equation may possess solutions X which are expressible as polynomials in A. On the other hand, there may exist solutions, but none expressible as a polynomial in A; and finally, there are equations of the type (1) which possess no solution at all.*

In 1928 Roth† found necessary and sufficient conditions that there may exist solutions of (1) expressible as polynomials in A, and he found the number of such solutions, in case any exists. He employed the theory of elementary divisors. In this paper Roth gave a bibliography which was quite complete up to that time. In 1931 Franklin attacked the problem through the canonical form, and found not only all solutions X that are expressible as polynomials in A, but also solutions that are not so expressible. Rutherford‡ also employed the canonical form. Still more recently Ingraham§ discussed the problem using the theory of elementary divisors.

Let $F_j(A)$, $(j=0, \cdots, m)$ be known polynomials in A with scalar coefficients, and consider the more general equation

---

* Franklin, *Algebraic matric equations*, Journal of Mathematics and Physics, Massachusetts Institute of Technology, vol. 10 (1932), pp. 289–314.

† Roth, *A solution of the matric equation* $P(X)=A$, Transactions of this Society, vol. 30 (1928), pp. 579–596.

‡ Rutherford, *On the canonical form of a rational integral function of a matrix*, Proceedings Edinburgh Mathematical Society, (2), vol. 3 (1932), pp. 135–143.

§ Ingraham, *On the rational solutions of the matrix equation* $P(X)=A$, Journal of Mathematics and Physics, vol. 13 (1934), pp. 46–50.

$$\sum_{j=0}^{m} F_j(A)X^{m-j} = 0, \tag{2}$$

considered by Roth* in a later paper and by Franklin.† This latter equation obviously includes (1) as a special case, and, since it presents little more difficulty than (1), we shall consider (2) first. Our problem then is to find necessary and sufficient conditions that there may exist solutions X of (2) which are expressible as polynomials in A, and to give a simple method for finding such solutions in case any exist. We shall not employ the theory of elementary divisors which leads to a somewhat complicated argument, nor the canonical form, since in actual practise the reduction of a matrix to the canonical form is itself quite tedious. On the other hand, we shall show that the principal *idempotent* and *nilpotent* matrices associated with the matrix A lend themselves quite readily to a solution of the problem.

2. *The Principal Idempotent and Nilpotent Matrices Associated with A.*‡ If λ is a scalar, the equation of degree n,

$$\psi(\lambda) = |\lambda I - A| = 0,$$

is called the *characteristic equation* of A. As is well known, A satisfies its own characteristic equation, but often this is not the equation of *lowest* degree that A satisfies. In fact, if $\theta(\lambda)$ denotes the highest common factor of all the $(n-1)$-rowed minors of $A-\lambda I$, and if we denote by $\phi(\lambda)$ the polynomial $\psi(\lambda)/\theta(\lambda)$, then $\phi(\lambda)=0$ is the equation of lowest degree which A satisfies. We shall hereafter refer to $\phi(\lambda)$ as the *reduced characteristic function* and to the equation $\phi(\lambda)=0$ as the *reduced* or *minimum equation* of A.

Let us suppose that $\phi(\lambda)$ when resolved into linear factors is of the form

$$\phi(\lambda) = \prod_{i=1}^{r} (\lambda - \alpha_i)^{\nu_i}, \qquad (\sum \nu_i = N), \tag{3}$$

* Roth, *On the equation $P(A,X)=0$ in matrices*, Transactions of this Society, vol. 35 (1933), pp. 689–708.

† Franklin, loc. cit.

‡ Wedderburn, *Lectures on Matrices*, American Mathematical Society Publications, 1934, pp. 23–29.

where the α's are distinct. For $r>1$, we write

$$h_i(\lambda) = \frac{\phi(\lambda)}{(\lambda - \alpha_i)^{\nu_i}}, \qquad (i = 1, \cdots, r).$$

We can then determine two scalar polynomials $M_i(\lambda)$ and $N_i(\lambda)$ of degrees not exceeding $\nu_i - 1$ and $N - \nu_i - 1$, respectively, such that

$$M_i(\lambda)h_i(\lambda) + N_i(\lambda)(\lambda - \alpha_i)^{\nu_i} \equiv 1.$$

If we write

$$\phi_i(\lambda) = M_i(\lambda)h_i(\lambda), \qquad (i = 1, \cdots, r),$$

and for $\phi_i(A)$ write ϕ_i, then ϕ_i is the *principal idempotent* element of A corresponding to the root α_i. The matrices ϕ_i satisfy the conditions

$$(4) \qquad \phi_i^k = \phi_i,$$

for any positive integer k;

$$(5) \qquad \phi_i\phi_j = 0, \quad (i \neq j); \qquad \sum_{i=1}^{r} \phi_i = I.$$

Moreover, these ϕ's are linearly independent and none is zero. Let us now denote by η_i the matric polynomial in A defined by the formula

$$\eta_i = \eta_i(A) = (A - \alpha_i I)\phi_i(A), \qquad (i = 1, 2, \cdots, r).$$

It is easily shown that these matrices η_i satisfy the conditions

$$(6) \qquad \eta_i^k \neq 0, \quad (k < \nu_i); \qquad \eta_i^{\nu_i} = 0;$$

$$(7) \qquad \eta_i\phi_i = \eta_i = \phi_i\eta_i; \qquad \eta_i\eta_j = 0, \quad (i \neq j);$$

and, moreover,

$$(8) \qquad A = \sum_{i=1}^{r} (\alpha_i\phi_i + \eta_i) = \sum \phi_i(\alpha_i + \eta_i).*$$

The matrix η_i is the *principal nilpotent* element of A corresponding to the root α_i. If $r=1$, so that $\phi(\lambda)$ reduces to $(\lambda - \alpha)^N$, we take $\phi = I$, $\eta = A - \alpha I$, and it will be seen at once that such of

* Here and throughout the remainder of the paper we have designated the scalar matrix $\alpha_i I$ merely by the symbol α_i.

the conditions (4), $\cdots$, (8) as are applicable hold also in this case.

3. LEMMA 1. *If the ϕ's and η's are defined as in the preceding section, and if $f(A)$ is any polynomial in A with scalar coefficients, then*

$$f(A) = \sum_{i=1}^{r} \phi_i f(\alpha_i + \eta_i) = \sum_{i=1}^{r} \phi_i f_i(\eta_i), \tag{9}$$

where the f_i are polynomials determined uniquely by $f(A)$. Conversely, any such expression is equal to a polynomial in A.

For $r > 1$, we have from (8), in view of (4), (5), and (7),

$$A^2 = \sum \phi_i(\alpha_i + \eta_i)^2,$$

and, in general, if k is any positive integer

$$A^k = \sum \phi_i(\alpha_i + \eta_i)^k.$$

Hence, if $f(A)$ is a polynomial with scalar coefficients, it follows that

$$f(A) = \sum \phi_i f(\alpha_i + \eta_i).$$

If now $f(\alpha_i + \eta_i)$ be expanded and written as a polynomial $f_i(\eta_i)$ in η_i, we have (9). Moreover, it is clear from the manner in which they arise that the f_i are unique.*

The converse follows at once since each of the ϕ's and η's is a polynomial in A. The lemma also holds obviously for $r = 1$.

LEMMA 2. *Let X be a matrix expressible as a polynomial in A, and therefore of the form on the right in* (9). *Necessary and sufficient conditions that X may be a solution of* (2) *are that the f_i satisfy the relations*

$$\sum_{j=0}^{m} F_j(\alpha_i + \eta_i) f_i^{m-j}(\eta_i) = 0, \qquad (i = 1, \cdots, r). \tag{10}$$

For, by Lemma 1, $F_j(A)$ is expressible in the form

$$F_j(A) = \sum_{i=1}^{r} \phi_i F_j(\alpha_i + \eta_i), \qquad (j = 0, \cdots, m).$$

---

* Otherwise, if we assume that $\sum \phi_i f_i(\eta_i) = \sum \phi_i \psi_i(\eta_i)$, it will follow, just as in the proof of the second part of Lemma 2, that $f_i = \psi_i$, $(i = 1, \cdots, r)$.

Also, if $X = f(A) = \sum_{\tau=0}^{r} \phi_\tau f_\tau(\eta_\tau)$, we have as above

$$X^{m-j} = \sum \phi_\tau f_\tau^{m-j}(\eta_\tau).$$

On substituting into (2) and making use of (4) and (5), we have

$$\sum_{i=1}^{r} \phi_i \sum_{j=0}^{m} F_j(\alpha_i + \eta_i) f_i^{m-j}(\eta_i) = 0. \tag{11}$$

The conditions (10) are therefore obviously sufficient. To prove that they are also necessary, multiply (11) through by ϕ_τ, and we obtain

$$\phi_i \sum_{j=0}^{m} F_j(\alpha_i + \eta_i) f_i^{m-j}(\eta_i) = 0, \qquad (i = 1, \cdots, r). \tag{12}$$

Now, by hypothesis, the f_i and all of the F's are polynomials. Hence, since $\eta_i^{\nu_i} = 0$, each of the equations (12) reduces to the form

$$\phi_i(a_0 + a_1\eta_i + \cdots + a_{\nu_i - 1}\eta_i^{\nu_i - 1}) = 0. \tag{13}$$

On multiplying through by η_i and recalling that $\phi_i\eta_i^k = (\phi_i\eta_i)^k$ we obtain

$$a_0(\phi_i\eta_i) + a_1(\phi_i\eta_i)^2 + \cdots + a_{\nu_i - 1}(\phi_i\eta_i)^{\nu_i} = 0.$$

Now the minimum equation of $\eta_i = \phi_i\eta_i$ is

$$(\phi_i\eta_i)^{\nu_i} = 0,$$

so that the matrices $\phi_i\eta_i, \cdots, (\phi_i\eta_i)^{\nu_i - 1}$ are linearly independent. Hence,

$$a_0 = a_1 = \cdots = a_{\nu_i - 2} = 0,$$

and it follows from (13) that we have also $a_{\nu_i - 1} = 0$. The lemma is therefore established.

4. *The Matric Equation* $P(A, X) = 0$. If for brevity we denote the left member of (2) by $P(A, X)$, the conditions (10) can clearly be written in the form

$$P[\alpha_i + \eta_i,\ f_i(\eta_i)] = 0. \tag{10'}$$

Since η_i behaves in all ways precisely as a scalar variable ξ, except that $\eta_i^{\nu_i} = 0$, and since a necessary and sufficient condition that (10′) hold is that the left member be divisible by $\eta_i^{\nu_i}$, we may state the following theorem.

THEOREM 1. *If α_i, $(i=1, \cdots, r)$, are the distinct roots, of multiplicities ν_i, respectively, of the minimum equation of A, necessary and sufficient conditions that a matrix X, expressible as a polynomial in A, satisfy* (2) *are that there exist polynomials $f_i(\xi)$, $(i=1, \cdots, r)$, such that the equations*

$$P[\alpha_i + \xi, f_i(\xi)] = 0, \qquad (i = 1, \cdots, r), \tag{14}$$

possess roots of multiplicities ν_i, respectively, at $\xi=0$. If such polynomials f_i exist, then $X = \sum \phi_i f_i(\eta_i)$ is a solution of (2).

Assuming that the conditions of Theorem 1 are satisfied, let us now proceed to find the polynomials f_i. Since $\eta_i^{\nu_i}=0$, the typical polynomial f_i may be taken in the form

$$f_i(\xi) = x_0 + x_1\xi + \cdots + x_{\nu-1}\xi^{\nu-1}. \tag{15}$$

From (14) we have then as a first necessary condition

$$P(\alpha_i, x_0) = 0. \tag{16}$$

Any root of this equation will serve as x_0. If $\nu_i=1$, f_i is determined. If $\nu_i>1$, we differentiate (14) as to ξ and put $\xi=0$, whence

$$P_{\alpha_i} + x_1 P_{x_0} = 0, \tag{17}$$

where the subscripts denote partial differentiation. If $P_{x_0}\neq 0$, that is, if there exists a simple root of (15) which may be chosen for x_0, then x_1 is uniquely determined. If, however, for every root x_0 of (16), we have $P_{x_0}=0$, it is necessary also that $P_{\alpha_i}=0$ at $\xi=0$. Differentiating (14) a second time and putting $\xi=0$, we have

$$P_{\alpha_i\alpha_i} + 2x_1 P_{\alpha_i x_0} + P_{x_0x_0}x_1^2 + 2P_{x_0}x_2 = 0,$$

and so on.

It will be noticed that the successive coefficients $x_1, x_2, \cdots$, of f_i enter these equations for the first time with the coefficient P_{x_0}. This leads to the Theorem of Franklin.*

THEOREM 2. *A sufficient condition that the equation $P(A, X)=0$ may have a solution X, expressible as a polynomial in A, is that*

* Franklin, loc. cit.

for each multiple root α_i *of the minimum equation of* A, *the equation* $P(\alpha_i, x_0) = 0$ *have at least one simple root.*

5. *Solution of the Matric Equation* $P(X) = A$. In particular, if our matric equation is of the form (1), the equation (14) becomes

$$P[f_i(\xi)] = \alpha_i + \xi. \tag{18}$$

In this case, (17) reduces to

$$x_1 P_{x_0} = 1,$$

so that the condition $P_{x_0} \neq 0$ is not only a *sufficient* condition that x_1 may exist, but a *necessary* one also. Since each simple root of (16) yields a unique polynomial f_i corresponding to it, it follows from the uniqueness feature of Lemma 1 that we are led to the Theorem of Roth.*

THEOREM 3. *Let* A *be a square matrix whose minimum equation has the* s *multiple roots* $\alpha_1, \cdots, \alpha_s$, *and the* t *simple roots* $\alpha_{s+1}, \cdots, \alpha_{s+t}$. *The matric equation* $P(X) = A$ *has a solution for* X *as a polynomial in* A *if, and only if, each of the equations*

$$P(x) = \alpha_i, \qquad (i = 1, \cdots, s),$$

has at least one simple root. If these equations have respectively $\mu_1, \cdots, \mu_s$ *simple roots and if* $\mu_{s+1}, \cdots, \mu_{s+t}$ *denote the number of distinct roots of the equations*

$$P(x) = \alpha_i, \quad (i = s + 1, \cdots, s + t),$$

the total number of solutions of (1) *for* X *as a polynomial in* A *is*

$$\mu_1 \mu_2 \cdots \mu_{s+t}.$$

COROLLARY. *If the minimum equation of* A *has all roots distinct, the matric equation* $P(X) = A$ *is always solvable for* X *as a polynomial in* A. *If* μ_i *denotes the number of distinct roots of the equation* $P(x) = \alpha_i$, *the total number of such solutions is* $\mu_1 \mu_2 \cdots \mu_r$.

THE UNIVERSITY OF NORTH CAROLINA

---

* Roth, Transactions of this Society, vol. 30 (1928), pp. 579–596.