CHAPTER II

ALGEBRAIC OPERATIONS WITH MATRICES. THE CHARACTERISTIC EQUATION

2.01 **Identities.** The following elementary considerations enable us to carry over a number of results of ordinary scalar algebra into the algebra of matrices. Suppose $f(\lambda_1, \lambda_2, \cdots, \lambda_r)$, $g(\lambda_1, \lambda_2, \cdots, \lambda_r)$ are integral algebraic functions of the scalar variables λ_i with scalar coefficients, and suppose that

$$f(\lambda_1, \lambda_2, \cdots, \lambda_r) = g(\lambda_1, \lambda_2, \cdots, \lambda_r)$$

is an algebraic identity; then, when $f(\lambda_1, \cdots, \lambda_r) - g(\lambda_1, \cdots, \lambda_r)$ is reduced to the standard form of a polynomial, the coefficients of the various powers of the λ's are zero. In carrying out this reduction no properties of the λ's are used other than those which state that they obey the laws of scalar multiplication and addition: if then we replace $\lambda_1, \lambda_2, \cdots, \lambda_r$ by *commutative* matrices $x_1, x_2, \cdots, x_r$, the reduction to the form 0 is still valid step by step and hence

$$f(x_1, x_2, \cdots, x_r) = g(x_1, x_2, \cdots, x_r).$$

An elementary example of this is

$$(1 - x^2) = (1 - x)(1 + x)$$

or, when $xy = yx$,

$$x^2 - y^2 = (x - y)(x + y).$$

Here, if $xy \neq yx$, the reader should notice that the analogue of the algebraic identity becomes

$$x^2 - y^2 = x(x + y) - (x + y)y,$$

which may also be written $x^2 - y^2 = (x - y)(x + y) + (yx - xy)$.

2.02 **Matric polynomials in a scalar variable.** By a matric polynomial in a scalar variable λ is meant a matrix that can be expressed in the form

$$P(\lambda) = p_0\lambda^r + p_1\lambda^{r+1} + \cdots + p_r \qquad (p_0 \neq 0), \tag{1}$$

where $p_0, p_1, \ldots, p_r$ are constant matrices. The coordinates of $P(\lambda)$ are scalar polynomials in λ and hence, if

$$Q(\lambda) = q_0\lambda^s + q_1\lambda^{s-1} + \cdots + q_s \qquad (q_0 \neq 0) \tag{2}$$

is also a matric polynomial, $P(\lambda) \equiv Q(\lambda)$ if, and only if, $r = s$ and the coefficients of corresponding powers of λ are equal, that is, $p_i = q_i$ $(i = 1, 2, \ldots, r)$. If $|q_0| \neq 0$, the degree of the product $P(\lambda)Q(\lambda)$ (or $Q(\lambda)P(\lambda)$) is exactly $r + s$ since the coefficient of the highest power λ^{r+s} which occurs in the product is p_0q_0

(or q_0p_0) which cannot be 0 if $p_0 \neq 0$ and $|q_0| \neq 0$. If, however, both $|p_0|$ and $|q_0|$ are 0, the degree of the product may well be less than $r + s$, as is seen from the examples

$$(e_{11}\lambda + 1)(e_{22}\lambda + 1) = e_{11}e_{22}\lambda^2 + (e_{11} + e_{22})\lambda + 1 = (e_{11} + e_{22})\lambda + 1,$$

$$\begin{Vmatrix} \lambda & 1 \\ \lambda & 1 \end{Vmatrix} \begin{Vmatrix} 1 & -1 \\ -\lambda & \lambda \end{Vmatrix} = 0.$$

Another noteworthy difference between matric and scalar polynomials is that, when the determinant of a matric polynomial is a constant different from 0, its inverse is also a matric polynomial: for instance

$$(e_{12}\lambda + 1)^{-1} = -e_{12}\lambda + 1,$$
$$[(e_{12} + e_{23})\lambda + 1]^{-1} = e_{13}\lambda^2 - (e_{12} + e_{23})\lambda + 1.$$

We shall call such polynomials *elementary polynomials.*

2.03 **The division transformation.** The greater part of the theory of the division transformation can be extended from ordinary algebra to the algebra of matrices; the main precaution that must be taken is that it must not be assumed that every element of the algebra has an inverse and that due allowance must be made for the peculiarities introduced by the lack of commutativity in multiplication.

THEOREM 1. *If $P(\lambda)$ and $Q(\lambda)$ are the polynomials defined by* (1) *and* (2), *and if $|q_0| \neq 0$, there exist unique polynomials $S(\lambda)$, $R(\lambda)$, $S_1(\lambda)$, $R_1(\lambda)$, of which S and S_1 if not zero, are of degree $r - s$ and the degrees of R and R_1 are $s - 1$ at most, such that*

$$P(\lambda) \equiv S(\lambda)Q(\lambda) + R(\lambda) \equiv Q(\lambda)S_1(\lambda) + R_1(\lambda).$$

If $r < s$, we may take $S_1 = S = 0$ and $R_1 = R = P$; in so far as the existence of these polynomials is concerned the theorem is therefore true in this case. We shall now assume as a basis for a proof by induction that the theorem is true for polynomials of degree less than r and that $r \leq s$. Since $|q_0| \neq 0$, q_0^{-1} exists and, as in ordinary scalar division, we have

$$P(\lambda) - p_0q_0^{-1}\lambda^{r-s}Q(\lambda) = (p_1 - p_0q_0^{-1}q_1)\lambda^{r-1} + \cdots = P_1(\lambda).$$

Since the degree of P_1 is less than r, we have by hypothesis $P_1(\lambda) = P_2(\lambda)Q(\lambda) + R(\lambda)$, the degrees of P_2 and R being less, respectively, than $r - s$ and s; hence

$$P(\lambda) = (p_0q_0^{-1}\lambda^{r-s} + P_2(\lambda))Q(\lambda) + R(\lambda) = S(\lambda)Q(\lambda) + R(\lambda)$$

as required by the theorem. The existence of the right hand quotient and remainder follows in the same way.

It remains to prove the uniqueness of S and R. Suppose, if possible, that $P = SQ + R = TQ + U$ where R and S are as above and T, U are poly-

nomials the degree of U being less than s; then $(S - T)Q = U - R$. If $S - T \neq 0$, then, since $|q_0| \neq 0$, the degree of the polynomial $(S - T)Q$ is at least as great as that of Q and is therefore greater than the degree of $U - R$. It follows immediately that $S - T = 0$, and hence also $U - R = 0$; which completes the proof of the theorem.

If Q is a scalar polynomial, that is, if its coefficients q are scalars, then $S = S_1$, $R = R_1$; and, if the division is exact, then $Q(\lambda)$ is a factor of each of the coordinates of $P(\lambda)$.

THEOREM 2. *If the matric polynomial* (1) *is divided on the right by* $\lambda - a$, *the remainder is*

$$p_0a^r + p_1a^{r-1} + \cdots + p_r$$

and, if it is divided on the left, the remainder is

$$a^rp_0 + a^{r-1}p_1 + \cdots + p_r.$$

As in ordinary algebra the proof follows immediately from the identity

$$\lambda^s - a^s = (\lambda - a)(\lambda^{s-1} + \lambda^{s-2}a + \cdots + a^{s-1})$$

in which the order of the factors is immaterial since λ is a scalar.

If $P(\lambda)$ is a scalar polynomial, the right and left remainders are the same and are conveniently denoted by $P(a)$.

2.04 Theorem 1 of the preceding section holds true as regards the existence of S, S_1, R, R_1, and the degree of R, R_1 even when $|q_0| = 0$ provided $|Q(\lambda)| \neq 0$. Suppose the rank of q_0 is $t < n$; then by §1.10 it has the form $\sum_1^t \alpha_i S \beta_i$ or, say, $h\left(\sum_1^t e_{ii}\right)k$ where h and k are non-singular matrices for which $he_i = \alpha_i$, $k'e_i = \beta_i$ $(i = 1, 2, \cdots, t)$. If $c_1 = \sum_{t+1}^n e_{ii}$, then

$$Q_1 = (c_1\lambda + 1)h^{-1}Q \tag{3}$$

is a polynomial whose degree is not higher than the degree s of Q since $c_1h^{-1}q_0 = 0$ so that the term in λ^{s+1} is absent. Now, if $\eta = |h^{-1}|$, then

$$|Q_1| = |c_1\lambda + 1|\,|h^{-1}|\,|Q| = (1 + \lambda)^{n-t}\eta\,|Q|,$$

so that the degree of $|Q_1|$ is greater than that of $|Q|$ by $n - t$. If the leading coefficient of Q_1 is singular, this process may be repeated, and so on, giving $Q_1, Q_2, \cdots$, where the degree of $|Q_i|$ is greater than that of $|Q_{i-1}|$. But the degree of each Q_i is less than or equal to s and the degree of the determinant of a polynomial of the sth degree cannot exceed ns. Hence at some stage the leading coefficient of, say, Q_j is not singular and, from the law of formation (3) of the successive Q's, we have $Q_j(\lambda) = H(\lambda)Q(\lambda)$, where $H(\lambda)$ is a matric polynomial.

By Theorem 1, Q_j taking the place of Q, we can find S^* and R, the latter of degree $s - 1$ at most, such that

$$P(\lambda) = S^*(\lambda)H(\lambda)Q(\lambda) + R(\lambda) = S(\lambda)Q(\lambda) + R(\lambda).$$

The theorem is therefore true even if $|q_0| = 0$ except that the quotient and remainder are not necessarily unique and the degree of S may be greater than $r - s$, as is shown by taking $P = \lambda^2 - 1$, $Q = e_{11}\lambda + 1$, when we have

$$P = (e_{22}\lambda^2 + e_{11}\lambda - 1)Q = (e_{22}\lambda^2 + e_{11}\lambda - 1 + e_{12})Q - e_{12}.$$

2.05 **The characteristic equation.** If x is a matrix, the scalar polynomial

$$f(\lambda) = |\lambda - x| = \lambda^n + a_1\lambda^{n-1} + \cdots + a_n \tag{4}$$

is called the *characteristic function* corresponding to x. We have already seen (§1.05 (15)) that the product of a matrix and its adjoint equals its determinant; hence

$$(\lambda - x)\ \text{adj}\ (\lambda - x) = |\lambda - x| = f(\lambda).$$

It follows that the polynomial $f(\lambda)$ is exactly divisible by $\lambda - x$ so that by the remainder theorem (§2.03, Theorem 2)

$$f(x) = 0. \tag{5}$$

As a simple example of this we may take $x = \begin{Vmatrix} \alpha & \beta \\ \gamma & \delta \end{Vmatrix}$. Here

$$f(\lambda) = (\lambda - \alpha)(\lambda - \delta) - \beta\gamma = \lambda^2 - (\alpha + \delta)\lambda + \alpha\delta - \beta\gamma,$$

and

$$f(x) = \begin{Vmatrix} \alpha^2 + \beta\gamma & \alpha\beta + \beta\delta \\ \gamma\alpha + \delta\gamma & \gamma\beta + \delta^2 \end{Vmatrix} - (\alpha + \delta)\begin{Vmatrix} \alpha & \beta \\ \gamma & \delta \end{Vmatrix} + (\alpha\delta - \beta\gamma)\begin{Vmatrix} 1 & 0 \\ 0 & 1 \end{Vmatrix} = 0.$$

The following theorem is an important extension of this result.

THEOREM 3. *If* $f(\lambda) = |\lambda - x|$ *and* $\theta(\lambda)$ *is the highest common factor of the first minors of* $|\lambda - x|$, *and if*

$$\varphi(\lambda) = f(\lambda)/\theta(\lambda), \tag{6}$$

the leading coefficient of $\theta(\lambda)$ *being* 1 (*and therefore also that of* $\varphi(\lambda)$), *then*

(i) $\varphi(x) = 0$;

(ii) *if* $\psi(\lambda)$ *is any scalar polynomial such that* $\psi(x) = 0$, *then* $\varphi(\lambda)$ *is a factor of* $\psi(\lambda)$, *that is,* $\varphi(\lambda)$ *is the scalar polynomial of lowest degree and with leading coefficient* 1 *such that* $\varphi(x) = 0$;

(iii) *every root of* $f(\lambda)$ *is a root of* $\varphi(\lambda)$.

The coordinates of $\text{adj}(\lambda - x)$ are the first minors of $|\lambda - x|$ and therefore by hypothesis $[\text{adj}(\lambda - x)]/\theta(\lambda)$ is integral; also

$$\frac{\text{adj}(\lambda - x)}{\theta(\lambda)}(\lambda - x) = \frac{f(\lambda)}{\theta(\lambda)} = \varphi(\lambda);$$

hence $\varphi(x) = 0$ by the remainder theorem.

If $\psi(\lambda)$ is any scalar polynomial for which $\psi(x) = 0$, we can find scalar polynomials $M(\lambda)$, $N(\lambda)$ such that $M(\lambda)\varphi(\lambda) + N(\lambda)\psi(\lambda) \equiv \zeta(\lambda)$, where $\zeta(\lambda)$ is the highest common factor of φ and ψ. Substituting x for λ in this scalar identity and using $\varphi(x) = 0 = \psi(x)$ we have $\zeta(x) = 0$; if, therefore, $\psi(x) = 0$ is a scalar equation of lowest degree satisfied by x, we must have $\psi(\lambda) = \zeta(\lambda)$, apart from a constant factor, so that $\psi(\lambda)$ is a factor of $\varphi(\lambda)$, say

$$\varphi(\lambda) = h(\lambda)\psi(\lambda). \tag{7}$$

Since $\psi(x) = 0$, $\lambda - x$ is a factor of $\psi(\lambda)$, say $\psi(\lambda) = (\lambda - x)g(\lambda)$, where g is a matric polynomial; hence

$$\psi(\lambda) = \frac{\varphi(\lambda)}{h(\lambda)} = \frac{f(\lambda)}{h(\lambda)\theta(\lambda)} = (\lambda - x)g(\lambda).$$

Hence

$$g(\lambda) = \frac{f(\lambda)}{\theta(\lambda)h(\lambda)(\lambda - x)} = \frac{\operatorname{adj}(\lambda - x)}{\theta(\lambda)h(\lambda)}$$

and this cannot be integral unless $h(\lambda)$ is a constant in view of the fact that $\theta(\lambda)$ is the highest common factor of the coordinates of $\operatorname{adj}(\lambda - x)$; it follows that $\psi(\lambda)$ differs from $\varphi(\lambda)$ by at most a constant factor.

A repetition of the first part of this argument shows that, if $\psi(x) = 0$ is any scalar equation satisfied by x, then $\varphi(\lambda)$ is a factor of $\psi(\lambda)$.

It remains to show that every root of $f(\lambda)$ is a root of $\varphi(\lambda)$. If λ_1 is any root of $f(\lambda) = |\lambda - x|$, then from $\varphi(\lambda) = g(\lambda)(\lambda - x)$ we have

$$\varphi(\lambda_1) = g(\lambda_1)(\lambda_1 - x)$$

so that the determinant, $[\varphi(\lambda_1)]^n$, of the scalar matrix $\varphi(\lambda_1)$ equals $|g(\lambda_1)|\,|\lambda_1 - x|$, which vanishes since $|\lambda_1 - x| = f(\lambda_1)$. This is only possible if $\varphi(\lambda_1) = 0$, that is, if every root of $f(\lambda)$ is also a root of $\varphi(\lambda)$.

The roots of $f(\lambda)$ are also called the *roots*[1] of x, $\varphi(\lambda)$ is called the *reduced characteristic function* of x, and $\varphi(x) = 0$ the *reduced equation* of x.

2.06 A few simple results are conveniently given at this point although they are for the most part merely particular cases of later theorems. If $g(\lambda)$ is a scalar polynomial, then on dividing by $\varphi(\lambda)$, whose degree we shall denote by ν, we may set $g(\lambda) = q(\lambda)\varphi(\lambda) + r(\lambda)$, where q and r are polynomials the degree of r being less than ν. Replacing λ by x in this identity and remembering that $\varphi(x) = 0$, we have[2] $g(x) = r(x)$, that is, any polynomial can be replaced by an equivalent polynomial of degree less than ν.

[1] They are also called the *latent* roots of x.

[2] If $g(\lambda)$ is a matric polynomial whose coefficients are not all commutative with x, the meaning of $g(x)$ is ambiguous; for instance, x may be placed on the right of the coefficients, or it may be put on the left. For such a polynomial we can say in general that it can be replaced by an equal polynomial in which no power of x higher than the $(\nu - 1)$th occurs.

If $g(\lambda)$ is a scalar polynomial which is a factor of $\varphi(\lambda)$, say $\varphi(\lambda) = h(\lambda)g(\lambda)$, then $0 = \varphi(x) = h(x)g(x)$. It follows that $|g(x)| = 0$; for if this were not so, we should have $h(x) = [g(x)]^{-1}\varphi(x) = 0$, whereas x can satisfy no scalar equation of lower degree than φ. Hence, if $g(\lambda)$ is a scalar polynomial which has a factor in common with $\varphi(x)$, then $g(x)$ is singular.

If a scalar polynomial $g(\lambda)$ has no factor in common with $\varphi(\lambda)$, there exist scalar polynomials $M(\lambda)$, $N(\lambda)$ such that $M(\lambda)g(\lambda) + N(\lambda)\varphi(\lambda) \equiv 1$. Hence $M(x)g(x) = 1$, or $[g(x)]^{-1} = M(x)$. It follows immediately that any finite rational function of x with scalar coefficients can be expressed as a scalar polynomial in x of degree $\nu - 1$ at most. It should be noticed carefully however that, if x is a variable matrix, the coefficients of the reduced polynomial will in general contain the variable coordinates of x and will not be integral in these unless the original function is integral. It follows also that $g(x)$ is singular only when $g(\lambda)$ has a factor in common with $\varphi(\lambda)$.

Finally we may notice here that similar matrices have the same reduced equation; for, if g is a scalar polynomial, $g(y^{-1}xy) = y^{-1}g(x)y$. As a particular case of this we have that xy and yx have the same reduced equation if, say, y is non-singular; for $xy = y^{-1}\cdot yx\cdot y$. If both x and y are singular, it can be shown[3] that xy and yx have the same characteristic equation, but not necessarily the same reduced equation as is seen from the example $x = e_{12}$, $y = e_{22}$.

2.07 **Matrices with distinct roots.** Because of its importance and comparative simplicity we shall investigate the form of a matrix all of whose roots are different before considering the general case. Let

$$f(\lambda) = |\lambda - x| = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n) \tag{8}$$

where no two roots are equal and set

$$f_i(\lambda) = \frac{(\lambda - \lambda_1)\cdots(\lambda - \lambda_{i-1})(\lambda - \lambda_{i+1})\cdots(\lambda - \lambda_n)}{(\lambda_i - \lambda_1)\cdots(\lambda_i - \lambda_{i-1})(\lambda_i - \lambda_{i+1})\cdots(\lambda_i - \lambda_n)} = \frac{f(\lambda)/f'(\lambda_i)}{\lambda - \lambda_i}. \tag{9}$$

By the Lagrange interpolation formula $\sum_i f_i(\lambda) \equiv 1$; hence

$$f_1(x) + f_2(x) + \cdots + f_n(x) = 1. \tag{10}$$

Further, $f(\lambda)$ is a factor of $f_i(\lambda)f_j(\lambda)$ $(i \neq j)$ so that

$$f_i(x)f_j(x) = 0 \qquad (i \neq j); \tag{11}$$

hence multiplying (10) by $f_i(x)$ and using (11) we have

$$[f_i(x)]^2 = f_i(x). \tag{12}$$

Again, $(\lambda - \lambda_i)f_i(\lambda) = f(\lambda)/f'(\lambda_i)$; hence $(x - \lambda_i)f_i(x) = 0$, that is,

$$xf_i(x) = \lambda_i f_i(x), \tag{13}$$

[3] For example, by replacing y by $y + \delta$, δ being a scalar, and considering the limiting case when δ approaches 0.

whence, summing with regard to i and using (10), we have

$$x = \lambda_1 f_1(x) + \lambda_2 f_2(x) + \cdots + \lambda_n f_n(x). \tag{14}$$

If we form x^r from (14), r being a positive integer, it is immediately seen from (11) and (12), or from the Lagrange interpolation formula, that

$$x^r = \lambda_1^r f_1 + \lambda_2^r f_2 + \cdots + \lambda_n^r f_n, \tag{15}$$

where f_i stands for $f_i(x)$, and it is easily verified by actual multiplication that, if no root is 0,

$$x^{-1} = \lambda_1^{-1} f_1 + \lambda_2^{-1} f_2 + \cdots + \lambda_n^{-1} f_n$$

so that (15) holds for negative powers also. The matrices f_i are linearly independent. For if $\Sigma \gamma_i f_i = 0$, then

$$0 = f_j \Sigma \gamma_i f_i = \gamma_j f_j^2 = \gamma_j f_j$$

whence every $\gamma_j = 0$ seeing that in the case we are considering $f(\lambda)$ is itself the reduced characteristic function so that $f_j(x) \neq 0$.

From these results we have that, if $g(\lambda)$ is any scalar rational function whose denominator has no factor in common with $\varphi(\lambda)$, then

$$g(x) = g(\lambda_1) f_1 + g(\lambda_2) f_2 + \cdots + g(\lambda_n) f_n. \tag{16}$$

It follows from this that the roots of $g(x)$ are $g(\lambda_i)$ $(i = 1, 2, \cdots, n)$. For setting $y = g(x)$, $\mu_i = g(\lambda_i)$, we have as above

$$\psi(y) = \Sigma \psi(\mu_i) f_i,$$

$\psi(\lambda)$ being a scalar polynomial. Now $\psi(y) f_i = \psi(\mu_i) f_i$; hence, if $\psi(y) = 0$, then also $\psi(\mu_i) = 0$ $(i = 1, 2, \cdots, n)$; and conversely. Hence if the notation is so chosen that $\mu_1, \mu_2, \cdots, \mu_r$ are the distinct values of μ_i, the reduced characteristic function of $y = g(x)$ is $\prod_1^r (\lambda - \mu_i)$.

2.08 If the determinant $|\lambda - x| = f(\lambda)$ is expanded in powers of λ, it is easily seen[4] that the coefficient a_r of λ^{n-r} is $(-1)^r$ times the sum of the principal minors of x of order r; this coefficient is therefore a homogeneous polynomial of degree r in the coordinates of x. In particular, $-a_1$ is the sum of the coordinates in the main diagonal: this sum is called the *trace* of x and is denoted by tr x.

If y is an arbitrary matrix, μ a scalar variable, and $z = x + \mu y$, the coefficients of the characteristic equation of z, say

$$z^n + b_1 z^{n-1} + \cdots + b_n = 0, \tag{17}$$

[4] For instance, by differentiating $|\lambda - x|$ $n - r$ times with respect to λ and then setting $\lambda = 0$.

are polynomials in μ of the form

$$b_s = a_{s0} + \mu a_{s1} + \cdots + \mu^s a_{ss}, \quad (a_{s0} = a_s, a_{00} = 1) \tag{18}$$

and the powers of z are also polynomials in μ, say

$$z^r = x^r + \mu \begin{Bmatrix} x & y \\ r-1 & 1 \end{Bmatrix} + \mu^2 \begin{Bmatrix} x & y \\ r-2 & 2 \end{Bmatrix} + \cdots + \mu^r y^r \tag{19}$$

where $\begin{Bmatrix} x & y \\ s & t \end{Bmatrix}$ is obtained by multiplying s x's and t y's together in every possible way and adding the terms so obtained, e.g.,

$$\begin{Bmatrix} x & y \\ 2 & 1 \end{Bmatrix} = x^2y + xyx + yx^2.$$

If we substitute (18) and (19) in (17) and arrange according to powers of μ, then, since μ is an independent variable, the coefficients of its several powers must be zero. This leads to a series of relations connecting x and y of the form

$$\sum_{i,j} a_{ij} \begin{Bmatrix} x & & y \\ n-s-i+j & & s-j \end{Bmatrix} = 0 \qquad (s = 0, 1, 2, \cdots) \tag{20}$$

where a_{ij} are the coefficients defined in (18) and $\begin{Bmatrix} x & & y \\ n-s-i+j & & s-j \end{Bmatrix}$ is replaced by 0 when $j > s$. In particular, if $s = 1$,

$$\begin{Bmatrix} x & y \\ n-1 & 1 \end{Bmatrix} + a_1 \begin{Bmatrix} x & y \\ n-2 & 1 \end{Bmatrix} + \cdots + a_{n-1}y + a_{11}x^{n-1} + \cdots + a_{n1} = 0$$

which, when $xy = yx$, becomes

$$f'(x)y = -(a_{11}x^{n-1} + \cdots + a_{n1}) = g(x).$$

When x has no repeated roots, $f'(\lambda)$ has no root in common with $f(\lambda)$ and $f'(x)$ has an inverse (cf. §2.06) so that $y = g(x)/f'(x)$ which can be expressed as a scalar polynomial in x; and conversely every such polynomial is commutative with x. We therefore have the following theorem:

THEOREM 4. *If x has no multiple roots, the only matrices commutative with it are scalar polynomials in x.*

2.09 Matrices with multiple roots. We shall now extend the main results of §2.07 to matrices whose roots are not necessarily simple. Suppose in the first place that x has only one distinct root and that its reduced characteristic function is $\varphi(\lambda) = (\lambda - \lambda_1)^\nu$, and set

$$\eta_1^i = \eta_i = (x - \lambda_1)^i = (x - \lambda_1)\eta_{i-1} \qquad (i = 1, 2, \cdots, \nu - 1);$$

then

$$\eta_1^\nu = 0, \quad x\eta_{\nu-1} = \lambda_1\eta_{\nu-1}, \quad x\eta_i = \lambda_1\eta_i + \eta_{i+1} \quad (i = 1, 2, \cdots, \nu - 2)$$

and

$$x^s = (\lambda_1 + \eta_1)^s = \lambda_1^s + s\lambda_1^{s-1}\eta_1 + \binom{s}{2}\lambda_1^{s-2}\eta_1^2 + \cdots$$

where the binomial expansion is cut short with the term $\eta_1^{\nu-1}$ since $\eta_1^\nu = 0$. Again, if $g(\lambda)$ is any scalar polynomial, then

$$g(x) = g(\lambda_1 + \eta_1) = g(\lambda_1) + g'(\lambda_1)\eta_1 + \cdots + \frac{g^{(\nu-1)}(\lambda_1)}{(\nu-1)!}\eta_1^{\nu-1}.$$

It follows immediately that, if $g^{(s)}(\lambda)$ is the first derivative of $g(\lambda)$ which is not 0 when $\lambda = \lambda_1$ and $(\kappa - 1)s < \nu \leq \kappa s$, then the reduced equation of $g(x)$ is

$$[g(x) - g(\lambda_1)]^\kappa = 0.$$

It should be noted that the first $\nu - 1$ powers of η_1 are linearly independent since $\varphi(\lambda)$ is the reduced characteristic function of x.

2.10 We shall now suppose that x has more than one root. Let the reduced characteristic function be

$$\varphi(\lambda) = \prod_{i=1}^{r} (\lambda - \lambda_i)^{\nu_i} \qquad (\Sigma\nu_i = \nu,\ r > 1) \tag{21}$$

and set

$$h_i(\lambda) = \varphi(\lambda)/(\lambda - \lambda_i)^{\nu_i}. \tag{22}$$

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

$$M_i(\lambda)h_i(\lambda) + (\lambda - \lambda_i)^{\nu_i}N_i(\lambda) \equiv 1, \qquad M_i(\lambda_i) \not\equiv 0.$$

If we set

$$\varphi_i(\lambda) = M_i(\lambda)h_i(\lambda), \tag{23}$$

then $1 - \Sigma\varphi_i(\lambda)$ is exactly divisible by $\varphi(\lambda)$ and, being of degree $\nu - 1$ at most, must be identically 0; hence

$$\sum_1^r \varphi_i(\lambda) = 1. \tag{24}$$

Again, from (22) and (23), $\varphi(\lambda)$ is a factor of $\varphi_i(\lambda)\varphi_j(\lambda)$ $(i \neq j)$ and hence on multiplying (24) by $\varphi_i(\lambda)$ we have

$$[\varphi_i(\lambda)]^2 \equiv \varphi_i(\lambda), \qquad \varphi_i(\lambda)\varphi_j(\lambda) \equiv 0, \bmod \varphi(\lambda) \qquad (i \neq j). \tag{25}$$

Further, if $g(\lambda)$ is a scalar polynomial, then

$$\begin{aligned} g(\lambda) &= \sum_1^r g(\lambda)\varphi_i(\lambda) \\ &= \sum_1^r \left[g(\lambda_i) + g'(\lambda_i)(\lambda - \lambda_i) + \cdots + \frac{g^{(\nu_i-1)}(\lambda_i)}{(\nu_i-1)!}(\lambda - \lambda_i)^{\nu_i-1}\right]\varphi_i(\lambda) + R \end{aligned} \tag{26}$$

where R has the form $\Sigma C_i(\lambda)(\lambda - \lambda_i)^{\nu_i}\varphi_i(\lambda)$, C_i being a polynomial, so that R vanishes when x is substituted for λ.

2.11 If we put x for λ in (23) and set φ_i for $\varphi_i(x)$, then (24) and (25) show that

$$\varphi_i^2 = \varphi_i, \quad \varphi_i\varphi_j = 0 \quad (i \neq j), \quad \sum_1^r \varphi_i = 1. \tag{27}$$

It follows as in §2.07 that the matrices φ_i are linearly independent and none is zero, since $\varphi_i(\lambda_i) \not\equiv 0$ so that $\varphi(\lambda)$ is not a factor of $\varphi_i(\lambda)$, which would be the case were $\varphi_i(x) = 0$. We now put x for λ in (26) and set

$$\eta_i = (x - \lambda_i)\varphi_i \qquad (i = 1, 2, \cdots, r). \tag{28}$$

Since the ν_ith power of $(\lambda - \lambda_i)\varphi_i(\lambda)$ is the first which has $\varphi(\lambda)$ as a factor, η_i is a nilpotent matrix of index ν_i (cf. §1.05) and, remembering that $\varphi_i^2 = \varphi_i$, we have

$$\eta_i^j = (x - \lambda_i)^j\varphi_i \not\equiv 0 \quad (j < \nu_i), \qquad \eta_i\varphi_i = \eta_i = \varphi_i\eta_i, \tag{29}$$

$$x\varphi_i = \lambda_i\varphi_i + \eta_i, \quad x\eta_i^j = \lambda_i\eta_i^j + \eta_i^{j+1}, \tag{30}$$

equation (26) therefore becomes

$$g(x) = \sum_1^r \left[g(\lambda_i)\varphi_i + g'(\lambda_i)\eta_i + \cdots + \frac{g^{(\nu_i - 1)}(\lambda_i)}{(\nu_i - 1)!}\eta_i^{\nu_i - 1} \right] \tag{31}$$

and in particular

$$x = \sum_1^r (\lambda_i\varphi_i + \eta_i) = \Sigma x_i. \tag{32}$$

The matrices φ_i and η_i are called the *principal idempotent* and *nilpotent elements* of x corresponding to the root λ_i. The matrices φ_i are uniquely determined by the following conditions: if ψ_i $(i = 1, 2, \cdots, r)$ are any matrices such that

$$\begin{aligned} &\text{(i)} \quad x\psi_i = \psi_i x, \\ &\text{(ii)} \quad (x - \lambda_i)\psi_i \text{ is nilpotent}, \\ &\text{(iii)} \quad \sum_i \psi_i = 1, \qquad \psi_i^2 = \psi_i \not\equiv 0, \end{aligned} \tag{33}$$

then $\psi_i = \varphi_i$ $(i = 1, 2, \cdots, r)$. For let $\theta_{ij} = \varphi_i\psi_j$; from (i) θ_{ij} also equals $\psi_j\varphi_i$. From (ii) and (28)

$$\eta_i = x\varphi_i - \lambda_i\varphi_i, \qquad \xi_j = x\psi_j - \lambda_j\psi_j$$

are both nilpotent and, since η_i and φ_i are polynomials in x, they are commutative with ψ_j and therefore with ξ_j; also

$$\begin{aligned} x\theta_{ij} &= \lambda_i\theta_{ij} + (x - \lambda_i)\varphi_i\psi_j = \lambda_i\theta_{ij} + \eta_i\psi_j \\ &= \lambda_j\theta_{ij} + (x - \lambda_j)\varphi_i\psi_j = \lambda_j\theta_{ij} + \xi_j\varphi_i. \end{aligned}$$

Hence $(\lambda_i - \lambda_j)\theta_{ij} = \xi_i\varphi_i - \eta_i\psi_j$. But if μ is the greater of the indices of ξ_i and η_i, then, since all the matrices concerned are commutative, each term of $(\xi_i\varphi_i - \eta_i\psi_j)^{2\mu}$ contains ξ_i^{μ} or η_i^{μ} as a factor and is therefore 0. If $\theta_{ij} \neq 0$, this is impossible when $i \neq j$ since θ_{ij} is idempotent and $\lambda_i - \lambda_j \neq 0$. Hence $\varphi_i\psi_j = 0$ when $i \neq j$ and from (iii)

$$\psi_j = \psi_j\Sigma\varphi_i = \psi_j\varphi_j = \varphi_j\Sigma\psi_i = \varphi_j$$

which proves the uniqueness of the φ's.

2.12 We shall now determine the reduced equation of $g(x)$. If we set g_i for $g(x)\varphi_i$, then

$$(34) \qquad \begin{aligned} g_i &= g(\lambda_i)\varphi_i + g'(\lambda_i)\eta_i + \cdots + \frac{g^{(\nu_i - 1)}(\lambda_i)}{(\nu_i - 1)!}\eta_i^{\nu_i - 1} \\ &= g(\lambda_i)\varphi_i + \zeta_i, \end{aligned}$$

say, and if s_i is the order of the first derivative in (34) which is not 0, then ζ_i is a nilpotent matrix whose index k_i is given by $k_i = 1 < \nu_i/s_i \leq k_i$.

If $\Phi(\lambda)$ is a scalar polynomial, and $\gamma_i = g(\lambda_i)$,

$$\Phi(g(x)) = \sum \Phi(g_i)\varphi_i = \sum \left[\Phi(\gamma_i)\varphi_i + \Phi'(\gamma_i)\zeta_i + \cdots + \frac{\Phi^{(k_i - 1)}(\gamma_i)}{(k_i - 1)!}\zeta_i^{k_i - 1}\right]$$

so that $\Phi(g(x)) = 0$ if, and only if, $g(\lambda_i)$ is a root of $\Phi(\lambda)$ of multiplicity k_i. Hence, if

$$\Psi(\lambda) = \Pi[\lambda - g(\lambda_i)]^{k_i}$$

where when two or more values of i give the same value of $g(\lambda_i)$, only that one is to be taken for which k_i is greatest, then $\Psi(\lambda)$ is the reduced characteristic function of $g(x)$. As a part of this result we have the following theorem.

THEOREM 5. *If $g(\lambda)$ is a scalar polynomial and x a matrix whose distinct roots are $\lambda_1, \lambda_2, \cdots, \lambda_r$, the roots of the matrix $g(x)$ are*[5]

$$g(\lambda_1), g(\lambda_2), \cdots, g(\lambda_r).$$

If the roots $g(\lambda_i)$ are all distinct, the principal idempotent elements of $g(x)$ are the same as those of x; for condition (33) of §2.11 as applied to $g(x)$ are satisfied by φ_i $(i = 1, 2, \cdots, r)$, and these conditions were shown to characterize the principal idempotent elements completely.

2.13 **The square root of a matrix.** Although the general question of functions of a matrix will not be taken up till a later chapter, it is convenient to give here one determination of the square root of a matrix x.

[5] That these are roots of $g(x)$ follows immediately from the fact that $\lambda - x$ is a factor of $g(\lambda) - g(x)$; but it does not follow so readily from this that the only roots are those given except, of course, when $r = n$ and all the quantities $g(\lambda_i)$ are distinct.

If α and β are scalars, $\alpha \neq 0$, and $(\alpha + \beta)^{\frac{1}{2}}$ is expanded formally in a Taylor series,

$$(\alpha + \beta)^{\frac{1}{2}} = \alpha^{\frac{1}{2}} \sum_0^\infty \delta_r \left(\frac{\beta}{\alpha}\right)^r$$

then, if $S_\nu = \alpha^{\frac{1}{2}} \sum_0^{\nu-1} \delta_r(\beta/\alpha)^r$, it follows that

$$S_\nu^2 = \alpha + \beta + \alpha T_\nu \tag{35}$$

where T_ν is a polynomial in β/α which contains no power of β/α lower than the νth. If a and b are commutative matrices and a is the square of a known non-singular matrix $a^{\frac{1}{2}}$, then (35) being an algebraic identity in α and β remains true when a and b are put in their place.

If $x_i = \lambda_i\varphi_i + \eta_i$ is the matrix defined in §2.11 (32), then so long as $\lambda_i \neq 0$, we may set $\alpha = \lambda_i\varphi_i$, $\beta = \eta_i$ since $\lambda_i\varphi_i = (\lambda_i^{\frac{1}{2}}\varphi_i)^2$; and in this case the Taylor series terminates since $\eta_i^{\nu_i} = 0$, that is, $T_{\nu_i} = 0$ and the square of the terminating series for $(\lambda_i\varphi_i + \eta_i)^{\frac{1}{2}}$ in powers of η_i equals $\lambda_i\varphi_i + \eta_i$. It follows immediately from (32) and (27) that, if x is a matrix no one of whose roots is 0, the square of the matrix

$$x^{\frac{1}{2}} = \sum_1^r \lambda_i^{\frac{1}{2}} \Big[\varphi_i + \tfrac{1}{2}\lambda_i^{-1}\eta_i - \cdots + (-1)^{\nu_i - 2} \frac{(2\nu_i - 4)!}{2^{2\nu_i - 3}(\nu_i - 2)!(\nu_i - 1)!} \left(\frac{\eta_i}{\lambda_i}\right)^{\nu_i - 1} \Big] \tag{36}$$

is x.

If the reduced equation of x has no multiple roots, (36) becomes

$$x^{\frac{1}{2}} = \Sigma \lambda_i^{\frac{1}{2}} \varphi_i \tag{37}$$

and this is valid even if one of the roots is 0. If, however, 0 is a multiple root of the reduced equation, x may have no square root as, for example, the matrix $\left\| \begin{matrix} 0 & 1 \\ 0 & 0 \end{matrix} \right\|$.

Formula (36) gives 2^r determinations of $x^{\frac{1}{2}}$ but we shall see later that an infinity of determinations is possible in certain cases.

2.14 **Reducible matrices.** If $x = x_1 + x_2$ is the direct sum of x_1 and x_2 and e_1, e_2 are the corresponding idempotent elements, that is,

$$e_i x = x_i = x e_i, \qquad e_i e_j = 0 \qquad (i \neq j;\ i, j = 1, 2),$$

then $x^r = x_1^r + x_2^r$ $(r \geq 2)$ and we may set as before $1 = x^0 = x_1^0 + x_2^0 = e_1 + e_2$. Hence, if $f(\lambda) = \lambda^m + b_1\lambda^{m-1} + \cdots + b_m$ is any scalar polynomial, we have

$$f(x) = e_1 f(x_1) + e_2 f(x_2) = f(x_1) + f(x_2) - b_m,$$

and if $g(\lambda)$ is a second scalar polynomial

$$f(x)g(x) = e_1 f(x_1)g(x_1) + e_2 f(x_2)g(x_2).$$

Now if $f_i(\lambda)$ is the reduced characteristic function of x_i regarded as a matrix in the space determined by e_i, then the reduced characteristic function of x_i as a matrix in the original fundamental space is clearly $\lambda f_i(\lambda)$ unless λ is a factor of $f_i(\lambda)$ in which case it is simply $f_i(\lambda)$. Further the reduced characteristic function of $x = x_1 + x_2$ is clearly the least common multiple of $f_1(\lambda)$ and $f_2(\lambda)$; for if

$$\psi(\lambda) = f_1(\lambda)g_1(\lambda) = f_2(\lambda)g_2(\lambda)$$

then

$$\begin{aligned}\psi(x_1 + x_2) &= e_1\psi(x_1) + e_2\psi(x_2) \\ &= e_1 f_1(x_1)g_1(x_1) + e_2 f_2(x_2)g_2(x_2) = 0.\end{aligned}$$