A SOLUTION OF THE MATRIC EQUATION $P(X)=A$*

BY
WILLIAM E. ROTH

I. Introduction

The equation $P(X)=A$, where $P(\lambda)$ is a polynomial in λ with scalar coefficients and A is a square matrix of order n, has received the attention of various writers. Perhaps the first to deal with the solution of such an equation of degree greater than the first was Cayley in his *Memoir on the theory of matrices.*† He there gave a solution for the equation $L=M^{1/2}$, where M was a known matrix of order two or of order three. The theory expounded in the remarkable memoirs of Cayley was further developed by Sylvester;‡ he gave a general solution of the equation $X^p=A$, but he did not give the deductions that led him to his results, nor did he discuss the conditions under which his solution applies. He asserted that the solutions of $X^p=A$ are p^μ in number, where μ is the number of distinct roots of the characteristic equation of A. The statement is correct for the kind of solutions he gave, namely, those expressible as polynomials in the given matrix. He recognized the existence of solutions not so expressible in case $X^p=I$, where I is the identical matrix, and later treated this particular case separately.§ In an article‖ which appeared in 1883, he called attention to the relationship of quaternions to matrices of the second order and gave a definition of the four units of quaternions in terms of matrices; and from then on he took up the discussion of quadratic equations in quaternions.¶

The work of Sylvester advanced the subject considerably; but the increased interest in mathematical foundations and in logical rigor led to new

---

* Presented to the Society, September 9, 1927; received by the editors December 19, 1927.

† See Philosophical Transactions of the Royal Society of London, vol. 148 (1858), pp. 17–37; or Collected Mathematical Papers, vol. II, pp. 475–496.

‡ Sylvester, *Sur les puissances et les racines de substitutions linéaires*, Comptes Rendus, vol. 94 (1882), pp. 55–59; or Mathematical Papers, vol. III, pp. 562–4.

§ Sylvester, *Sur les racines des matrices unitaires*, Comptes Rendus, vol. 94 (1882), pp. 396–9; or Mathematical Papers, vol. III, pp. 565–7.

‖ Sylvester, *On the involution and evolution of quaternions*, Philosophical Magazine, vol. 16 (1883), pp. 394–396; or Mathematical Papers, vol. IV, pp. 112–114.

¶ Sylvester, *Sur la solution explicite de l'équation quadratique de Hamilton en quaternions ou en matrices du second ordre*, Comptes Rendus, vol. 99 (1884), pp. 555–8, 621–631; or Mathematical Papers, vol. IV, pp. 188–198.

treatments of some of his problems; the principal contributions to the algebra of matrices from the modern point of view were made by Frobenius. His solution of the binomial equation,* $X^2=A$, where A is a non-singular square matrix, has found its way into modern textbooks.† Dickson‡ gives practically the same solution extended to any degree in X. Frobenius stated that his solution may readily be extended to apply when A is a singular matrix, but how this can be done is not clear from his discussion and apparently has never been accomplished by his method. The solutions of Frobenius are p^μ in number and are expressible as polynomials in the given matrix; in both these respects his results agree with those of Sylvester.

The introduction of the Weierstrass§ elementary divisors opened a new mode of attack upon the problems we are considering here. This was employed by Kreis‖ in his general solution of the equation $P(X)=A$ defined above, for which he gave solutions that are expressible as polynomials in the given matrix; his results are expressed in terms of the Weierstrass elementary divisors and associated normal forms. Later he¶ treated the binomial equation $X^p=A$ separately and gave a criterion for the existence of solutions when A is non-singular or singular. A course similar to that of Kreis was followed by Cecioni,** who solved the equation $X^p=A$, but not the more general equation, $P(X)=A$. Cecioni calls the solutions formed by Frobenius "soluzioni singolari" and says that these and only these can be expressed as linear aggregates of powers of the given matrix; he gives a criterion by means of which one should be able to determine the existence of such solutions. In this, however, he seems to have been led into an error as will be pointed out later. He further considers the possible solutions in a field F which contains the elements of A. The paper of Cecioni, like those of Kreis, is very difficult to read because of the difficulties involved by the use of elementary divisors.

---

* Frobenius, *Über die cogredienten Transformation der bilinearen Formen*, Sitzungsbericht der Königlichen Preussischen Akademie der Wissenschaften, 1896.

† See Muth, *Theorie und Anwendung der Elementarteiler*, Leipzig, Teubner, 1899. Bôcher, *Introduction to Higher Algebra*, New York, Macmillan, 1907.

‡ Dickson, *Modern Algebraic Theories*, Chicago, Sanborn, 1926.

§ Weierstrass, *Zur Theorie der bilinearen und quadratischen Formen*, Monatsberichte der Königlichen Preussischen Akademie der Wissenschaften, 1868, pp. 310–338.

‖ Kreis, *Contribution à la Théorie des Systèmes linéaires*, Zürich Thesis, 1906.

¶ Kreis, *Auflösung der Gleichung* $X^m=A$, Vierteljahrschrift der Naturforschenden Gesellschaft in Zürich, 53te Jahrgang, 1908, pp. 366–376.

** Cecioni, *Sopra alcune operazioni algebriche sulle matrici*, Annali della Reale Scuola Normale Superiore di Pisa, vol. 11 (1909), pp. 1–40.

The present paper will give such solutions of the equation $P(X)=A$ as are expressible as polynomials in A, and a criterion for the existence of such solutions will be established. The method here to be developed has an advantage over that of Frobenius, in that it applies as well in certain cases where A is singular and is not restricted to the binomial equation. It is based upon the given equation and the polynomial $\psi(\lambda)$, for which $\psi(A)=0$, which may or may not be the equation of lowest degree satisfied by A. The use of infinite expansions such as were used by Frobenius and others is entirely avoided, thus removing the doubt which must arise when that method is used, in as much as a series in a given matrix may not be convergent when it is convergent for scalar quantitites. However, the method does not give all the solutions that may occur when there exists an equation $\psi(A)=0$ of degree lower than the order of A. This fact was pointed out by Sylvester and becomes evident if we regard the equation $P(X)=kI$; the solutions of this equation expressible as polynomials in I must necessarily have the form $X=\alpha I$, where α is a root of the equation $P(\lambda)=k$, whereas other solutions are known to exist according to the results of Kreis and Cecioni cited above.

Frobenius* gave two theorems that are closely related to Theorem II of the present paper but they cannot be used to prove the existence of a solution for the equation $P(X)=A$. At any rate Frobenius did not refer to them when he solved the equation $X^2=A$, even though the theorems were published previous to the appearance of his solution of this equation.

II. Preliminary theorems

THEOREM I. *If $\psi(\lambda)$ is a polynomial of degree $m>1$ in λ, and the distinct roots of $\psi(\lambda)=0$ are α_j, $j=1, 2, \cdots, s$; if $P(\lambda)$ is a polynomial in λ of degree $p>1$, whose leading coefficient is unity and whose constant term is zero; and if the equation $P(\lambda)-\alpha_j=0$ has at least one simple root for each α_j, $j=1, 2, \cdots, s$, which is a multiple root of $\psi(\lambda)=0$; then polynomials $\phi_i(\lambda)$, $i=1, 2, \cdots, p$, of degree m in λ exist such that*

$$\prod_{i=1}^{p}\phi_i(\lambda)\equiv\psi(P(\lambda)),$$

and such that at least one, $\phi_k(\lambda)$, $1\leqq k\leqq p$, has no quadratic factor in λ in common with any $P(\lambda)-\alpha_j$, $j=1, 2, \cdots, s$.

---

* Frobenius, *Über lineare Substitutionen und bilineare Formen*, Crelle's Journal, vol. 84 (1878), pp. 1–63.

Suppose that $\psi(\lambda)$ is given by the identity

$$\psi(\lambda) \equiv \prod_{j=1}^{s}(\lambda - \alpha_j)^{\epsilon_j}, \tag{1}$$

where α_j, $j=1, 2, \cdots, s$, are the distinct roots of $\psi(\lambda)=0$ and where $\sum_{j=1}^{s} \epsilon_j = m$. We have, by hypothesis,

$$P(\lambda) \equiv \lambda^p + h_1\lambda^{p-1} + h_2\lambda^{p-2} + \cdots + h_{p-1}\lambda\ ; \tag{2}$$

and we assume further that $P(\lambda)-\alpha_j$, $j=1, 2, \cdots, s$, is given by

$$P(\lambda) - \alpha_j \equiv \prod_{i=1}^{p}(\lambda - \beta_{ij}) \qquad (j = 1,2, \cdots, s)\ ; \tag{3}$$

then $-h_1, h_2, \cdots, (-1)^{p-1}h_{p-1}, (-1)^{p-1}\alpha_j$ are the elementary symmetric functions of the roots β_{ij}, $i=1, 2, \cdots, p$, of the equations $P(\lambda)-\alpha_j=0$, $j=1, 2, \cdots, s$. From (1) and (3), it follows that

$$\begin{aligned}\psi(P(\lambda)) &\equiv \prod_{j=1}^{s}(P(\lambda) - \alpha_j)^{\epsilon_j}, \\ &\equiv \prod_{j=1}^{s}\Big[\prod_{i=1}^{p}(\lambda - \beta_{ij})\Big]^{\epsilon_j}.\end{aligned}$$

The factors of the right member of this identity may be rearranged in a number of ways so as to give

$$\psi(P) \equiv \prod_{i=1}^{p}\Big[\prod_{j=1}^{s}(\lambda - \beta_{ij})^{\epsilon_j}\Big], \tag{4}$$

where here and in the following the abbreviated notation, $\psi(P)$, will be used to denote $\psi(P(\lambda))$ regarded as a polynomial in λ unless the contrary is specifically stated, and the corresponding notation will be used with the same significance for other polynomials in $P(\lambda)$. Now let

$$\phi_i(\lambda) \equiv \prod_{j=1}^{s}(\lambda - \beta_{ij})^{\epsilon_j} \qquad (i = 1,2, \cdots, p)\ ; \tag{5}$$

then $\phi_i(\lambda)$, $i=1, 2, \cdots, p$, will be polynomials of degree $\sum_{j=1}^{s} \epsilon_j = m$ in λ; and by (4),

$$\psi(P) \equiv \prod_{i=1}^{p}\phi_i(\lambda). \tag{6}$$

It still remains to be shown that at least one $\phi_k(\lambda)$, $1 \leqq k \leqq p$, has no factor of the second degree in λ in common with any polynomial $P(\lambda)-\alpha_j$, $j=1, 2,$

$\cdots, s$. To this end we write the roots of each equation $P(\lambda)-\alpha_j=0$, $j=1, 2, \cdots, s$, in a separate column thus:

$$\begin{array}{cccc} P(\lambda)-\alpha_1=0, & P(\lambda)-\alpha_2=0, & \cdots, & P(\lambda)-\alpha_s=0; \\ \beta_{11}, & \beta_{12}, & \cdots, & \beta_{1s}; \\ \beta_{21}, & \beta_{22}, & \cdots, & \beta_{2s}; \\ \cdots & \cdots & \cdots & \cdots \\ \beta_{p1}, & \beta_{p2}, & \cdots, & \beta_{ps}. \end{array}$$

Then the β's occurring in any one column are distinct from those occurring in any other column; for suppose $\beta_{rh}=\beta_{tk}$, $h\neq k$; then by (3), $P(\beta_{rh})=\alpha_h$, and $P(\beta_{tk})=\alpha_k$, but since $\beta_{rh}=\beta_{tk}$, it follows that $\alpha_h=\alpha_k$, $h\neq k$. This is impossible because α's with different subscripts are distinct. Then no β of any one column of the above table is repeated in any other column. On the other hand, β's in the same column are not necessarily distinct from each other. The definition of $\phi_i(\lambda)$, $i=1, 2, \cdots, p$, given in (5) shows that $\lambda-\beta_{ij}$, $j=1, 2, \cdots, s$, has the same exponent, ϵ_j, in $\phi_i(\lambda)$ that the corresponding factor, $\lambda-\alpha_j$, has in $\psi(\lambda)$; consequently it may be said that a certain polynomial $\phi_k(\lambda)$, $1\leqq k\leqq p$, is formed by taking for its factors one β from each column of the table above; each factor $\lambda-\beta_{kj}$ so taken is given the same exponent ϵ_j in $\phi_k(\lambda)$ that $\lambda-\alpha_j$ has in $\psi(\lambda)$. Thus there is a one-to-one correspondence between the factors $(\lambda-\beta_{kj})^{\epsilon_j}$ of $\phi_k(\lambda)$ and $(\lambda-\alpha_j)^{\epsilon_j}$ of $\psi(\lambda)$. The only way in which any $\phi_k(\lambda)$ can have a factor of the second degree in λ in common with some $P(\lambda)-\alpha_j$, $j=1, 2, \cdots, s$, is that $\lambda-\beta_{kj}$ be a multiple factor of $P(\lambda)-\alpha_j$, where $\lambda-\alpha_j$ is a multiple factor of $\psi(\lambda)$. But by hypothesis we know that $P(\lambda)-\alpha_j$ has at least one simple factor, when $\lambda-\alpha_j$ is a multiple factor of $\psi(\lambda)$. Consequently it is possible to construct at least one polynomial $\phi_k(\lambda)$ which will have no factor of the second degree in λ in common with any $P(\lambda)-\alpha_j$, $j=1, 2, \cdots, s$. The remaining polynomials $\phi_i(\lambda)$, $i=1, 2, \cdots, k-1, k+1, \cdots, p$, must then be formed in one of $(p-1)^s$ ways, not necessarily all distinct, from the $p-1$ elements remaining in each column of the above table; together with $\phi_k(\lambda)$, they will satisfy identity (6).

In general, we can say the polynomial $\phi_k(\lambda)$, having the above properties, may be formed in $\prod_{j=1}^{s}\mu_j$ distinct ways, where s is the number of distinct roots of $\psi(\lambda)=0$, and where μ_j is the number of distinct roots of $P(\lambda)-\alpha_j=0$ when α_j is a simple root of $\psi(\lambda)=0$ and the number of simple roots of $P(\lambda)-\alpha_j=0$ when α_j is a multiple root of $\psi(\lambda)=0$. If some polynomial $P(\lambda)-\alpha_j$ has only multiple factors when α_j is a multiple root of $\psi(\lambda)=0$, then the

corresponding $\mu_j=0$ and no polynomial $\phi_k(\lambda)$ can be formed satisfying the conditions above.

THEOREM II. *Under the conditions of Theorem* I *and with the polynomial,* $\phi_k(\lambda)$, *whose existence was there established, there exist polynomials* $H_k(\lambda)$, $Z_k(\lambda)$, *and* $T_k(\lambda)$, *the latter of degree less than* m *in* λ, *such that*

$$H_k(\lambda)\phi_k(\lambda) \equiv \lambda - T_k(P(\lambda)),$$

and that

$$Z_k(\lambda)\psi(\lambda) \equiv \lambda - P(T_k(\lambda)).$$

We have, as under the preceding theorem,

$$\psi(\lambda) \equiv \prod_{j=1}^{s}(\lambda - \alpha_j)^{\epsilon_j}, \tag{1}$$

where the α_j, $j=1, 2, \cdots, s$, are distinct;

$$P(\beta_{kj}) = \alpha_j \qquad (j = 1,2, \cdots, s); \tag{7}$$

and

$$\phi_k(\lambda) \equiv \prod_{j=1}^{s}(\lambda - \beta_{kj})^{\epsilon_j}. \tag{5}$$

Furthermore, $\phi_k(\lambda)$ has no quadratic factor in common with any polynomial $P(\lambda)-\alpha_j$, $j=1, 2, \cdots, s$, and because of (6) we can write the identity

$$\psi(P) \equiv \phi_k(\lambda)Q(\lambda), \tag{8}$$

where $Q(\lambda)$ is a polynomial in λ.

We shall show that every polynomial $\psi_1(\lambda)$, not identically zero, and such that

$$\psi_1(P) \equiv \phi_k(\lambda)Q_1(\lambda), \tag{9}$$

is divisible by $\psi(\lambda)$, and consequently there exists no polynomial $\psi_1(\lambda)$ of degree lower than m in λ which can satisfy an identity of this kind. Let $\psi_1(\lambda)$ be any polynomial that satisfies the identity (9), and substitute β_{kj}, $j=1, 2, \cdots, s$, for λ in this identity; then because of (5) and (7),

$$\psi_1(\alpha_j) = 0 \qquad (j = 1,2, \cdots, s). \tag{10}$$

That is, the distinct factors of $\psi(\lambda)$ must also be factors of $\psi_1(\lambda)$. If we can now show that the multiplicity of any factor $(\lambda-\alpha_\tau)$, $1 \leqq \tau \leqq s$, of $\psi_1(\lambda)$ is ϵ_τ, then according to (1) $\psi_1(\lambda)$ must be divisible by $\psi(\lambda)$. In case all factors of $\psi(\lambda)$ are simple, this assertion is already proved. If $(\lambda-\alpha_\tau)^{\epsilon_\tau}$, $\epsilon_\tau>1$, is a factor of $\psi(\lambda)$, then $(\lambda-\beta_{k\tau})^{\epsilon_\tau}$ is a factor of $\phi_k(\lambda)$; we shall now show

that $\psi_1(\lambda)$ must have the factor $(\lambda-\alpha_\tau)^{\epsilon_\tau}$ in common with $\psi(\lambda)$. For this purpose we differentiate the members of (9) with respect to λ, and we find

$$\psi_1'(P)P'(\lambda) \equiv \{\phi_k(\lambda)Q_1(\lambda)\}'.$$

Substitute $\beta_{k\tau}$ for λ in this identity; since $\epsilon_\tau>1$, the right member is zero and we have

$$\psi_1'(\alpha_\tau)P'(\beta_{k\tau}) = 0.$$

Now $(\lambda-\beta_{k\tau})^{\epsilon_\tau}$, $\epsilon_\tau>1$, is a factor of $\phi_k(\lambda)$ and since $P(\lambda)-\alpha_\tau$ has the factor $\lambda-\beta_{k\tau}$, $P'(\lambda)$ cannot have this factor, for the polynomials $P(\lambda)-\alpha_\tau$ and $\phi_k(\lambda)$ have no quadratic factor in λ in common. Therefore

$$P'(\beta_{k\tau}) \neq 0, \tag{11}$$

and we must have

$$\psi_1'(\alpha_\tau) = 0. \tag{12}$$

Thus $(\lambda-\alpha_\tau)^2$ is a factor of $\psi_1(\lambda)$. If now $\epsilon_\tau>2$, we take the second derivatives with respect to λ of the members of (9), and thus obtain the identity

$$\psi_1''(P)P'^2(\lambda) + \psi_1'(P)P''(\lambda) \equiv \{\phi_k(\lambda)Q_1(\lambda)\}''.$$

Substitute $\beta_{k\tau}$ for λ in this identity; then the right member is zero and because of (11) and (12)

$$\psi_1''(\alpha_\tau) = 0.$$

This equation together with (10) and (12) permits us to conclude that $(\lambda-\alpha_\tau)^3$ is a factor of $\psi_1(\lambda)$, if $\epsilon_\tau>2$. To show that this procedure may be continued step by step to justify the conclusion that $\psi_1(\lambda)$ has the factor $(\lambda-\alpha_\tau)^{\epsilon_\tau}$, $1\leqq\tau\leqq s$, we assume that rth derivatives of the members of (9) with respect to λ satisfy the identity

$$\begin{aligned}\psi_1^{(r)}(P)P'^r(\lambda) + \psi_1^{(r-1)}(P)R_r(\lambda) + \psi_1^{(r-2)}(P)S_r(\lambda) + \cdots \\ + \psi_1''(P)U_r(\lambda) + \psi_1'(P)P^{(r)}(\lambda) \equiv \{\phi_k(\lambda)Q_1(\lambda)\}^{(r)},\end{aligned} \tag{13}$$

in which the leading term of the left member is $\psi_1^{(r)}(P)\,P'^r(\lambda)$ and the remaining terms are in $\psi_1^{(r-1)}(P)$, $\psi_1^{(r-2)}(P)$, $\cdots$, $\psi_1'(P)$ multiplied by polynomials in λ. This formula holds for $r=1$ and for $r=2$; to show that it holds in general we differentiate its members with respect to λ. This gives us

$$\begin{aligned}\psi_1^{(r+1)}(P)P'^{r+1}(\lambda) &+ \psi_1^{(r)}(P)\left[rP'^{r-1}(\lambda)P''(\lambda) + P'(\lambda)R_r(\lambda)\right] \\ &+ \psi_1^{(r-1)}(P)\left[R_r'(\lambda) + P'(\lambda)S_r(\lambda)\right] + \cdots \\ &+ \psi_1''(P)\left[U_r'(\lambda) + P'(\lambda)P^{(r)}(\lambda)\right] + \psi_1'(P)P^{(r)}(\lambda) \\ &\equiv \{\phi_k(\lambda)Q_1(\lambda)\}^{(r+1)},\end{aligned}$$

which is again of the general form (13) with r replaced by $r+1$. That formula is therefore valid.

Now we assume we have shown by successive steps that

$$\psi_1(\alpha_\tau) = \psi_1'(\alpha_\tau) = \psi_1''(\alpha_\tau) = \cdots = \psi_1^{(r-1)}(\alpha_\tau) = 0,$$

where $r \leqq \epsilon_\tau - 1$; then letting $\lambda = \beta_{k\tau}$ in (13), the right member will be zero for $(\lambda - \beta_{k\tau})^{\epsilon_\tau}$ is a factor of $\phi_k(\lambda)$ and because of the relations just written and because of (11), we must likewise have

$$\psi_1^{(r)}(\alpha_\tau) = 0.$$

Then $(\lambda - \alpha_\tau)^{r+1}$ must be a factor of $\psi_1(\lambda)$, if $(\lambda - \alpha_\tau)^r$ is and if $(\lambda - \beta_{k\tau})^{\epsilon_\tau}$, $\epsilon_\tau \geqq r+1$, is a factor of $\phi_k(\lambda)$. We are therefore permitted to conclude that $\psi_1(\lambda)$ has the factors $(\lambda - \alpha_\tau)^{\epsilon_\tau}$, $1 \leqq \tau \leqq s$, and because $\alpha_1, \alpha_2, \cdots, \alpha_s$ are distinct, that $\psi_1(\lambda)$ is divisible by $\prod_{j=1}^{s} (\lambda - \alpha_j)^{\epsilon_j}$, that is, by $\psi(\lambda)$. Furthermore, no polynomial $\psi_1(\lambda)$ of lower degree than m, the degree of $\psi(\lambda)$, exists which can satisfy an identity of the form given by (9); and any $\psi_1(\lambda)$ of degree m which satisfies identity (9) must be of the form $\psi_1(\lambda) \equiv c\psi(\lambda)$, where c is an arbitrary constant not zero, and is therefore uniquely determined save for a constant factor. We are now ready to proceed with the proof of the present theorem.

Let $\beta_{k\tau}$ be any root of $\phi_k(\lambda) = 0$; dividing $[P(\lambda) - \alpha_\tau]^\sigma$ by $\phi_k(\lambda)$, we find for every positive integer σ

$$[P(\lambda) - \alpha_\tau]^\sigma \equiv \phi_k(\lambda) Q_\sigma(\lambda) + (\lambda - \beta_{k\tau}) R_\sigma(\lambda), \tag{14}$$

where $Q_\sigma(\lambda)$ and $R_\sigma(\lambda)$ are polynomials in λ, and the degree of $R_\sigma(\lambda)$ in λ will not exceed $m-2$; $\lambda - \beta_{k\tau}$ is a factor of the remainder because it is a factor of both $P(\lambda) - \alpha_\tau$ and $\phi_k(\lambda)$. For $\sigma = m-1$, the left side of (14) is a polynomial of degree $p(m-1)$. Since we supposed $p > 1$ and $m > 1$, it follows that $p(m-1) \geqq m$, the degree of $\phi_k(\lambda)$, so that

$$Q_{m-1}(\lambda) \not\equiv 0.$$

On the basis of (14) we form the sum

$$\sum_{\sigma=1}^{m-1} t_\sigma [P(\lambda) - \alpha_\tau]^\sigma \equiv \phi_k(\lambda) \sum_{\sigma=1}^{m-1} t_\sigma Q_\sigma(\lambda) + (\lambda - \beta_{k\tau}) \sum_{\sigma=1}^{m-1} t_\sigma R_\sigma(\lambda), \tag{15}$$

where t_σ, $\sigma = 1, 2, \cdots, m-1$, are arbitrary constants. If it were possible to choose $t_1, t_2, \cdots, t_{m-1}$ not all zero so that

$$\sum_{\sigma=1}^{m-1} t_\sigma R_\sigma(\lambda) \equiv 0, \tag{16}$$

we would have for these values of $t_1, t_2, \cdots, t_{m-1}$

$$\sum_{\sigma=1}^{m-1} t_\sigma [P(\lambda) - \alpha_\tau]^\sigma \equiv \phi_k(\lambda) \sum_{\sigma=1}^{m-1} t_\sigma Q_\sigma(\lambda),$$

so that $\sum_{\sigma=1}^{m-1} t_\sigma(\lambda-\alpha_\tau)^\sigma$ would be a polynomial of degree less than m in λ satisfying the identity (9). Since no such polynomial can satisfy that condition, (16) is impossible, and $R_\sigma(\lambda)$, $\sigma=1, 2, \cdots, m-1$, are linearly independent. This fact permits us to define the constants $t_1, t_2, \cdots, t_{m-1}$ so that

$$\sum_{\sigma=1}^{m-1} t_\sigma R_\sigma(\lambda) \equiv 1. \tag{17}$$

If we suppose that

$$R_\sigma(\lambda) \equiv r_{1\sigma}\lambda^{m-2} + r_{2\sigma}\lambda^{m-3} + \cdots + r_{m-1,\sigma},$$

then t_σ, $\sigma=1, 2, \cdots, m-1$, must satisfy the $m-1$ non-homogeneou equations

$$\sum_{\sigma=1}^{m-1} r_{i\sigma} t_\sigma = 0 \qquad (i = 1,2, \cdots, m-2),$$

$$\sum_{\sigma=1}^{m-1} r_{m-1,\sigma} t_\sigma = 1.$$

The determinant of the coefficients of this system of equations is

$$\begin{vmatrix} r_{11} & r_{12} & \cdots & r_{1\,m-1} \\ r_{21} & r_{22} & \cdots & r_{2\,m-1} \\ \cdots & \cdots & \cdots & \cdots \\ r_{m-1\,1} & r_{m-1\,2} & \cdots & r_{m-1\,m-1} \end{vmatrix},$$

and is not zero, for $R_1(\lambda)$, $R_2(\lambda)$, $\cdots$, $R_{m-1}(\lambda)$ whose coefficients form the elements of the first, second, $\cdots$, and last column of this determinant are linearly independent. Consequently the constants $t_1, t_2, \cdots, t_{m-1}$ may be uniquely determined so that (17) is satisfied, and (15) may be written as

$$\sum_{\sigma=1}^{m-1} t_\sigma [P(\lambda) - \alpha_\tau]^\sigma \equiv \phi_k(\lambda) \sum_{\sigma=1}^{m-1} t_\sigma Q_\sigma(\lambda) + \lambda - \beta_{k\tau}.$$

If we let

$$T_k(\lambda) \equiv \beta_{k\tau} + \sum_{\sigma=1}^{m-1} t_\sigma(\lambda - \alpha_\tau)^\sigma, \tag{18}$$

and

$$H_k(\lambda) \equiv -\sum_{\sigma=1}^{m-1} t_\sigma Q_\sigma(\lambda),$$

this identity becomes

$$H_k(\lambda)\phi_k(\lambda) \equiv \lambda - T_k(P(\lambda)), \tag{19}$$

where $T_k(\lambda)$ is of degree lower than m in λ.

Finally, by (3),

$$P(\lambda) - \alpha_r \equiv \prod_{i=1}^{p}(\lambda - \beta_{ir}),$$

and if here we replace λ in the right member by $H_k(\lambda)\phi_k(\lambda)+T_k(P)$ according to (19), we obtain the identity

$$P(\lambda) - \alpha_r \equiv P(T_k(P)) - \alpha_r + K_k(\lambda)\phi_k(\lambda),$$

or

$$K_k(\lambda)\phi_k(\lambda) \equiv P(\lambda) - P(T_k(P)),$$

where $K_k(\lambda)\phi_k(\lambda)$ is the aggregate of all terms of the product that contain the factor $\phi_k(\lambda)$ explicitly. This identity is of the form (9), where $\lambda - P(T_k(\lambda))$ replaces $\psi_1(\lambda)$; $\lambda - P(T_k(\lambda))$ is consequently divisible by $\psi(\lambda)$, so that we may write

$$Z_k(\lambda)\psi(\lambda) \equiv \lambda - P(T_k(\lambda)), \tag{20}$$

where $Z_k(\lambda)$ is a polynomial in λ. This completes the proof of the present theorem.

III. Solution of the equation $P(X)=A$

THEOREM III. *If $\psi(\lambda)$ is a polynomial of degree $m>1$ in λ, and the distinct roots of $\psi(\lambda)=0$ are α_j, $j=1, 2, \cdots, s$; if $P(\lambda)$ is a polynomial of degree $p>1$ in λ whose leading coefficient is unity and whose constant term is zero; if the equation $P(\lambda)-\alpha_j=0$, $j=1, 2, \cdots, s$, has at least one simple root for every α_j which is a multiple root of $\psi(\lambda)=0$; and if*

$$\psi(A) = 0,$$

where A is a square matrix of order n; then there exists at least one matrix X also of order n such that

$$P(X) = A,$$

*and such that X is expressible as a polynomial in A with scalar coefficients.**

* The theorem holds as well for $m=1$ and for $p=1$, but in either case the results would be trivial and do not merit separate treatment.

Under the same hypotheses as those here stated, we have proved in Theorem II that a polynomial $T_k(\lambda)$ exists such that

$$\lambda - P(T_k(\lambda)) \equiv \psi(\lambda)Z_k(\lambda), \tag{20}$$

where $Z_k(\lambda)$ is a polynomial in λ. According to the present theorem we have $\psi(A)=0$, and since powers of a single matrix and their products with scalars obey the commutative and distributive laws of ordinary algebra, it is evident that if λ above be replaced by A and λ^0 by I, we have

$$A - P(T_k(A)) = 0,$$

or

$$P(T_k(A)) = A.$$

Therefore the equation

$$P(X) = A$$

has a solution which is given by

$$X_k = T_k(A), \tag{21}$$

where $T_k(\lambda)$ is a polynomial in λ with scalar coefficients. This completes the proof of our theorem.*

The existence of X is not only proved in this theorem but its form in terms of A is explicitly given by (21). Indeed the solution X_k of the equation $P(X)=A$ corresponding to the polynomial $\phi_k(\lambda)$, which has no factor of the second degree in common with any polynomial $P(\lambda)-\alpha_j$, $j=1, 2, \cdots, s$, is distinct from that corresponding to any other such polynomial of the $\prod_{j=1}^{s}\mu_j$ that may have this property. (See page 583.) Before we prove the uniqueness of the solutions, we shall show that

$$\phi_k(X_k) = 0.$$

According to (5) and (19),

$$\begin{aligned}\phi_k(\lambda) &\equiv \prod_{j=1}^{s}[T_k(P) - \beta_{kj} + H_k(\lambda)\phi_k(\lambda)], \\ &\equiv \phi_k(T_k(P)) + L_k(\lambda)H_k(\lambda)\phi_k(\lambda),\end{aligned}$$

* If $\psi(\lambda)$ is the polynomial of lowest degree such that $\psi(A)=0$, then the conditions of the present theorem are also necessary for the existence of X expressible as a polynomial in A. The writer's attention was called to this fact by Professor Ingraham. The logic by which we may prove that the conditions are necessary is virtually given by Frobenius, *Über lineare Substitutionen und bilineare Formen*, Crelle's Journal, vol. 84 (1878), p. 13, if we regard his $g(r)$ as the known function, $f(r)$ as the unknown function, and $\psi(r)$ as the polynomial of lowest degree such that $\psi(A)=0$. In our notation these polynomials would be $P(\lambda)$, $T_k(\lambda)$, and $\psi(\lambda)$ respectively. The method leads to the construction of $T_k(\lambda)$ by means of solving for the coefficients of $T_k(\lambda)$ a linear system of equations with non-zero determinant.

where $L_k(\lambda)H_k(\lambda)\phi_k(\lambda)$ is the aggregate of all terms of the product that have $H_k(\lambda)\phi_k(\lambda)$ as a factor. This identity is again of the form (9); in the present case all terms save $\phi_k(T_k(P))$, which corresponds to $\psi_1(P(\lambda))$ of that identity, contain $\phi_k(\lambda)$ explicitly and therefore $\phi_k(T_k(\lambda))$ is divisible by $\psi(\lambda)$. We may therefore write the identity above in the form

$$\phi_k(\lambda) \equiv \psi(P)M_k(P) + L_k(\lambda)H_k(\lambda)\phi_k(\lambda),$$

where $M_k(\lambda)$ is a polynomial in λ. Now by (19) and (21)

$$H_k(X_k)\phi_k(X_k) = X_k - T_k(A) = 0,$$

and since $\psi(A)=0$, then by substituting X_k for λ and I for λ^0 in the identity established above, we have

$$\phi_k(X_k) = 0,$$

for X_k is a solution of the equation $P(X)=A$. Consequently we have shown that $\phi_k(\lambda)$ is a polynomial that is satisfied by the matrix X_k obtained by the above method.

Thus far no restriction upon the polynomial $\psi(\lambda)$, save that $\psi(A)=0$, has been made. However, in order to prove the uniqueness of the solution obtained for each polynomial $\phi_k(\lambda)$ of the permitted class, we assume that $\psi(\lambda)$ of degree m is the polynomial of lowest degree satisfied by A. Suppose further that $\phi_k(\lambda)$ and $\phi_l(\lambda)$ are two distinct polynomials of the $\prod_{j=1}^{s}\mu_j$ that have no quadratic factor in common with any $P(\lambda)-\alpha_j$, $j=1, 2, \cdots, s$, and that determine the same solution X for the equation $P(X)=A$. Then we would have

$$\phi_k(X) = 0, \quad \phi_l(X) = 0;$$

two equations of degree m satisfied by X. Consequently it will be possible to determine constants c_1 and c_2 so that

$$F(\lambda) \equiv c_1\phi_k(\lambda) + c_2\phi_l(\lambda),$$

where $F(\lambda)$ will be a polynomial not identically zero and of degree $m'<m$, since $\phi_k(\lambda)$ and $\phi_l(\lambda)$ are distinct. But substituting X for λ and I for λ^0 we have $F(X)=0$. Then X satisfies a polynomial of degree m'. Suppose

$$F(\lambda) \equiv \prod_{j=1}^{m'}(\lambda - \gamma_i)$$

and that

$$P(\gamma_j) = \delta_j \qquad (j = 1,2, \cdots, m'),$$

where neither the γ_j nor the δ_j are necessarily distinct. Let

$$\Psi(\lambda) \equiv \prod_{j=1}^{m'} (\lambda - \delta_j) ;$$

now according to the definition of δ_j, each polynomial $P(\lambda)-\delta_j$ will have $\lambda-\gamma_j$ as a factor and the product of the m' polynomials $P(\lambda)-\delta_j$, $j=1, \cdots, m'$, will be divisible by $F(\lambda)$, i.e.,

$$\Psi(P) \equiv \prod_{j=1}^{m'} (P(\lambda) - \delta_j),$$

$$\equiv F(\lambda)N(\lambda),$$

where $N(\lambda)$ is a polynomial in λ. Now making the usual substitution X for λ we have

$$\Psi(A) = F(X)N(X) = 0,$$

since $P(X)=A$ and $F(X)=0$. But $\Psi(\lambda)$ is a polynomial of degree m', whereas the polynomial $\psi(\lambda)$ of degree m was assumed as that of lowest degree which vanished for A, so that the above equation is impossible. Hence the solutions X_k and X_l corresponding to distinct polynomials $\phi_k(\lambda)$ and $\phi_l(\lambda)$ are distinct.

The number of distinct solutions of the equation $P(X)=A$, expressible as polynomials in A, is therefore given by $\prod_{j=1}^{s} \mu_j$, where s is the number of distinct roots of the equation $\psi(\lambda)=0$, where $\psi(\lambda)$ is the polynomial of lowest degree for which $\psi(A)=0$, and where μ_j is the number of distinct roots of $P(\lambda)-\alpha_j=0$, when α_j is a simple root of $\psi(\lambda)=0$, and the number of simple roots of $P(\lambda)-\alpha_j=0$, when α_j is a multiple root of $\psi(\lambda)=0$.

Though the solution obtained by the method developed above is not restricted by the condition that $\psi(\lambda)$ is the polynomial of lowest degree for which $\psi(A)=0$, there is clearly no advantage in employing a polynomial of higher degree since such a course would only increase the labor involved in solving a particular example and may not give all the solutions that are possible. This last statement becomes evident if we recall that it is entirely possible in certain cases for α_r to be a simple root of $\psi(\lambda)=0$, where $\psi(\lambda)$ is of lower degree than n, the order of A, and where $\psi(A)=0$, but for the characteristic equation of A to have α_r as a multiple root. If in such a case $P(\lambda)-\alpha_r=0$ have multiple roots the characteristic equation would permit fewer solutions by the above method than would be possible on the basis of $\psi(\lambda)=0$.

The equation $X^p = A$ has p^s solutions when A is non-singular, for then all roots of the equation $\lambda^p = \alpha_j$ are distinct and $\mu_j = p$, $j = 1, 2, \cdots, s$. On the other hand, when A is singular and the equation $\psi(\lambda) = 0$, where $\psi(\lambda)$ is such that $\psi(A) = 0$, has a simple root $\alpha_1 = 0$, then the equation $X^p = A$ has p^{s-1} solutions of the kind here studied; for in this case the equation $\lambda^p = \alpha_1$ has only the root $\lambda = 0$, hence $\mu_1 = 1$, whereas the other $s-1$ equations, $\lambda^p - \alpha_j = 0$, $j = 2, 3, \cdots, s$, will each have p distinct solutions; if, however, $\alpha_1 = 0$ is a multiple root of $\psi(\lambda) = 0$, where $\psi(A) = 0$ is the equation of lowest degree satisfied by A, then since $\lambda^p - \alpha_1 = 0$ has no simple roots, we see that according to the definition of μ_j above, $\mu_1 = 0$, and thus our method would in this case give no solutions of the equation $X^p = A$.

The conclusion here reached that the equation $X^p = A$, $|A| = 0$, always has solutions if a polynomial $\psi(\lambda)$, such that $\psi(A) = 0$, exists which does not have λ^2 as a factor, contradicts a statement made by Cecioni.* This contradiction will be exhibited explicitly by means of a numerical example. (See Example 2 below.)

It may be shown that for a given matrix A, singular or non-singular, the coefficients $h_1, h_2 \cdots, h_{p-2}$ of the polynomial

$$P(\lambda) \equiv \lambda^p + h_1 \lambda^{p-1} + h_2 \lambda^{p-2} + \cdots + h_{p-2} \lambda^2 + h_{p-1} \lambda$$

may be taken entirely arbitrarily and h_{p-1} may be chosen in an infinity of ways so that the equation $P(X) = A$ will have solutions expressible as polynomials in the given matrix A.

The solution developed above for the equation $P(X) = A$ permits the following conclusion. If the elements of A belong to the algebraic field F which contains the coefficients of $P(\lambda)$ and if $\psi(\lambda)$ is the characteristic function or the polynomial of lowest degree such that $\psi(A) = 0$, then the solutions of the equation $P(X) = A$ that are expressible as polynomials in A belong to the field formed by the adjunction of the roots of $\psi(P(\lambda)) = 0$ to the field F. This conclusion follows from the fact that the coefficients of $\psi(\lambda)$ belong to the field that contains the elements of A, and that the coefficients of $\phi_k(\lambda)$ belong to the field containing the roots of $\psi(P(\lambda)) = 0$.

IV. Examples

Example 1. Given

$$P(X) \equiv X^2 - 3X = A,$$

* Cecioni, loc. cit., p. 85.

where

$$A = \begin{pmatrix} -3 & 3 & 6 \\ -\frac{7}{3} & 5 & 2 \\ -\frac{4}{3} & 4 & 0 \end{pmatrix};$$

X must be found satisfying this equation.

The polynomial $\psi(\lambda)$, such that $\psi(A)=0$, is in this case

$$\psi(\lambda) \equiv \lambda^3 - 2\lambda^2 - 8\lambda.$$

The roots of $\psi(\lambda)=0$ are 0, -2, and 4. The table giving the roots of $P(\lambda)-\alpha_j=0, j=1, 2, 3$, is

$$\begin{array}{lll} P(\lambda) = 0, & P(\lambda) + 2 = 0, & P(\lambda) - 4 = 0, \\ 0, & 1, & -1, \\ 3, & 2, & 4. \end{array}$$

$\phi_k(\lambda)$ may be formed in the following 2^3 ways:

$$\begin{aligned} \phi_1(\lambda) &\equiv \lambda(\lambda-1)(\lambda+1), & \phi_2(\lambda) &\equiv (\lambda-3)(\lambda-2)(\lambda-4), \\ \phi_1'(\lambda) &\equiv \lambda(\lambda-1)(\lambda-4), & \phi_2'(\lambda) &\equiv (\lambda-3)(\lambda-2)(\lambda+1), \\ \phi_1''(\lambda) &\equiv \lambda(\lambda-2)(\lambda+4), & \phi_2''(\lambda) &\equiv (\lambda-3)(\lambda-1)(\lambda+1), \\ \phi_1'''(\lambda) &\equiv \lambda(\lambda-2)(\lambda+1), & \phi_2'''(\lambda) &\equiv (\lambda-3)(\lambda-1)(\lambda-4). \end{aligned}$$

Each of the polynomials above is such that it has no quadratic factor in common with any polynomial $P(\lambda)$, $P(\lambda)+2$, $P(\lambda)-4$, and consequently for each there exists a unique matrix X such that $P(X)=A$ and $\phi_i^{(r)}(X)=0$, $i=1, 2, r=0, 1, 2, 3$. The solutions may be obtained directly from the pairs of equations $P(X)=A$ and $\phi_i^{(r)}(X)=0$, and this process would be entirely legitimate after the results above. We will, however, for the present develop the solution that satisfies $\phi_2(\lambda)$ to illustrate the theory as developed above.

In this case $\phi_2(\lambda)\equiv\lambda^3-9\lambda^2+26\lambda-24$, and according to (16), where in the present case $P(\lambda)-\alpha\equiv\lambda^2-3\lambda+2$, we have

$$\begin{aligned} P(\lambda) + 2 &\equiv (\lambda-2)(\lambda-1), \\ [P(\lambda)+2]^2 &\equiv \phi_2(\lambda)[\lambda+3] + (\lambda-2)(14\lambda-38). \end{aligned}$$

Then

$$\sum_{\sigma=1}^{2} t_\sigma R_\sigma(\lambda) \equiv t_1(\lambda-1) + t_2(14\lambda-38) \equiv 1,$$

and consequently

$$t_1 = 14/24, \quad t_2 = -1/24.$$

Then

$$T_2(\lambda) \equiv 2 + \frac{14}{24}(\lambda + 2) - \frac{1}{24}(\lambda + 2)^2,$$

$$\equiv -\frac{1}{24}(\lambda^2 - 10\lambda - 72),$$

and we consequently must have

$$X_2 = -\frac{1}{24}(A^2 - 10A - 72I)$$

as a solution of $P(X) = A$, according to the above theory. In the same way we get the solutions corresponding to each of the polynomials above. These are here tabulated in order:

$$X_1 = \frac{1}{24}(A^2 - 10A), \qquad X_2 = -\frac{1}{24}(A^2 - 10A - 72I),$$

$$X_1' = \frac{1}{4}A^2, \qquad X_2' = -\frac{1}{4}(A^2 - 12I),$$

$$X_1'' = \frac{1}{3}(A^2 - A), \qquad X_2'' = -\frac{1}{3}(A^2 - A - 9I),$$

$$X_1''' = \frac{1}{8}(A^2 - 6A), \qquad X_2''' = -\frac{1}{8}(A^2 - 6A - 24I),$$

where $X_i^{(r)}$ is the solution whose characteristic function in each case is $\phi_i^{(r)}(\lambda)$, $i = 1, 2$, and $r = 0, 1, 2, 3$.

EXAMPLE 2. We propose to find solutions of the equation $X^2 = A$, where

$$A = \begin{pmatrix} 1 & -8 & 0 & 8 \\ 1 & 1 & 0 & -1 \\ 4 & 0 & 4 & -4 \\ 1 & -8 & 0 & 8 \end{pmatrix}$$

is a singular matrix.

The characteristic function of A is a polynomial $\psi(\lambda)$ of lowest degree such that $\psi(A) = 0$. The roots of $\psi(\lambda) \equiv \lambda^4 - 14\lambda^3 + 49\lambda^2 - 36\lambda = 0$ are $\alpha_1 = 0$, $\alpha_2 = 1$, $\alpha_3 = 4$, $\alpha_4 = 9$. The roots of $\lambda^2 - \alpha_j = 0$, $j = 1, 2, 3, 4$, are

$$\begin{array}{cccc} \lambda^2 = 0, & \lambda^2 - 1 = 0, & \lambda^2 - 4 = 0, & \lambda^2 - 9 = 0, \\ 0, & 1, & 2, & 3, \\ 0, & -1, & -2, & -3; \end{array}$$

whence the following polynomials ϕ are formed:

$$\begin{array}{ll} \phi_1(\lambda) \equiv \lambda(\lambda-1)(\lambda-2)(\lambda-3), & \phi_2(\lambda) \equiv \lambda(\lambda+1)(\lambda+2)(\lambda+3), \\ \phi_1'(\lambda) \equiv \lambda(\lambda-1)(\lambda-2)(\lambda+3), & \phi_2'(\lambda) \equiv \lambda(\lambda+1)(\lambda+2)(\lambda-3), \\ \phi_1''(\lambda) \equiv \lambda(\lambda-1)(\lambda+2)(\lambda-3), & \phi_2''(\lambda) \equiv \lambda(\lambda+1)(\lambda-2)(\lambda+3), \\ \phi_1'''(\lambda) \equiv \lambda(\lambda+1)(\lambda-2)(\lambda-3), & \phi_2'''(\lambda) \equiv \lambda(\lambda-1)(\lambda+2)(\lambda+3). \end{array}$$

Each one of these polynomials is such that it has not a factor of the second degree in common with any $\lambda^2-\alpha_j, j=1, 2, 3, 4$. Here we have, according to our method,

$$\begin{aligned} \lambda^2 &\equiv \lambda^2, \\ \lambda^4 &\equiv \phi_1(\lambda) + \lambda(6\lambda^2 - 11\lambda + 6), \\ \lambda^6 &\equiv \phi_1(\lambda)[\lambda^2 + 6\lambda + 25] + \lambda(90\lambda^2 - 239\lambda + 150). \end{aligned}$$

Then

$$\sum_{\sigma=1}^{3} t_\sigma R_\sigma(\lambda) \equiv t_1\lambda + t_2(6\lambda^2 - 11\lambda + 6) + t_3(90\lambda^2 - 239\lambda + 150) \equiv 1,$$

and

$$t_1 = 74/60, \quad t_2 = -15/60, \quad t_3 = 1/60.$$

Then

$$T_1(\lambda) \equiv \frac{1}{60}(\lambda^3 - 15\lambda^2 + 74\lambda),$$

or

$$X_1 = \frac{1}{60}(A^3 - 15A^2 + 74A).$$

X_1 is a solution of $\phi_1(X)=0$ and $X^2=A$, and its negative is a solution of $\phi_2(X)=0$ and $X^2=A$. In this manner we get as the remaining solutions:

$$X_1' = \mp\frac{1}{6}(A^2 - 7A),$$

$$X_1'' = \mp\frac{1}{12}(A^3 - 11A^2 + 22A),$$

$$X_1''' = \mp\frac{1}{30}(2A^3 - 25A^2 + 53A),$$

where the upper sign is that of the solution obtained for $\phi_1^{(r)}(X)=0$ and $X^2=A$ and the lower of that for $\phi_2^{(r)}(X)=0$ and $X^2=A$. These solutions are, respectively,

$$X_1 = -X_2 = \begin{pmatrix} 1 & -2 & 0 & 2 \\ 1 & 1 & 0 & -1 \\ 2 & 0 & 2 & -2 \\ 1 & -2 & 0 & 2 \end{pmatrix}, \quad X_1' = -X_2' = \begin{pmatrix} 1 & 4 & 0 & -4 \\ 1 & 1 & 0 & -1 \\ 2 & 0 & 2 & -2 \\ 1 & 4 & 0 & -4 \end{pmatrix},$$

$$X_1'' = -X_2'' = \begin{pmatrix} 1 & -2 & 0 & 2 \\ 1 & 1 & 0 & -1 \\ -2 & 0 & -2 & 2 \\ 1 & -2 & 0 & 2 \end{pmatrix}, \quad X_1''' = -X_2''' = \begin{pmatrix} -1 & -4 & 0 & 4 \\ -1 & -1 & 0 & 1 \\ 2 & 0 & 2 & -2 \\ -1 & -4 & 0 & 4 \end{pmatrix}.$$

Thus the conclusion reached in the preceding section, that the equation $X^p=A$ may have solutions expressible as polynomials in A even when $|A|=0$, is verified by a particular example when $p=2$.

UNIVERSITY OF WISCONSIN,
MADISON, WIS.