CHAPTER VIII

FUNCTIONS OF MATRICES

8.01 **Matric polynomials.** The form of a polynomial in a matrix has already been discussed in §2.11 but we repeat the principal formulas here for convenience. If x is a matrix whose reduced characteristic function is

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

and

$$m_i(\lambda) = \varphi(\lambda)/(\lambda - \lambda_i)^{\nu_i}, \qquad M_i(\lambda)m_i(\lambda) + (\lambda - \lambda_i)^{\nu_i}N_i(\lambda) \equiv 1,$$

$$\varphi_i(\lambda) = M_i(\lambda)m_i(\lambda), \tag{2}$$

$$\varphi_i = \varphi_i(x), \qquad h_i = (x - \lambda_i)\varphi_i, \tag{3}$$

and if $g(\lambda)$ is a scalar polynomial in λ, then

$$g(x) = \sum_{i=1}^{r} \left[g(\lambda_i)\varphi_i + g'(\lambda_i)h_i + \cdots + \frac{g^{(\nu_i - 1)}(\lambda_i)h_i^{\nu_i - 1}}{(\nu_i - 1)!} \right]. \tag{4}$$

This formula can still be interpreted when the coefficients of $g(\lambda)$ are matrices, but in this case the notation $g(x)$ is ambiguous. Let $g(\lambda) = a_0 + a_1\lambda + \cdots + a_m\lambda^m$; then

$$a_0 + a_1x + \cdots + a_mx^m \quad \text{and} \quad a_0 + xa_1 + \cdots + x^ma_m$$

are called, respectively, the *dextro-* and *laevo-lateral* polynomials corresponding to $g(\lambda)$. It is clear that (4) holds for a dextro-lateral polynomial and will give the corresponding laevo-lateral polynomial if $g(\lambda_i)\varphi_i$, $g'(\lambda_i)h_i$, etc., are replaced by $\varphi_ig(\lambda_i)$, $h_ig'(\lambda_i)$, etc.

8.02 **Infinite series.** If $a_0, a_1, \cdots$ are matrices and λ a scalar variable the coordinates of the matrix

$$g(\lambda) = a_0 + a_1\lambda + a_2\lambda^2 + \cdots \tag{5}$$

are scalar infinite series in λ; and if each of these series converges for mod λ less than ρ, we say that the series (5) converges. When this condition is satisfied, the series

$$g(x) = a_0 + a_1x + a_2x^2 + \cdots \tag{6}$$

converges for any matrix x for which the absolute value of the root of greatest absolute value is less than ρ. For if g_m is the sum of the first m terms of (6), then by (4) $g_m = \sum_i g_{mi}$ where

$$g_{mi} = g_m(\lambda_i)\varphi_i + g_m'(\lambda_i)h_i + \cdots + \frac{g_m^{(\nu_i - 1)}(\lambda_i)h_i^{\nu_i - 1}}{(\nu_i - 1)!}.$$

The matrices φ_i, h_i are independent of m and, since the absolute value of each λ_i is less than ρ, $g_m(\lambda_i)$, $g_m'(\lambda_i)$, $\cdots$, $g_m^{(\nu_i - 1)}(\lambda_i)$ converge to $g(\lambda_i)$, $g'(\lambda_i)$, $\cdots$, $g^{(\nu_i - 1)}(\lambda_i)$ when m approaches infinity.

As an illustration of such a series we may define $\exp x$ and $\log(1 + x)$ by

$$\exp x = \epsilon^x = 1 + \frac{x}{1} + \frac{x^2}{2!} + \cdots \tag{7}$$

$$\log(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} + \cdots \tag{8}$$

The first of these converges for every matrix x, the second for matrices all of whose roots are less than 1 in absolute value.

The usual rules for adding series and for multiplying series whose coefficients are commutative with x and with each other hold for matric series. For instance we can show by the ordinary proof that, if $xy = yx$, then $\epsilon^{x+y} = \epsilon^x\epsilon^y$ but this will not usually be the case if $xy \neq yx$.

8.03 **The canonical form of a function.** In the case of multiform functions (4) does not always give the most general determination of the function which is only obtained by taking into account the partial as well as the principal elements of the variable x. As in §3.06 suppose that x has the canonical form

$$x = \left\| \begin{matrix} a_1 & & & & \\ & a_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & a_r \end{matrix} \right\| \tag{9}$$

where a_i is a block of terms

$$\begin{matrix} \lambda_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 & 0 \\ \cdots\cdots & & & \cdots & \cdots & \\ \cdots\cdots & & & \cdots & \cdots & \\ 0 & 0 & 0 & \cdots & \lambda_i & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda_i \end{matrix} \qquad (r_i \text{ rows and columns}). \tag{10}$$

It is convenient to let a_i stand also for the matrix derived from (9) by replacing every a_j by 0 except when $j = i$. We can then write

$$x = \Sigma a_i, \qquad a_i a_j = 0 \qquad (i \neq j) \tag{11}$$

and we may set

$$a_i = \lambda_i e_i + z_i \tag{12}$$

where (cf. §3.07) $e_i^2 = e_i$, z_i is a nilpotent matrix of index r_i, and

$$\sum_i e_i = 1, \qquad e_i z_i = z_i = z_i e_i, \qquad e_i e_j = 0, \qquad e_i z_j = 0 = z_j e_i \qquad (i \neq j).$$

The part of z_i which is not 0 is given by the oblique line of 1's in (10); z^2 is obtained by moving all the 1's one place to the right except the last which disappears, and in general z^{m-1} has a line of 1's starting in the mth column of (10) and running parallel to the main diagonal till it meets the boundary of the block.

It is now easy to see the form of a scalar polynomial $g(x)$ or of a convergent power series with scalar coefficients; for

$$g(x) = \Sigma g(a_i)e_i = \Sigma\left[g(\lambda_i)e_i + g'(\lambda_i)z_i + \cdots + \frac{g^{(r_i-1)}(\lambda_i)z_i^{r_i-1}}{(r_i-1)!}\right] \tag{13}$$

and the block of terms in $g(x)$ which corresponds to a_i in (10) is, omitting the subscripts for clearness,

$$\begin{array}{ccccccc} g(\lambda) & g'(\lambda) & \frac{g''(\lambda)}{2!} & \frac{g'''(\lambda)}{3!} & \cdots & \frac{g^{(r-1)}(\lambda)}{(r-1)!} \\ & g(\lambda) & g'(\lambda) & \frac{g''(\lambda)}{2!} & \cdots & \frac{g^{(r-2)}(\lambda)}{(r-2)!} \\ & & g(\lambda) & g'(\lambda) & \cdots & \frac{g^{(r-3)}(\lambda)}{(r-3)!} \\ & & & & \cdots & \\ & & & & \cdot\cdot & \\ & & & & & g'(\lambda) \\ & & & & & g(\lambda) \end{array} \tag{14}$$

where all the terms to the left of the main diagonal are 0, the coordinates in the first row are as indicated, and all those on a line parallel to the main diagonal are the same as the one where this line meets the first row.

If the characteristic function is the same as the reduced function, no two blocks of terms in (9) correspond to the same root and e_i, z_i are the principal idempotent and nilpotent elements of x corresponding to λ_i and (13) is the same as (4). This is not the case when the same root occurs in more than one

of the blocks (10) and, when this is so, the a_i are not necessarily uniquely determined. For instance let

$$x = \begin{Vmatrix} \gamma & 0 & 0 \\ 0 & \gamma & 1 \\ 0 & 0 & \gamma \end{Vmatrix} = \gamma + e_{23}. \tag{15}$$

Here we have

$$a_1 = \gamma e_{11}, \qquad a_2 = \gamma(e_{22} + e_{33}) + e_{23},$$
$$e_1 = e_{11}, \qquad e_2 = e_{22} + e_{33}, \qquad z_1 = 0, \qquad z_2 = e_{23}.$$

But if

$$f_{11} = e_{11} - e_{13} \qquad f_{12} = e_{12} \qquad f_{13} = e_{13}$$
$$f_{21} = e_{21} - e_{23} \qquad f_{22} = e_{22} \qquad f_{23} = e_{23}$$
$$f_{31} = e_{31} - e_{13} + e_{11} - e_{33}, \qquad f_{32} = e_{32} + e_{12}, \qquad f_{33} = e_{33} + e_{13},$$

the f's form a set of matric units and

$$x = \gamma f_{11} + \gamma(f_{22} + f_{33}) + f_{23}$$

so that we might have chosen $f_1 = f_{11}$, $f_2 = f_{22} + f_{33} = e_{22} + e_{33} + e_{13}$ as idempotent elements in place of e_1 and e_2.

It should be carefully noted that f_1, f_2 are not commutative with e_1, e_2 and in consequence different determinations of a multiform function may not be commutative with each other. For instance, if x is the matrix given in (15) with $\gamma \neq 0$, and $\gamma^{\frac{1}{2}}$ is a particular determination of the square root of γ, we have already seen in §2.13 that determinations of $x^{\frac{1}{2}}$ are given by

$$\begin{aligned} u_1 &= \gamma^{\frac{1}{2}} e_{11} - \gamma^{\frac{1}{2}}(e_{22} + e_{33}) - e_{23}/2\gamma^{\frac{1}{2}} \\ u_2 &= \gamma^{\frac{1}{2}} f_{11} - \gamma^{\frac{1}{2}}(f_{22} + f_{33}) - f_{23}/2\gamma^{\frac{1}{2}} \\ &= \gamma^{\frac{1}{2}}(e_{11} - e_{13}) - \gamma^{\frac{1}{2}}(e_{22} + e_{33} + e_{13}) - e_{23}/2\gamma^{\frac{1}{2}} \\ &= u_1 - 2\gamma^{\frac{1}{2}} e_{13}, \end{aligned}$$

and these two values of $x^{\frac{1}{2}}$ are not commutative.

8.04 **Roots of 0 and 1.** The reduced equation of a nilpotent matrix of index m is $x^m = 0$ and this matrix can therefore be defined as a primitive mth root of 0; the index m cannot be greater than n and it exceeds 1 unless $x = 0$. The canonical form of x must contain at least one block of order $r_1 = m$, similar to (10) but with $\lambda_i = 0$, and a number of like blocks of orders, say, r_i ($i = 2, 3, \cdots$) where $r_i \leq r_1$ and $\sum_1 r_i = n$. This gives rise to a series of distinct types in number equal to the number of partitions of $n - m$ into parts no one of which exceeds m, and x is a primitive mth root if, and only if, it is similar to one of these types.

If x is a primitive mth root of 1, its reduced characteristic function is a factor of $\lambda^{m'} - 1$ and hence x has simple elementary divisors. Let ϵ be a scalar primitive mth root of 1, and let $f_1, f_2, \cdots, f_s$ be idempotent matrices of ranks $r_1, r_2, \cdots$ for which $f_i f_j = 0$ $(i \neq j)$, $\Sigma f_i = 1$; for instance, if $\rho_i = \sum_{j=1}^{i-1} r_j$, we may set

$$f_i = \sum_{\rho_i+1}^{\rho_{i+1}} e_{jj} \qquad (i = 1, 2, \cdots, s; \rho_{s+1} = n).$$

The canonical form for x is then

$$\epsilon^{t_1} f_1 + \epsilon^{t_2} f_2 + \cdots + \epsilon^{t_s} f_s \tag{16}$$

where the exponents t_i are all different modulo m and at least one ϵ^{t_i}, say the first, is primitive. Any primitive mth root of 1 is then similar to a matrix of the form (16), and conversely.

8.05 The equation $y^m = x$; algebraic functions. Let $\lambda_1, \lambda_2, \cdots, \lambda_s$ be the distinct roots of x and $\mu_i = \lambda_i^{1/m}$ a particular determination of the mth root of λ_i for $i = 1, 2, \cdots, s$; then, if $y^m = x$, the roots of y are all of the form $\epsilon^{t_i}\mu_i$, where ϵ is a primitive scalar mth root of 1. Suppose that the roots of y are

$$\begin{aligned} &\mu_{11} = \mu_1, \quad \mu_{12} = \epsilon^{t_{12}}\mu_1, \quad \cdots, \quad \mu_{1r_1} = \epsilon^{t_{1r_1}}\mu_1, \\ &\qquad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\mu_{s1} = \mu_s, \quad \mu_{s2} = \epsilon^{t_{s2}}\mu_s, \quad \cdots, \quad \mu_{sr_s} = \epsilon^{t_{sr_s}}\mu_s, \end{aligned} \tag{17}$$

and let a particular choice of the partial idempotent and nilpotent elements corresponding to μ_{ij} be f_{ijk} and h_{ijk} $(k = 1, 2, \cdots)$; also let the index of h_{ijk} be n_{ijk}. Then

$$y = \Sigma(\mu_{ij} f_{ijk} + h_{ijk}) \tag{18}$$

and hence

$$y^m = \Sigma(\mu_{ij} f_{ijk} + h_{ijk})^m = \Sigma(\lambda_i f_{ijk} + g_{ijk}) \tag{19}$$

where g_{ijk} is the nilpotent matrix

$$g_{ijk} = (\mu_{ij} f_{ijk} + h_{ijk})^m - \mu_{ij}^m f_{ijk}. \tag{20}$$

Further, if $\lambda_i \neq 0$, (20) can be solved for h_{ijk} as a polynomial in g_{ijk}; for we can write (20) in the form

$$g_{ijk} = \alpha_1 h_{ijk} + \alpha_2 h_{ijk}^2 + \cdots \tag{21}$$

and, since $\alpha_1 = m\mu_{ij}^{m-1} \neq 0$, the ordinary process for inverting a power series shows that we can satisfy (21) by a series of the form

$$h_{ijk} = \beta_1 g_{ijk} + \beta_2 g_{ijk}^2 + \cdots, \qquad (\beta_1 \neq 0), \tag{22}$$

there being here no question of convergence since any power series in a nilpotent matrix terminates. It follows from (21) and (22) that the indices of g_{ijk} and h_{ijk} are the same.

We shall now show that the matrices f_{ijk} and g_{ijk} form a set of partial idempotent and nilpotent elements of x provided always that x is not singular. If this were not so, then f_{ijk} must be the sum of two or more partial idempotent elements; for the sake of brevity we shall assume that it is the sum of two since the proof proceeds in exactly the same way if more components are taken. Let $f_{ijk} = d_1 + d_2$ where d_1 and d_2 are partial idempotent elements of x and let c_1, c_2 be the corresponding nilpotent elements; then

$$g_{ijk} = c_1 + c_2, \qquad c_1c_2 = 0 = c_2c_1.$$

Hence also $h_{ijk} = b_1 + b_2$, $b_1b_2 = 0 = b_2b_1$ where b_α is obtained by putting c_α for g_{ijk} in (22); and this is impossible since we assumed that f_{ijk} and h_{ijk} were partial idempotent and nilpotent elements of y. We have therefore the following theorem.

THEOREM 1. *If x is a non-singular matrix, any determination of $y = x^{1/m}$ can be obtained by expressing x in terms of partial idempotent and nilpotent elements, say $x = \Sigma(\lambda_i f_i + g_i)$ and putting*

$$y = \Sigma(\lambda_i f_i + g_i)^{1/m} = \Sigma\lambda_i^{1/m}(f_i + m^{-1}\lambda_i^{-1}g_i + \tfrac{1}{2}m^{-1}(m^{-1} - 1)\lambda_i^{-2}g^2 + \cdots).$$

Here the binomial series terminates and $\lambda_i^{1/m}$ is a determination of the mth root of λ_i which may be different for different terms of the summation if this root occurs with more than one partial element.

There is thus a two-sided multiplicity of mth roots of x; the $\lambda_i^{1/m}$ have m possible determinations in each term and also there is in certain cases an infinity of ways of choosing the set of partial elements. Since the canonical form is independent of the actual choice of the set of partial elements out of the possible sets, any choice of such a set can be derived from any other such set by transforming it by a matrix u; and since x itself is the same no matter what set of partial elements is chosen, we have $uxu^{-1} = x$, that is, u is commutative with x. It follows from the development given in §§7.01, 7.02 and 7.04 that a matrix u which is commutative with *every* partial idempotent element is a polynomial in x.

8.06 We must now consider the case in which x is singular and in doing so it is sufficient to discuss mth roots of a nilpotent matrix; for the principal idempotent element of x which corresponds to a root μ is the sum of those principal idempotent elements of y which correspond to those roots whose mth power is μ, so that the principal idempotent element corresponding to the root 0 is the same for both x and y. Let the elementary divisors of y be $\lambda^{m_1}, \lambda^{m_2}, \cdots, \lambda^{m_p}$; then

$$y = y_1 + y_2 + \cdots + y_p$$

where y_i is a nilpotent matrix of index m_i, and we may suppose the fundamental basis so chosen that the significant part of y_i is

$$\begin{matrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & 0 \end{matrix} \qquad (m_i \text{ rows and columns}). \tag{23}$$

To simplify the notation we shall consider for the moment only one part y_i and replace it by y and m_i by n so that $y^n = 0$ and

$$y = \sum_1^{n-1} e_i S e_{i+1}.$$

If we now form the mth power of y, then $y^m = 0$ if $m \geq n$ and if $m < n$

$$y^m = \sum_1^{n-m} e_i S e_{i+m}.$$

If we define r and k by

$$(r-1)m + k = n \leq rm \qquad (k > 0) \tag{24}$$

then $r \geq 2$ and

$$y^m e_i = 0,\ y^m e_{i+m} = e_i,\ y^m e_{i+2m} = e_{i+m},\ \cdots,\ y^m e_{i+(r-1)m} = e_{i+(r-2)m} \qquad (i = 1, 2, \cdots, k) \tag{25}$$

giving k chains of order r of invariant vectors, and similarly for $i = k+1$, $k+2$, $\cdots$, m, we have $m - k$ chains whose order is $r - 1$ since for these values of i the last equation in (25) is missing. If we set u and v for blocks of terms like (23) only with r and $r-1$ rows and columns, respectively, then we can find a non-singular matrix P which permutes the rows and columns in y^m so that

$$P^{-1}y^mP = \left\| \begin{matrix} u & & & & & & \\ & u & & & & & \\ & & \ddots & & & & \\ & & & u & & & \\ & & & & v & & \\ & & & & & \ddots & \\ & & & & & & v \end{matrix} \right\| \qquad (k\ u\text{'s and } m-k\ v\text{'s}). \tag{26}$$

We are now in a position to consider the solution of $y^m = x$ where x is a nilpotent matrix of index r. In the elementary divisors of x suppose p_1 expo-

nents equal r, p_2 equal $r-1$, and in general p_j equal $r-j+1$ $(j=1, 2, \cdots, r)$; here the p's are integers equal to or greater than 0 such that

$$\sum_1 (r-j+1)p_j = n, \qquad p_1 \neq 0.$$

The maximum possible exponent for any elementary divisor of y is rm; let q_{ji} $(i=1, 2, \cdots, m; j=1, 2, \cdots, r)$ be the number of exponents which equal

$$(r-j+1)m - i + 1 = (r-j)m + (m-i+1).$$

Forming y^m and using (24) and the results of (25) we then see that

$$\begin{aligned} (27) \qquad & q_{j-1,2} + 2q_{j-1,3} + \cdots + (m-1)q_{j-1,m} + mq_{j1} + (m-1)q_{j2} \\ & \qquad + \cdots + q_{jm} = p_j \qquad (j=1, 2, \cdots, r;\ q_{0i} = 0) \\ & \sum_{i,j} [(r-j+1)m - i + 1]\, q_{ji} = n. \end{aligned}$$

These relations form a set of Diophantine equations for the q's. When a set of q's have been found, we can find the matrix P (cf. (26)) for each part of y^m and then set $y = \Sigma R^{-1} y_i R$ where R has the form $\Sigma P_i Q_i$, Q_i being commutative with $P_i^{-1} y_i^m P_i$ and so chosen that R is not singular.

8.07 **The exponential and logarithmic functions.** The function $\exp y = \epsilon^y$ has already been defined in §8.02 by the series

$$\epsilon^y = 1 + \sum_1^\infty y^m/m!$$

or in §8.03 in terms of the partial units of y. Let the distinct roots of y be $\mu_1, \mu_2, \cdots, \mu_s$ and let a choice of the partial idempotent and nilpotent elements corresponding to μ_i be f_{ij}, h_{ij} $(j=1, 2, \cdots, k_i)$ so that

$$(28) \qquad f_i = \sum_{j=1}^{k_i} f_{ij}, \qquad h_i = \sum_{j=1}^{k_i} h_{ij} \qquad (i=1, 2, \cdots, s)$$

are the principal idempotent and nilpotent elements of y. If we set $x = \epsilon^y$, we have

$$\begin{aligned} (29) \qquad x = \epsilon^y &= \sum_i \sum_j \left[\epsilon^{\mu_i} f_{ij} + h_{ij} + \frac{h_{ij}^2}{2!} + \cdots + \frac{h_{ij}^{\nu_{ij}-1}}{(\nu_{ij}-1)!} \right] \\ &= \sum_i \sum_j (\epsilon^{\mu_i} f_{ij} + g_{ij}) \end{aligned}$$

where ν_{ij} is the index of h_{ij} and

$$(30) \qquad g_{ij} = h_{ij} + \frac{h_{ij}^2}{2} + \cdots + h_{ij}^{\nu_{ij}-1}/(\nu_{ij}-1)!.$$

The index of g_{ij} is clearly ν_{ij}. Solving (30) for h_{ij}, we have the formal solution $h_{ij} = \log(1 + g_{ij})$ and on using this or inverting the power series in (30) we get

$$h_{ij} = g_{ij} - \tfrac{1}{2}g_{ij}^2 + \cdots + (-1)^{\nu_{ij}} g_{ij}^{\nu_{ij}-1}/(\nu_{ij} - 1). \tag{31}$$

As in §8.05 it follows that f_{ij}, g_{ij} form a set of partial elements for x and, when x is given so that $y = \log x$, the method there used gives the following theorem.

THEOREM 2. *If x is a non-singular matrix whose distinct roots are $\lambda_1, \lambda_2, \cdots, \lambda_r$, and if $\log \lambda_1, \log \lambda_2, \cdots, \log \lambda_r$ are particular determinations of the logarithms of these roots, then the general determination of $\log x$ is found as follows. Take any set of partial elements of x, say f_{ij}, g_{ij} $(i = 1, 2, \cdots, r; j = 1, 2, \cdots, k_i)$ where f_{ij}, g_{ij} correspond to λ_i and the index of g_{ij} is ν_{ij}, let h_{ij} be the nilpotent matrix defined by* (31), *and let k_{ij} be any integers, then*

$$\log x = \sum_i \sum_j [(\log \lambda_i + k_{ij}\pi\sqrt{-1})f_{ij} + h_{ij}]. \tag{32}$$

The discussion of the relation between different determinations of $\log x$ is practically the same as for $x^{1/m}$ and need not be repeated.

If f_i and h_i are defined by (28), a particular determination of $\log x$ is given by

$$\operatorname{Log} x = \sum_i [(\log \lambda_i + k_i\pi\sqrt{-1})f_i + h_i]. \tag{33}$$

This form of $\log x$ has the same principal elements as x provided $\log \lambda_i + k_i \neq \log \lambda_j + k_j$ for any $i \neq j$, and even when this condition is not satisfied, it is convenient to refer to (33) as a *principal* determination of $\log x$. This determination is the one given by the series (cf. §8.02 (8))

$$\log x = (x - 1) - \tfrac{1}{2}(x - 1)^2 + \tfrac{1}{3}(x - 1)^3 - \cdots \tag{34}$$

provided each k_i is 0 and the principal determination of $\log \lambda_i$ is used. The series converges only when the roots of $x - 1$ are all less than 1 in absolute value.

8.08 **The canonical form of a matrix in a given field.** If the coefficients of a matrix are restricted to lie in a given field of rationality, the canonical form used in the preceding sections requires some modification. The definition of the invariant factors is rational as are also the theorems regarding similar matrices which were derived from them in Chapter 3; and hence if X and x are rational matrices which have the same invariant factors there exists a rational matrix P for which $P^{-1}xP = X$. The definition of elementary divisors requires only the natural alteration of substituting powers of irreducible polynomials for $(\lambda - \lambda_i)^{\nu_{ij}}$.

Let

$$\alpha(\lambda) = \lambda^m + a_1\lambda^{m-1} + \cdots + a_m$$

be a scalar polynomial in a field F which is irreducible in F; then the matrix of order m defined by

$$x_a = \begin{matrix} -a_1 & -a_2 & \cdots & -a_{m-1} & -a_m \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & \cdots & 1 & 0 \end{matrix} \tag{35}$$

has $\alpha(\lambda)$ as its characteristic function; since $\alpha(\lambda)$ is irreducible in F, it follows immediately that x_a is an irreducible matrix in F and that $\alpha(\lambda)$ is also the reduced characteristic function. It is easily seen that the invariant factors of $\lambda - x$ are given by $m - 1$ units followed by $\alpha(\lambda)$.

Again, if we consider

$$x_a = \begin{matrix} x_a & 1_m & & & \\ & x_a & 1_m & & \\ & & \cdot & \cdot & \\ & & & \cdot & \cdot \\ & & & & \cdot \quad 1_m \\ & & & & x_a \end{matrix} \qquad (r \text{ rows and columns}) \tag{36}$$

which is a matrix of order rm, we see as in §8.03 (14) that, if $g(\lambda)$ is a scalar polynomial in λ,

$$g(x_a) = \begin{matrix} g(x_a) & g'(x_a) & \cdots & g^{(r-1)}(x_a)/(r-1)! \\ & g(x_a) & \cdots & g^{(r-2)}(x_a)/(r-2)! \\ & & \cdots & \cdot \\ & & \cdot\cdot & \cdot \\ & & \cdot & g'(x_a) \\ & & & g(x_a). \end{matrix}$$

It follows that, if $g(x_a) = 0$, we must have $g^{(r-1)}(x_a) = 0$ and therefore $\alpha(\lambda)$ is a factor of $g^{(r-1)}(\lambda)$ so that $[\alpha(\lambda)]^r$ is a factor of $g(\lambda)$. But if we put $g(\lambda) = [\alpha(\lambda)]^r$ the first $(r-1)$ derivatives of $g(\lambda)$ have $\alpha(\lambda)$ as a factor and so vanish when λ is replaced by x_a; hence $g(x_a) = 0$. It follows that the reduced characteristic function of x_a is $[\alpha(\lambda)]^r$ and, since the degree of this polynomial equals the order rm of x_a, it is also the characteristic function so that the invariant factors of x_a are given by 1 repeated $rm - 1$ times followed by $[\alpha(\lambda)]^r$. The argument used in §3.06 then gives the following theorem.

THEOREM 3. *Let $\alpha_1(\lambda)$, $\alpha_2(\lambda)$, $\cdots$, $\alpha_k(\lambda)$ be polynomials, not necessarily distinct, which are rational and irreducible in a field F and whose degrees are m_1, m_2, $\cdots$, m_k respectively; and let r_1, r_2, $\cdots$, r_k be any positive integers such that*

$\sum_1^k r_i m_i = n$; *then, if* x_{α_i} *is the matrix of order* $r_i m_i$ *formed from* $\alpha_i(\lambda)$ *in the same way as* x_α *in* (36) *is formed from* $\alpha(\lambda)$, *the matrix of order* n *defined by*

$$x = \begin{matrix} x_{\alpha_1} & & & \\ & x_{\alpha_2} & & \\ & & \cdot & \\ & & & \cdot \\ & & & x_{\alpha_k} \end{matrix} \tag{37}$$

has $\alpha_1(\lambda)$, $\alpha_2(\lambda)$, $\cdots$, $\alpha_k(\lambda)$ *as its elementary divisors in* F.

If X is any matrix with the same elementary divisors as x, it follows from §3.04 Theorem 5 that we can find a rational nonsingular matrix P such that $PXP^{-1} = x$. We may therefore take (37) as a canonical form for a matrix in the given field F.

8.09 **The absolute value of a matrix.** The absolute value of a matrix $a = \| a_{po} \|$ is most conveniently defined as

$$|a| = \left(\sum_{p,q=1}^{n} a_{pq}\bar{a}_{pq} \right)^{\frac{1}{2}} \tag{38}$$

where the heavy bars are used to distinguish between the absolute value and the determinant $|a|$. It must be carefully noted that the absolute value of a scalar matrix λ is not the same as the ordinary absolute value or modulus of λ, the relation between them being

$$|\lambda| = n^{\frac{1}{2}} \bmod \lambda. \tag{39}$$

It follows immediately from (37) that

$$\bmod (|a| - |b|) \leq |a + b| \leq |a| + |b|; \tag{40}$$

and from

$$\sum_{p,r} a_{pr}\bar{a}_{pr} \sum_{q\,s} b_{sq}\bar{b}_{sq} = \sum_{p,q} \left[\tfrac{1}{2} \sum_{r,s} (a_{pr}\bar{b}_{sq} - a_{ps}b_{rq})(\bar{a}_{pr}b_{sq} - \bar{a}_{ps}b_{rq}) + \sum_r a_{pr}b_{rq} \sum_s \bar{a}_{ps}\bar{b}_{sq} \right],$$

we have

$$|ab| \leq |a|\,|b|. \tag{41}$$

Since the trace of $a\bar{a}'$ is $\Sigma a_{pq}\bar{a}_{pq}$, the absolute value of a might also have been defined by

$$|a|^2 = \operatorname{tr} a\bar{a}' = \operatorname{tr} \bar{a}'a. \tag{42}$$

From this we see immediately that, if a is unitary, that is, $a\bar{a}' = 1$, then

$$|a| = n^{\frac{1}{2}}, \qquad |ab| = |b| \tag{43}$$

where b is any matrix.

No matter what matrix a is, $a\bar{a}'$ is a positive hermitian matrix, semi-definite or definite according as a is or is not singular; the roots $g_1, g_2, \cdots, g_n$ of $a\bar{a}'$ are therefore real and not negative. If we set

$$p_r = \Sigma g_1 g_2 \cdots g_r, \qquad p \equiv p_n = (\mathrm{mod}\ |a|)^2, \qquad s \equiv p_1 = |a|^2 = \Sigma g_i,$$

then

$$r^r p_r \leq \Sigma(g_1 + g_2 + \cdots + g_r)^r \leq \binom{n-1}{r-1} s^r, \qquad p_r \geq \binom{n}{r} p^{\binom{n-1}{r-1}/\binom{n}{r}} = \binom{n}{r} p^{r/n},$$

whence[1]

$$\binom{n}{r} p^{r/n} \leq p_r \leq r^{-r}\binom{n-1}{r-1} s^r. \tag{44}$$

If $C^r(a)$ is the rth supplementary compound of a (cf. §5.03), and $\alpha = p^{\frac{1}{2}}$ is put for mod $|a|$, then $p_r = |C^r(a)|^2$ and we may write for (44)

$$\binom{n}{r} \alpha^{2r/n} \leq |C^r(a)|^2 \leq r^{-r}\binom{n-1}{r-1} |a|^{2r}$$

and, since $(a')^{-1} = C^{n-1}(a)/|a|$, we have

$$n^{\frac{1}{2}} \alpha^{-1/n} \leq |a^{-1}| \leq (n-1)^{-\frac{n-2}{2}} |a|^{n-1}/\alpha \tag{45}$$

provided $|a| \neq 0$. This inequality enables us to deal with expressions involving negative powers of a.

Since $a^{-1} = \exp(-\log a)$, we also have

$$|a^{-1}b| = |(1 - \log a + \tfrac{1}{2!}(\log a)^2 - \cdots)b| \leq (1 + \tfrac{1}{2!}|\log a|^2 + \cdots)|b|$$

and therefore

$$|a^{-1}b| \leq \epsilon^{|\log a|}|b|. \tag{46}$$

Putting $b = 1$ we also have

$$|a^{-1}| \leq n^{\frac{1}{2}} - 1 + \epsilon^{|\log a|} \quad \text{and} \quad |a^{-1}| \leq n^{\frac{1}{2}}\epsilon^{|\log a|}. \tag{47}$$

It is also sometimes convenient to note as a consequence of (41) with $b = a^{-1}$ that

$$|a^{-1}| \geq n^{\frac{1}{2}}|a|^{-1}. \tag{48}$$

[1] If $r = n$, (44) gives Hadamard's expression for the maximum value of mod $|a|$.

8.10 **Infinite products.** As an illustration of the use of the preceding section we shall now investigate briefly the convergence of an infinite product. Let

$$P_m = (1 + a_1)(1 + a_2) \cdots (1 + a_m) \tag{49}$$
$$Q_m = (1 + |a_1|)(1 + |a_2|) \cdots (1 + |a_m|),$$

then, if c is an arbitrary matrix,

$$|P_m - 1| \leq Q_m - 1 < \epsilon^{\Sigma|a_i|} - 1, \tag{50}$$

$$|P_m c| \leq \epsilon^{\Sigma|a_i|}|c|, \tag{51}$$

$$|P_m - P_k| \leq Q_m - Q_k \leq \sum_k^m |a_i| \epsilon^{\sum_1^m |a_i|}. \tag{52}$$

For on expanding P_m we have

$$P_m = 1 + \sum_p a_p + \sum_{p<q} a_p a_q + \sum_{p<q<r} a_p a_q a_r + \cdots$$

therefore

$$|P_m - 1| \leq \sum_p |a_p| + \sum_{p<q} |a_p||a_q| + \cdots = Q_m - 1$$

$$|P_m c| \leq (1 + \sum_p |a_p| + \sum_{p<q} |a_p||a_q| + \cdots)|c| = Q_m|c| < \epsilon^{\Sigma|a_i|}|c|.$$

The proof of (52) follows in the same manner.

Hence P_m converges when Q_m does, for which it is sufficient that Σa_i is absolutely convergent in the sense that $\Sigma|a_i|$ converges.

8.11 **The absolute value of a tensor.** If w is a tensor of grade r, we define the absolute value of w by[2]

$$\text{mod } w = (r! S\bar{w}w)^{\frac{1}{2}}. \tag{53}$$

We shall for the most part consider only vectors of grade 1 as the extension to tensors of higher grade is usually immediate.

If x and y are any vectors, we derive from (53)

$$\text{mod } (x + y) \leq \text{mod } x + \text{mod } y, \qquad \text{mod } Sxy \leq \text{mod } x \text{ mod } y. \tag{54}$$

If A is a matrix,

$$(\text{mod } Ax)^2 = S\bar{A}\bar{x}Ax = S\bar{x}\bar{A}'Ax.$$

By §6.02

$$\bar{A}'A = \Sigma g_i \bar{a}_i S a_i$$

[2] The $r!$ enters here only because of the numerical factor introduced in defining Suv (cf. §5.16).

where the g's are real and not negative, and $S\bar{a}_i a_j = \delta_{ij}$ so that mod $a_i = 1$; hence

$$\begin{aligned} S\bar{x}\bar{A}'Ax &= \Sigma g_i S\bar{x}\bar{a}_i S a_i x = \Sigma g_i \,(\text{mod } Sa_i x)^2 \\ &\leq \Sigma g_i \,(\text{mod } a_i \text{ mod } x)^2 = (\Sigma g_i)(\text{mod } x)^2 \\ &= |A|^2 \,(\text{mod } x)^2, \end{aligned}$$

or

$$\text{mod } Ax \leq |A| \text{ mod } x. \tag{55}$$

From (54) we then have

$$\text{mod } SyAx \leq |A| \text{ mod } x \text{ mod } y. \tag{56}$$

8.12 **Matric functions of a scalar variable.** If the coordinates of a matrix $a(t) = \| a_{pq}(t) \|$ are functions of a scalar variable t, the matrix itself is called a matric function of t. The derivative, when it exists, is defined as

$$\frac{da}{dt} = \lim_{h \to 0} \frac{a(t+h) - a(t)}{h} = \left\| \frac{da_{pq}}{dt} \right\| \tag{57}$$

h being a scalar. The fundamental rules of differentiation are

$$\frac{d(a+b)}{dt} = \frac{da}{dt} + \frac{db}{dt}, \qquad \frac{d(ab)}{dt} = \frac{da}{dt} b + a \frac{db}{dt}, \qquad \frac{da'}{dt} = \left(\frac{da}{dt}\right)'$$

to which we may add, when $|a| \neq 0$,

$$\frac{da^{-1}}{dt} = -a^{-1} \frac{da}{dt} a^{-1}. \tag{58}$$

Other examples are

$$\frac{da^2}{dt} = \frac{da}{dt} a + a \frac{da}{dt}, \qquad \frac{da^3}{dt} = \frac{da}{dt} a^2 + a \frac{da}{dt} a + a^2 \frac{da}{dt}$$

and in general, if m is any positive integer,

$$\frac{da^m}{dt} = \begin{Bmatrix} a & da/dt \\ m-1 & 1 \end{Bmatrix}. \tag{59}$$

Under the usual conditions each of the coordinates of $a(t)$ is expansible as a Taylor series and this is therefore also true of $a(t)$. If $f(t)$ is a scalar function, $f(a)$ may or may not have a meaning. For instance, if $f(t)$ can be expanded in a power series which converges for mod $t < \alpha$, then the same power series[3] in

[3] If $g(t) = \Sigma u_n t^n$, u_n scalar, the series intended here is $\Sigma u_n a^n$. Other definitions are possible, e.g., if we set

$$G(a) = \sum_n u_n \sum_i c_{i0}^{(n)} a c_{i1}^{(n)} a c_{i2}^{(n)} \cdots c_{i,n-1}^{(n)} a c_{in}^{(n)}$$

where $\sum c_{i0}^{(n)} c_{i1}^{(n)} \cdots c_{in}^{(n)} = 1$, we still have $G(t) = g(t)$.

a converges when $|a| < \alpha$; but if $f(t)$ is defined by a Fourier series which is not differentiable, $f(a)$ will not have a meaning when the elementary divisors of a are not simple, as is seen immediately on referring to the form of §8.03 (14). If $f(a)$ and $f'(a)$ have a meaning and if da/dt is commutative with a, we have $df(a)/dt = f'(a)da/dt$. For instance, if x is a constant matrix and $a = t - x$, then

$$\frac{d \log (t - x)}{dt} = \frac{1}{t - x} \tag{60}$$

as is also easily proved directly.

The integral of $a(t)$ is defined as follows. If C is a regular contour in the t-plane, we shall set

$$\int_C a(t)dt = \left\| \int_C a_{pq}(t)dt \right\| \tag{61}$$

or if $t_1, t_2, \cdots$ is a series of points on C and t_i' a point on the arc (t_i, t_{i+1}), and if the number of points is increased indefinitely in such a way that mod $(t_{i+1} - t_i)$ approaches 0 for every interval, then

$$\int_C a(t)dt = \lim \Sigma a(t_i')(t_{i+1} - t_i). \tag{62}$$

The conditions for the existence of this limit are exactly the same as in the scalar theory.

If M is the least upper bound of $|a|$ on C and L is the length of C, it follows in the usual manner that

$$\left| \int_C a(t)dt \right| \leq \int |a(t)| \bmod dt \leq ML. \tag{63}$$

As an illustration of these definitions we shall now employ contour integration to prove some of our earlier results. If x is an arbitrary constant matrix and C a circle with center $t = 0$ and radius greater than $|x|$, then all the roots of x lie inside C and on C the series

$$\frac{1}{t - x} = \frac{1}{t} + \frac{x}{t^2} + \frac{x^2}{t^3} + \cdots$$

is uniformly convergent. Hence

$$\frac{1}{2\pi i} \int_C \frac{dt}{t - x} = \frac{1}{2\pi i} \sum_0^\infty x^n \int_C \frac{dt}{t^{n+1}} = 1 \tag{64}$$

a result which may also be derived from the definition of log $(t - x)$ in §8.07 and

$$\int_C \frac{dt}{t - x} = [\log (t - x)]_C.$$

We then have

$$\frac{1}{2\pi i}\int_C \frac{tdt}{t - x} = \frac{1}{2\pi i}\int_C \left(1 + \frac{x}{t - x}\right) dt = x$$

and in general, if $g(t)$ is a scalar function of t which is analytic in a region enclosing C,

$$g(x) = \frac{1}{2\pi i}\int_C \frac{g(t)}{t - x}\, dt. \tag{65}$$

Suppose now that $| t - x | = \theta(t)\varphi(t)$, adj $(t - x) = \theta(t)a(t)$ where $\theta(t)$ is the highest common factor of $| t - x |$ and the coordinates of adj $(t - x)$. We then have

$$g(x) = \frac{1}{2\pi i}\int_C \frac{g(t)}{\varphi(t)}\, a(t)dt \tag{66}$$

and under the given conditions this vanishes if, and only if, $g(t)/\varphi(t)$ has no singularities inside C, that is, if φ is a factor of g. We have therefore the theorem of §2.05 that $\varphi(t)$ is the reduced characteristic function of x and that $g(x) = 0$ only when $\varphi(t)$ is a factor of $g(t)$.

Since $a(t) \equiv \varphi(t)/(t - x)$ is a polynomial in x with scalar coefficients and with degree 1 less than the degree of $\varphi(t)$, say

$$a(t) = \alpha_1 x^{m-1} + \alpha_2 x^{m-2} + \cdots + \alpha_m,$$

equation (66) shows that $g(x)$ can be expressed as a polynomial in x, namely,

$$g(x) = \frac{1}{2\pi i}\Sigma x^{m-s}\int_C \frac{g(t)\alpha_s(t)}{\varphi(t)}\, dt. \tag{67}$$

We may also note that (66) leads to the interpolation formula §8.01 (4) if the integral is expanded in terms of the residues at the zeros of $\varphi(t)$.

All of these results can be extended to unilateral series in x with matric coefficients if care is taken to use $g(t)\ (t - x)^{-1}$ or $(t - x)^{-1}\ g(t)$ according as dextro- or laevo-lateral series are desired.

8.13 **Functions of a variable vector.** Before considering functions of a variable matrix, we shall consider briefly those of a variable vector; for more extended and systematic treatments the reader is referred to treatises on vector and tensor analysis.

The differential of a function of a variable in any non-commutative algebra

was defined by Hamilton as follows. Let $f(x)$ be a function of a variable x, dx a variable independent of x and t a scalar variable; then

$$df(x) = \lim_{t\to 0} \frac{f(x + tdx) - f(x)}{t}. \tag{68}$$

We shall assume tacitly hereafter that this limit exists for all the functions we shall consider.

An immediate consequence of (68) is that $df(x)$ is linear and homogeneous in dx. Hence, if $x = \Sigma\xi_i e_i$, $dx = \Sigma d\xi_i e_i$, then

$$df(x) = \lim_{t\to 0} \frac{f(\Sigma(\xi_i + td\xi_i)e_i) - f(\Sigma\xi_i e_i)}{t} = \Sigma \frac{\partial f}{\partial \xi_i} d\xi_i.$$

This leads to Hamilton's differential operator

$$\nabla = \Sigma e_i \frac{\partial}{\partial \xi_i} \tag{69}$$

in terms of which we may write (68) in the form

$$df(x) = (Sdx\nabla)f(x). \tag{70}$$

In using this operator it is frequently convenient to place it after its operand and, when this is done, some artifice is necessary to indicate the connection between them. This is done by attaching the same subscript to both; the method of doing this will be clear from the following examples in which $a = \Sigma\alpha_i e_i$, $b = \Sigma\beta_i e_i$ are vectors and $A = \| a_{ij} \|$ is a matrix.

$$\nabla a = \Sigma \frac{\partial \alpha_j}{\partial \xi_i} e_i e_j, \qquad a_\alpha \nabla_\alpha = \Sigma \frac{\partial \alpha_i}{\partial \xi_j} e_i e_j,$$

$$\nabla Sab = \nabla_\alpha Sa_\alpha b + \nabla_\alpha Sab_\alpha = \sum_i \left(\sum_j \left(\frac{\partial \alpha_j}{\partial \xi_i} \beta_j + \frac{\partial \beta_j}{\partial \xi_i} \alpha_j \right) \right) e_i,$$

$$S\nabla_\alpha \nabla_\beta Sa_\alpha b_\beta = \sum_{i,j} \frac{\partial \alpha_i}{\partial \xi_j} \frac{\partial \beta_i}{\partial \xi_j},$$

$$A_\alpha \nabla_\alpha = \sum_i \left(\sum_j \frac{\partial a_{ij}}{\partial \xi_j} \right) e_i, \qquad \nabla_\alpha A_\alpha = \Sigma \frac{\partial a_{jk}}{\partial \xi_i} e_i e_j Se_k(\ \ ),$$

$$Sa_\alpha A \nabla_\alpha = \sum_{i\,j} a_{ij} \frac{\partial \alpha_i}{\partial \xi_j}, \qquad SaA_\alpha \nabla_\alpha = \sum_{i,j} \alpha_i \frac{\partial a_{ij}}{\partial \xi_j}.$$

We can now consider the effect of a change of variable from x to[4] $\bar{x}$. Let $\bar{x} = \Sigma\bar{\xi}_i e_i$, $\nabla = \Sigma e_i \partial/\partial\bar{\xi}_i$; then

$$d\bar{x} = \bar{x}_\alpha S\nabla_\alpha dx = Jdx \tag{71}$$

[4] Here $\bar{x}$ denotes a new variable and not the conjugate imaginary. Instead of considering a change of variable, we may regard $\bar{x}$ as a vector function of x.

where

$$J = \bar{x}_\alpha S\nabla_\alpha = \| \partial\bar{\xi}_i/\partial\xi_j \| \tag{72}$$

is the Jacobian matrix of the transformation. Similarly

$$dx = x_\alpha S\nabla_\alpha d\bar{x} = \bar{J}d\bar{x}.$$

Hence

$$J\bar{J} = 1. \tag{73}$$

Again, since $Sdx\nabla = Sd\bar{x}\bar{\nabla} = SdxJ'\bar{\nabla}$, hence

$$\nabla = J'\bar{\nabla}, \qquad \bar{\nabla} = (J')^{-1}\nabla = \bar{J}'\nabla. \tag{74}$$

From (70) and (72) we see that the differentials of J and J' are given by

$$\begin{aligned} dJ &= Sdx\nabla_\alpha \cdot J_\alpha, \qquad \text{and} \qquad dJ = d\bar{x}_\alpha S\nabla_\alpha = J_\alpha dx S\nabla_\alpha \\ dJ' &= Sdx\nabla_\alpha \cdot J'_\alpha, \qquad dJ' = \nabla_\alpha SJ_\alpha dx = \nabla_\alpha Sdx J'_\alpha. \end{aligned} \tag{75}$$

This leads to the notion of contravariant and covariant vectors. If u is a vector function of x and $\bar{u}$ the corresponding[5] function after the change of variable, u is called contravariant if

$$\bar{u} = Ju, \tag{76}$$

and covariant when

$$\bar{u} = \bar{J}'u, \qquad u = J'\bar{u}. \tag{77}$$

If d_1, d_2 denote two independent variations so that $d_1(d_2x) = d_2(d_1x)$, then

$$\begin{aligned} d_2d_1\bar{x} = d_2(Jd_1x) &= d_2Jd_1x + Jd_1d_2x \\ &= J_\alpha d_2x S\nabla_\alpha d_1x + Jd_1d_2x. \end{aligned} \tag{78}$$

Hence second differentials are neither contra- nor co-variant.

If A is a matrix whose coordinates are functions of x, the bilinear differential form Sd_1xAd_2x when transformed becomes

$$Sd_1\bar{x}\bar{A}d_2\bar{x} = Sd_1xJ'\bar{A}Jd_2x$$

so that, if this form is invariant, that is, $Sd_1\bar{x}\bar{A}d_2\bar{x} = Sd_1xAd_2x$, we must have

$$A = J'\bar{A}J, \qquad A' = J'\bar{A}'J, \qquad Adx = J'\bar{A}d\bar{x}. \tag{79}$$

[5] As will be seen below, this does not necessarily mean merely the result of substituting $\bar{x}$ for x in the coordinates of u.

Hence when $\bar{A}$ is defined in this manner, Adx is a covariant vector. If by analogy with (78) we form a second differential of this vector and of $A'dx$, we get, using $d\bar{x} = Jdx$,

$$d_2(Ad_1x) = d_2J'\bar{A}d_1\bar{x} + J'd_2(\bar{A}d_1\bar{x})$$
$$d_1(A'd_2x) = d_1J'\bar{A}'d_2\bar{x} + J'd_1(\bar{A}'d_2\bar{x}).$$

From (75) $d_2J' = \nabla_\alpha S J_\alpha d_2x$, $d_1J' = \nabla_\alpha S d_1x J'_\alpha$; hence after a simple reduction

$$d_2(Ad_1x) + d_1(A'd_2x) = \nabla_\alpha S d_1x(J'\bar{A}'J_\alpha + J'_\alpha\bar{A}'J)d_2x + J'(d_2(\bar{A}d_1\bar{x}) + d_1(\bar{A}'d_2\bar{x}))$$
$$= \nabla_\alpha S d_1x(A'_\alpha - J'\bar{A}'_\alpha J)d_2x + J'(d_2(\bar{A}d_1\bar{x}) + d_1(\bar{A}'d_2\bar{x}))$$

which may be written

$$\begin{aligned} a &\equiv d_2(Ad_1x) + d_1(A'd_2x) - \nabla_\alpha S d_1x A'_\alpha d_2x \\ &= J'(d_2(\bar{A}d_1\bar{x}) + d_1(\bar{A}'d_2\bar{x}) - \nabla_\alpha S d_1\bar{x}\bar{A}'_\alpha d_2\bar{x}) \\ &= J'\bar{a} \end{aligned} \tag{80}$$

so that a is a covariant vector. This vector may also be written

$$a = d_2Ad_1x + d_1A'd_2x - \nabla_\alpha S d_1x A'_\alpha d_2x + (A + A')d_1d_2x.$$

Using a notation suggested by the Christoffel symbols we now write

$$\begin{aligned} [A; d_1x, d_2x] &= \tfrac{1}{2}(d_2Ad_1x + d_1A'd_2x - \nabla_\alpha S d_1x A'_\alpha d_2x) \\ &= \sum_{i\,j\,k} \tfrac{1}{2}\left(\frac{\partial a_{kj}}{\partial \xi_i} + \frac{\partial a_{ki}}{\partial \xi_j} - \frac{\partial a_{ji}}{\partial \xi_k}\right) d_1\xi_i d_2\xi_j e_k \end{aligned} \tag{81}$$

$$\left\{\begin{matrix} A \\ d_1x,\ d_2x \end{matrix}\right\} = (A + A')^{-1}(d_2Ad_1x + d_1A'd_2x - \nabla_\alpha S d_1x A'_\alpha d_2x) \tag{82}$$
$$= 2(A + A')^{-1}[A; d_1x, d_2x]$$

provided that $|A + A'| \neq 0$. If we now set

$$b = \left\{\begin{matrix} A \\ d_1x,\ d_2x \end{matrix}\right\} + d_1d_2x$$

and use the relation $(A + A')^{-1}J' = J^{-1}(\bar{A} + \bar{A}')^{-1}$, we have from (80)

$$\bar{b} = Jb \tag{83}$$

so that b is contravariant.

If we set $A = \frac{1}{2}(A + A') + \frac{1}{2}(A - A') = B + C$, we get from (81) and (80)

$$[A; d_1x, d_2x] = [B; d_1x, d_2x] + [C; d_1x, d_2x]$$
$$[B; d_1x, d_2x] + Bd_1d_2x = J'([\bar{B}; d_1\bar{x}, d_2\bar{x}] + \bar{B}d_1d_2\bar{x}) \tag{84}$$
$$[C; d_1x, d_2x] = J'[\bar{C}; d_1\bar{x}, d_2\bar{x}].$$

We shall require two transverses of the Christoffel matrices; these are defined by

$$Sb[A; a, c] = Sc[A; a, b]', \qquad Sb\left\{\begin{matrix} A \\ a, c \end{matrix}\right\} = Sc\left\{\begin{matrix} A \\ a, b \end{matrix}\right\}'$$

whence

$$2[A; a, b]' = \nabla_\alpha SaA'_\alpha b + A_\alpha bSa\nabla_\alpha - A_\alpha aSb\nabla_\alpha$$

$$\left\{\begin{matrix} A \\ a, b \end{matrix}\right\}' = [A; a, (A + A')^{-1}b]'. \tag{85}$$

To illustrate partial differentiation we shall consider functions which depend not only on x but also on a contravariant variable vector $u = \Sigma\omega_i e_i$. Since $\bar{u} = Ju$ and $J = \| \partial\bar{\xi}_i/\partial\xi_j \|$ is independent of u, we have

$$\frac{\partial}{\partial\omega_i} = \sum \frac{\partial\bar{\omega}_j}{\partial\omega_i}\frac{\partial}{\partial\bar{\omega}_j} = \sum \frac{\partial\bar{\xi}_j}{\partial\xi_i}\frac{\partial}{\partial\bar{\omega}_j}$$

$$\frac{\partial}{\partial\xi_i} = \sum \frac{\partial\bar{\xi}_j}{\partial\xi_i}\frac{\partial}{\partial\bar{\xi}_j} + \sum \frac{\partial\bar{\omega}_j}{\partial\xi_i}\frac{\partial}{\partial\bar{\omega}_j}.$$

Hence, if $\nabla' = \Sigma e_i \partial/\partial\omega_i$, $\bar{\nabla}' = \Sigma e_i \partial/\partial\bar{\omega}_i$, then

$$\nabla' = J'\bar{\nabla}', \qquad \nabla = J'\bar{\nabla} + d_u J' \cdot \bar{\nabla}' \tag{86}$$

where $d_u J' \equiv Su\nabla_\alpha \cdot J'_\alpha$. Here ∇' is covariant but ∇ is neither covariant nor contravariant, which means that formulae dependent on it will not usually be invariant in form under a change of variable. This difficulty is avoided as follows. If we combine (83) with (78) and replace $d_1 x$, $d_2 x$ by contravariant vectors a, b, then

$$d_b Ja = d_a Jb = J_\alpha bS\nabla_\alpha a = J\left\{\begin{matrix} A \\ a, b \end{matrix}\right\} - \left\{\begin{matrix} \bar{A} \\ \bar{a}, \bar{b} \end{matrix}\right\}. \tag{87}$$

Hence

$$d_a J'b = \left\{\begin{matrix} A \\ a, J'b \end{matrix}\right\}' - J'\left\{\begin{matrix} \bar{A} \\ \bar{a}, b \end{matrix}\right\}' \tag{87'}$$

and therefore

$$\nabla = J'\bar{\nabla} + \left\{\begin{matrix} A \\ u, \nabla' \end{matrix}\right\}' - J'\left\{\begin{matrix} \bar{A} \\ \bar{u}, \bar{\nabla}' \end{matrix}\right\}'$$

whence

$$D \equiv \nabla - \left\{\begin{matrix} A \\ u, \nabla' \end{matrix}\right\}' = J'\left(\bar{\nabla} - \left\{\begin{matrix} \bar{A} \\ \bar{u}, \bar{\nabla}' \end{matrix}\right\}\right) = J'\bar{D}; \tag{88}$$

D is therefore a covariant differential operator.

Let v be any contravariant vector and set $d_u v = Su\nabla_\alpha \cdot v_\alpha$; then, if f is any function of x and u,

$$d_v f \equiv (Sv\nabla_\alpha + Sd_u v\nabla'_\alpha)f_\alpha = (SvD_\alpha + S\delta_u v\nabla'_\alpha)f_\alpha \tag{89}$$

where $\delta_u v$ is the contravariant vector defined by

$$\delta_u v = d_u v + \left\{ \begin{matrix} A \\ u, v \end{matrix} \right\} = Vu. \tag{90}$$

The tensor corresponding to the matrix V is known as the covariant derivative of v.

8.14 **Functions of a variable matrix.** The general theory of analytic functions of a variable matrix $x = \| \xi_{ij} \|$ is co-extensive with that of n^2 scalar variables and hence is so general as to be void of properties peculiar to matrices. This follows immediately from the obvious relation

$$\xi_{ij} = \sum_{p=1}^{n} e_{pi} x e_{jp}$$

which expresses the (i, j) coordinate as a linear function of x.

The differential operator[6] corresponding to x is

$$\Delta = \left\| \frac{\partial}{\partial \xi_{ij}} \right\|. \tag{91}$$

It is often convenient to have a special notation for the transverse Δ' and when this is so we shall set

$$\partial = \left\| \frac{\partial}{\partial \xi_{ji}} \right\| = \Delta'. \tag{92}$$

These operators may stand after their operands and the same convention as was used for subscripts attached to ∇ will also be used for Δ and ∂ when necessary.

The fundamental property of Δ is

$$df = \operatorname{tr}(dx\Delta')f = \operatorname{tr}(dx\partial)f \tag{93}$$

where f is any function of x and $\operatorname{tr}(A)$ stands for the trace of the matrix A. This result follows immediately from

$$\operatorname{tr}(dx\Delta') \equiv \sum_i \sum_k d\xi_{ik} \frac{\partial}{\partial \xi_{ik}}.$$

[6] This operator first occurs in a paper by Taber (1890, (84)) who however did not make any systematic use of it. Macaulay in a tract published in 1893 (110) but written in 1887 used Δ consistently in applying quaternions to physical problems; he used the notation $\Box$ for Δ. Later Born (385) used the same operator to great effect in his theory of quantum matrices. Turnbull (436) uses Ω for Δ'.

8.15 **Differentiation formulae.** We collect here the principal formulae of differentiation; in each case the operand is x or x' and the dummy subscript is omitted except when the meaning is ambiguous without it. To simplify the expressions we set α and β for the traces of the matrices a and b, and ξ_r for the trace of x^r.

$$\begin{aligned} \Delta axb &= a'b = a'xb'\Delta, \qquad \Delta ax'b = \alpha b = bx'a\Delta \\ \Delta' axb &= \alpha b = bxa\Delta', \qquad \Delta' ax'b = a'b = a'x'b'\Delta' \end{aligned} \tag{94}$$

$$\begin{aligned} \Delta x^r &= x^{r-1} + x'x^{r-2} + (x')^2x^{r-3} + \cdots + (x')^{r-1} = \Delta'(x')^r \\ x^r\Delta &= x^{r-1} + x^{r-2}x' + x^{r-3}(x')^2 + \cdots + (x')^{r-1} = (x')^r\Delta' \\ \Delta' x^r &= nx^{r-1} + \xi_1 x^{r-2} + \xi_2 x^{r-3} + \cdots + \xi_r = x^r\Delta' \end{aligned} \tag{95}$$

$$\begin{aligned} \Delta \operatorname{tr}(axb) &= \Delta \operatorname{tr}(bax) = a'b' \\ \partial \operatorname{tr}(axb) &= ba \end{aligned} \tag{96}$$

$$\Delta \operatorname{tr}(x^r) \equiv \Delta\xi_r = r(x')^{r-1}, \qquad \partial \operatorname{tr}(x^r) = rx^{r-1}. \tag{97}$$

$$\begin{aligned} \operatorname{tr}(\Delta)axb &= ab = \operatorname{tr}(\partial)axb \\ \operatorname{tr}(\Delta)x^r &= rx^{r-1} = \operatorname{tr}(\partial)x^r \end{aligned} \tag{98}$$

$$\begin{aligned} \partial \operatorname{tr}\begin{Bmatrix} x & a \\ r & s \end{Bmatrix} &= (r+s)\begin{Bmatrix} x & a \\ r-1 & s \end{Bmatrix} = \operatorname{tr}(\partial)\begin{Bmatrix} x & a \\ r & s \end{Bmatrix} \\ \Delta \operatorname{tr}\begin{Bmatrix} x & a \\ r & s \end{Bmatrix} &= (r+s)\begin{Bmatrix} x' & a' \\ r-1 & s \end{Bmatrix}. \end{aligned} \tag{99}$$

$$\begin{aligned} \Delta'_\alpha x_\alpha^{r+1} - x\Delta'_\alpha x_\alpha^r &= \operatorname{tr}(x^r) \equiv \xi_r \\ (\Delta_\alpha x_\alpha - x'\Delta_\alpha)\, x_\alpha^r &= (x')^r. \end{aligned} \tag{100}$$

The proofs of these formulae are all very similar and we shall consider here only the most important leaving the remainder to the reader. If $a = \| a_{ij} \|$, $b = \| b_{ij} \|$, then

$$\Delta axb = \left\| \frac{\partial}{\partial \xi_{ip}} a_{pq}\xi_{qr}b_{rj} \right\| = \left\| \sum_p a_{pi}b_{pj} \right\| = a'b;$$

hence also

$$a'x'b'\Delta' = (\Delta bxa)' = (b'a)' = a'b.$$

The remaining parts of (94) follow in the same way. It follows also from (94) that

$$\begin{aligned} \Delta_\alpha x_\alpha^r &= \Delta_\alpha x_\alpha x^{r-1} + \Delta_\alpha x x_\alpha^{r-1} = x^{r-1} + \Delta_\alpha x x_\alpha x^{r-2} + \Delta_\alpha x^2 x_\alpha^{r-2} \\ &= x^{r-1} + x'x^{r-2} + \Delta_\alpha x^2 x_\alpha x^{r-3} + \Delta_\alpha x^3 x_\alpha^{r-3}, \end{aligned}$$

and so on; the remaining parts of (95) follow from (94) in the same way.

To prove (96) we notice first that tr (ab) = tr (ba) and hence in tr $(a_1a_2 \cdots a_r)$ the factors may be permuted cyclically. Then, if $c = \| c_{ij} \|$,

$$\partial \operatorname{tr}(cx) = \sum_{i,j} e_{ij} \frac{\partial}{\partial \xi_{ji}} \sum_{p,q} c_{pq}\xi_{qp} = \sum c_{ij}e_{ij} = c.$$

Formula (97) follows by repeated application of (96); thus

$$\begin{aligned}\Delta \operatorname{tr}(x^r) &= \Delta_\alpha \operatorname{tr}(x_\alpha x^{r-1}) + \Delta_\alpha \operatorname{tr}(x_\alpha^{r-1}x) \\ &= (x')^{r-1} + \Delta_\alpha \operatorname{tr}(x_\alpha x^{r-1}) + \Delta_\alpha \operatorname{tr}(x_\alpha^{r-2}x^2) \\ &= (x')^{r-1} + (x')^{r-1} + \cdots \\ &= r(x')^{r-1}.\end{aligned}$$

The remaining formulae are proved in the same way.

If $\Sigma\alpha_r\lambda^r$ is a scalar power series and $f(x) = \Sigma\alpha_r x^r$, then from (97) and (98)

$$\partial \operatorname{tr}(f(x)) = \Sigma r\alpha_r x^{r-1} = f'(x) = \operatorname{tr}(\partial) f(x)$$

so that the operators ∂ tr ( ) and tr (∂) have the same effect on such functions. Similarly, if $F(x) = \Sigma\alpha_r \begin{Bmatrix} x & a \\ r & s_r \end{Bmatrix}$, it follows from (99) that

$$\partial \operatorname{tr}(F) = \operatorname{tr}(\partial)F = \Sigma(r + s_r) \begin{Bmatrix} x & a \\ r-1 & s_r \end{Bmatrix}. \tag{101}$$

8.16 As an illustration of the application of the formulae of the preceding section we shall give some parts of the theory of quantum matrices which are applicable to matrices of finite order.

Let $q_1, q_2, \cdots, q_f; p_1, \cdots, p_f$ be the coordinates of a dynamical system and $\mathfrak{H}$ the Hamiltonian function; these coordinates satisfy the system of ordinary partial differential equations

$$\dot{q}_i = \frac{\partial \mathfrak{H}}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial \mathfrak{H}}{\partial q_i} \qquad (i = 1, 2, \cdots, f). \tag{102}$$

We may suppose that $f = n^2$, a perfect square; for, if $(n-1)^2 < f < n^2$, we can introduce $2(n^2 - f)$ additional coordinates $q_{f+1}, p_{f+1}, \cdots, q_{n^2}, p_{n^2}$ which do not occur in $\mathfrak{H}$ so that these variables equated to constants are solutions of the extended system. When this is done, we can order the q's and p's in square arrays $\| q_{ij} \|$, $\| p_{ij} \|$ in such a way that p_{ij} corresponds to q_{ji} for all i and j. Equation (102) then becomes

$$\dot{q}_{ij} = \frac{\partial \mathfrak{H}}{\partial p_{ji}}, \qquad \dot{p}_{ij} = -\frac{\partial \mathfrak{H}}{\partial q_{ji}}$$

or, if the matrices $\| q_{ij} \|$ and $\| p_{ji} \|$ are denoted by q and p and the corresponding transverse differential operators by ∂_q and ∂_p,

$$\dot{q} = \partial_p \mathfrak{H}, \qquad \dot{p} = \partial_q \mathfrak{H}. \tag{103}$$

If we transform (103) by the substitution

$$q = \epsilon^{wt}Q\epsilon^{-wt}, \qquad Q = \| Q_{ij} \|, \qquad p = \epsilon^{wt}P\epsilon^{-wt}, \qquad P = \| P_{ij} \|,$$

where w is a constant matrix, we get

$$\dot{q} = \epsilon^{wt}(\dot{Q} + wQ - Qw)\epsilon^{-wt}, \qquad \dot{p} = \epsilon^{wt}(\dot{P} + wP - Pw)\epsilon^{-wt}.$$

Also, if t is kept constant,

$$\operatorname{tr}(dq\partial_q) = \operatorname{tr}(\epsilon^{wt}dQ\epsilon^{-wt}\partial_q) = \operatorname{tr}(dQ\epsilon^{-wt}\partial_q\epsilon^{wt})$$

with a similar relation for p. Hence

$$\partial_q = \epsilon^{wt}\partial_Q\epsilon^{-wt}, \qquad \partial_p = \epsilon^{wt}\partial_P\epsilon^{-wt}.$$

Using these results in (103) we get

$$(104) \qquad (Q + wQ - Qw) = \partial_P\mathfrak{H}, \qquad (\dot{P} + wP - Pw) = -\partial_Q\mathfrak{H},$$

$\mathfrak{H}$ being expressed in terms of P and Q and, if necessary, also t. Now from (96)

$$Qw - wQ = \partial_P\operatorname{tr}[w(PQ - QP)] \qquad -(Pw - wP) = \partial_Q\operatorname{tr}[w(PQ - QP)]$$

and hence, if

$$(105) \qquad \mathfrak{K} \equiv \mathfrak{H} + \operatorname{tr}[w(PQ - QP)] = \mathfrak{H} + \operatorname{tr}[w(pq - qp)],$$

we have in place of (103)

$$(106) \qquad \dot{Q} = \partial_P\mathfrak{K}, \qquad \dot{P} = -\partial_Q\mathfrak{K},$$

so that the transformation is canonical.

If $\dot{Q} = 0 = \dot{P}$ in (104), then

$$wQ - Qw = \partial_P\mathfrak{H}, \qquad wP - Pw = -\partial_Q\mathfrak{H}$$

or on restoring the exponential factor

$$(107) \qquad wq - qw = \partial_p\mathfrak{H}, \qquad wp - pw = -\partial_q\mathfrak{H}.$$

When $\mathfrak{H}$ is given, these are algebraic equations which can be solved for p and q; the solution will of course generally contain arbitrary parameters.

Under the same assumptions (106) becomes

$$(108) \qquad \partial_P\mathfrak{K} = 0 = \partial_Q\mathfrak{K},$$

and if P, Q are independent variables, the only solution is $\mathfrak{K}$ = constant and the only solution for $\mathfrak{H}$ in (107) then has the form

$$\mathfrak{H}^* = \operatorname{tr}[w(pq - qp)]$$

apart from an additive constant. Equation (108) may then be written

$$(109) \qquad \partial_p(\mathfrak{H} - \mathfrak{H}^*) = 0 = \partial_q(\mathfrak{H} - \mathfrak{H}^*).$$

Now, if $r_1, r_2, \cdots, r_m$ are the parameters in the solution of (107) we have

$$\frac{\partial(\mathfrak{H} - \mathfrak{H}^*)}{\partial r_k} = \sum \frac{\partial p_{ij}}{\partial r_k} \frac{\partial(\mathfrak{H} - \mathfrak{H}^*)}{\partial p_{ij}} + \sum \frac{\partial q_{ij}}{\partial r_k} \frac{\partial(\mathfrak{H} - \mathfrak{H}^*)}{\partial q_{ij}}$$

which vanishes in virtue of (109). Hence, if $\mathfrak{H}$ is expressed in terms of r_1, $r_2, \cdots, r_m$ by using the solutions of (107), it will differ by an additive constant at most from $-$ tr $[w(pq - qp)]$.