

A Matrix Form of Taylor's Theorem.

By Professor H. W. TURNBULL.

(Received 1st October 1929. Read 2nd November 1929.)

INTRODUCTION.

The following pages continue a line of enquiry begun in a work *On Differentiating a Matrix*, (*Proceedings of the Edinburgh Mathematical Society* (2) 1 (1927), 111-128), which arose out of the Cayley operator $|\Omega| \equiv \left| \frac{\partial}{\partial x_{ji}} \right|$, where x_{ij} is the ij^{th} element of a square matrix $[x_{ij}]$ of order n , and all n^2 elements are taken as independent variables. The present work follows up the implications of Theorem III in the original, which stated that

$$\Omega s(X^r) = r X^{r-1}, \quad \Omega = \left[\frac{\partial}{\partial x_{ji}} \right], \quad X = [x_{ij}],$$

where $s(X^r)$ is the sum of the principal diagonal elements in the matrix X^r . This is now written $\Omega_s X^r = r X^{r-1}$ and Ω_s is taken as a fundamental operator analogous to ordinary differentiation, but applicable to matrices of any *finite* order n . In particular it leads to a matrix form of Taylor's Theorem (§ 8),

$$f(X + A) = f(X) + \Omega_s A f(X) + \frac{1}{2!} (\Omega_s A)^2 f(X) + \dots$$

where A is an arbitrary constant matrix *which need not commute with the variable X* . The function $f(X)$ is a scalar function of X , and is not a general matrix function: even so, $f(X + A)$ is essentially a function of two matrices X and A , and therefore is vastly more complicated than $f(X)$ itself.

The question of change of variable arises and leads to various results which generalize on the formulae $y = f(x)$, $\frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx}$. Incidentally it appears that the operator $\Omega X - X \Omega$ is an absolute matrix differential invariant for scalar functional transformations $Y = f(X)$, from a matrix X to a matrix Y of n^2 independent variables.

In the last paragraph the analogies and the disparities between the present calculus and that of q -numbers in quantum algebra receive attention.

§ 1. CYCLIC PROPERTIES OF THE DIAGONAL IN A MATRIX.

Let $X = [x_{ij}]$ be a square matrix of order n whose elements x_{ij} are scalar; then the sum s_X of its leading diagonal elements is Σx_{ii} . If $Y = [y_{ij}]$ is a matrix of the same order, we may form the product matrix $XY = [\Sigma x_{ik} y_{kj}]$; and the sum of its leading diagonal elements is given by

$$s_{XY} = \Sigma x_{ij} y_{ji}, \dots \dots \dots (1)$$

summed for both i and j from 1 to n . Adopting the double suffix convention of tensor analysis we can write

$$s_{XY} = x_{ij} y_{ji} = s_{YX} \dots \dots \dots (2)$$

Under such a convention the further expression $x_{ij} y_{jk} z_{ki}$ denotes a triple sum, with i, j, k each running from 1 to n : and manifestly $y_{ij} z_{jk} x_{ki}$ denotes exactly the same triple sum, whereas $x_{ik} y_{kl} z_{ij}$ denotes a double sum which is the ij^{th} representative element in the product matrix XYZ . We infer the following cyclic property of the triple product matrix XYZ :

$$s_{XYZ} = s_{YZX} = s_{ZXY} \dots \dots \dots (3)$$

Likewise for a matrix product of r factors $XY \dots Z$, the scalar quantity $s_{XY \dots Z}$ is unchanged by a cyclic permutation of these factors. This quantity is given by

$$s_{XY \dots Z} = x_{ij} y_{jk} \dots z_{qi} \dots \dots \dots (4)$$

and the summation on the right consists of n^r terms.

For the purpose of matrix differentiation this scalar function s_X of the elements in the diagonal of the matrix X is very important. Let it be called the *spur* of X , and also be denoted by $s(X)$. Thus

$$s(XY \dots Z) = s(Y \dots ZX) = \dots = s(ZXY \dots). \dots \dots (4')$$

The purpose of these r equivalent ways of writing the same function of a product matrix will appear in the sequel. But for the present it is worth noting that, in the ordinary differential calculus, we could differentiate a product uvw , of functions of a single variable ξ , as follows:

$$\frac{d(uvw)}{d\xi} = \frac{du}{d\xi} \cdot vw + \frac{dv}{d\xi} \cdot wu + \frac{dw}{d\xi} \cdot uv. \dots \dots (5)$$

This may look artificial, but it serves to illustrate the more difficult non-commutative case. It suggests a rule: *to differentiate a product*

of r factors, form a sum of r terms by bringing each different factor cyclically to the fore, and differentiate the first factor of each term, alone.

The transposition law. If $X' = [x_{ji}]$ denotes the transposed matrix of X , the corresponding functions, such as $s(X'Y')$, are obtained by transposing suitable suffixes. For example

$$\left. \begin{aligned} s_{X'Y'} &= x_{ji} y_{ij} = s_{XY}, & s_{X'Y} &= x_{ji} y_{ji} = s_{XY'} \\ s_{ZYX} &= s_{YXZ} = s_{XZY} = s_{Z'Y'X'} = s_{(XYZ)'} \end{aligned} \right\} \dots (6)$$

There are obviously two ways round a circuit: and if we compare (3) with (6), we see that the two ways round correspond with the two modes of writing a matrix—the direct and the transposed. The same is true for a product $XY\dots Z$ of r factors.

The addition law. If $Y = [y_{ij}]$ and $Z = [z_{ij}]$ then $Y + Z = [y_{ij} + z_{ij}]$, so that

$$s(Y + Z) = y_{ii} + z_{ii} = s(Y) + s(Z). \dots (7)$$

Hence the process of forming the spur of a matrix is distributive with the process of addition. Again, if a is a scalar number, we have

$$s(aY) = a y_{ii} = a s(Y): \dots (8)$$

and if I is the unit matrix and $f(X)$ is a polynomial of a matrix X with constant scalar coefficients

$$f(X) = a_0 I + a_1 X + a_2 X^2 + \dots + a_p X^p, \dots (9)$$

it follows that

$$s(f(X)) = a_0 n + a_1 s(X) + a_2 s(X^2) + \dots + a_p s(X^p). \dots (10)$$

§ 2. THE OPERATORS $\Omega, \Omega', \Omega_s, \Omega'_s$.

If we treat the n^2 elements x_{ij} as independent variables and form the matrix differential operators

$$\Omega = \left[\frac{\partial}{\partial x_{ji}} \right], \quad \Omega' = \left[\frac{\partial}{\partial x_{ij}} \right], \quad \dots (1)$$

then we interpret ΩY by the product rule —

$$\Omega Y = \left[\frac{\partial}{\partial x_{ji}} \right] [y_{ij}] = \left[\frac{\partial}{\partial x_{kj}} y_{kj} \right], \quad \dots (2)$$

and we interpret the effect of operating on a scalar quantity f by

$$\Omega f = \left[\frac{\partial f}{\partial x_{ji}} \right]. \quad \dots (3)$$

The suffix k in (2) is a dummy, denoting a sum of n terms in the

ij^{th} element of the matrix. In particular, let us operate on $s(XY)$, which may be written $x_{pq} y_{qp}$ (§ 1 (2)). Then

$$\frac{\partial s(XY)}{\partial x_{ji}} = \frac{\partial x_{pq}}{\partial x_{ji}} \cdot y_{qp} + x_{pq} \frac{\partial y_{qp}}{\partial x_{ji}} = y_{ij} + x_{pq} \frac{\partial y_{qp}}{\partial x_{ji}}, \dots\dots (4)$$

for all the other terms disappear. In fact $\partial x_{pq} / \partial x_{ji}$ is zero unless $p = j, q = i$. Hence by forming the whole matrix, we have

$$\Omega s(XY) = \left[\frac{\partial s(XY)}{\partial x_{ji}} \right] = Y + \left[x_{pq} \frac{\partial y_{qp}}{\partial x_{ji}} \right] \dots\dots\dots (5)$$

Similarly, if $A = [a_{ij}]$ is a matrix of constant elements, we have

$$\Omega s(AY) = 0 + \left[a_{pq} \frac{\partial y_{qp}}{\partial x_{ji}} \right] = \left[a_{pq} \frac{\partial y_{qp}}{\partial x_{ji}} \right], \dots\dots\dots (6)$$

and, for the transposed matrix X' , we have

$$\Omega s(X'Y) = \left[\frac{\partial x_{qp} y_{qp}}{\partial x_{ji}} \right] = Y' + \left[x_{qp} \frac{\partial y_{qp}}{\partial x_{ji}} \right] \dots\dots\dots (7)$$

It is preferable to rewrite (5) as

$$\Omega s(XY) = Y + \left[\frac{\partial y_{pq}}{\partial x_{ji}} x_{qp} \right], \dots\dots\dots (5')$$

where the last matrix is identically equal to the original $\left[x_{pq} \frac{\partial y_{qp}}{\partial x_{ji}} \right]$.

More generally, if $Z = UVW$, and $Z = [z_{ij}] = [u_{iq} v_{qr} w_{rj}]$, we have

$$\Omega s(Z) = \left[\frac{\partial u_{pq} v_{qr} w_{rp}}{\partial x_{ji}} \right] = \left[\frac{\partial u_{pq}}{\partial x_{ji}} \cdot v_{qr} w_{rp} \right] + \left[\frac{\partial v_{pq}}{\partial x_{ji}} \cdot w_{qr} u_{rp} \right] + \left[\frac{\partial w_{pq}}{\partial x_{ji}} \cdot u_{qr} v_{rp} \right] (8)$$

and this puts into evidence the *cyclic rule for differentiating a product*, given in § 1. If, for example, Y, U, V, W are each equal to X , the only non-zero terms in (5') and (8) occur where $p = j, q = i$. Consequently these formulae give

$$\begin{aligned} \Omega s(X^2) &= X + [x_{ij}] = 2X, \\ \Omega s(X^3) &= XX + XX + XX = 3X^2, \end{aligned}$$

and in general,¹

$$\Omega s(X^r) = rX^{r-1}. \dots\dots\dots (9)$$

One object of the present communication is to extend this formula (9) to the case when r is not merely a positive integer

¹ Cf. *Proc. Edinburgh Math. Soc.* (2) **1** (1928), 111–128 (128).

but is any rational number, and X^r is replaced by any regular function $\phi(X)$.

Next, we must modify the cyclic rule for differentiating a product, so as to include *transposed* factors, such as occur in (7). Comparing (5) with (7), we note that the cyclic rule is still at work, provided only that when a transposed factor is differentiated, *all the other factors are transposed in the resulting term*.

For, example, if $Z = X' V W = [x_{qi} v_{qr} w_{rj}]$, then

$$\begin{aligned} \Omega s(Z) &= \left[\frac{\partial x_{qp} v_{qr} w_{rp}}{\partial x_{ji}} \right] = [v_{jr} w_{ri}] + \left[\frac{\partial v_{pq}}{\partial x_{ji}} \cdot w_{qr} x_{pr} \right] + \left[\frac{\partial w_{pq}}{\partial x_{ji}} \cdot x_{rq} v_{rp} \right] \\ &= (VW)' + \text{etc.} \end{aligned} \tag{10}$$

Here the effect of differentiating the factor X' is to reverse the VW , and the term could be written $W'V'$. It is impossible to state the other two terms more precisely, unless the characters of V and W are defined. Let us confine ourselves to matrix products of factors consisting entirely of X, X' and constants A .

Examples. (i) $\Omega s(X'X) = 2X'$.

For, by (7), the operation on X' deletes X' and transposes the other factor X to X' ; and by (5) the operation on X deletes X without further alteration.

$$\begin{aligned} \text{(ii) } \Omega s(X'AXBX') &= (AXBX')' + (BX'X'A) + (X'AXB)' \\ &= XB'X'A' + BX'^2A + B'X'A'X. \end{aligned}$$

In this, two of the five factors in the operand are constants, giving zero terms, by (6). The cyclic rule is at work, together with transposition of the cycle, for the two cases X' . The practical point to bear in mind is to begin the cycle with the factor immediately after the deleted factor.

Change of notation to Ω_s . Hitherto there have been separate operations—that of forming the spur $s(Y)$ of a matrix Y , and the operation Ω . We can merge these two processes and call the joint operation Ω_s : from a matrix Y it generates a *derived* matrix Y_1 .

Hence by definition, if $Y = [y_{ij}]$, $\Omega = \left[\frac{\partial}{\partial x_{ji}} \right]$,

$$\Omega_s Y = \Omega s(Y) = \left[\frac{\partial y_{pp}}{\partial x_{ji}} \right] = Y_1. \tag{11}$$

There is an analogous transposed matrix operator Ω'_s defined by

$$\Omega'_s Y = \Omega'_s (Y) = \left[\frac{\partial y_{pp}}{\partial x_{ij}} \right] \dots\dots\dots(12)$$

but it does not appear to have the fundamental importance of Ω_s .

§ 3. FUNCTIONS OF MATRICES AND THEIR DERIVATES.

We must now distinguish between *three* types of function occurring in non-commutative algebra:

- I $f(X)$ the scalar function of a single argument X ,
- II $F(X)$ the matrix function of a single argument X ,
- III $\sigma(X)$ the scalar of X .

By I is meant a function composed of one matrix argument X together with scalar coefficients. It may be a polynomial, a rational, a regular function, and so on. For example

$$\exp(X) = 1 + X + \frac{X^2}{2!} + \frac{X^3}{3!} + \dots + \frac{X^r}{r!} + \dots \dots\dots(1)$$

is a scalar function. Strictly speaking $f(X)$ is a function of *two* matrices, X itself together with the unit matrix I . The unit in the first term of the expansion of $\exp X$ is understood to signify the unit matrix.

By II is meant a function involving *several* matrices, where X alone is the argument and all others are constants. A capital letter F will be used to distinguish this type of function from the scalar function. For example

$$F(X) = AX^2 + XBX + X^2C \dots\dots\dots(2)$$

is a particular quadratic matrix function of X .

By III is meant a scalar quantity, such as $s(X^2)$, whose value is completely determined by the value of the matrix X . The spur of X is a particular scalar of X .

It will be seen that, in the extreme case when $n = 1$, all matrices revert to ordinary scalar algebra, and there is no *raison d'être* for distinguishing between these three types of function, any more than between a matrix and its transposed.

We may obviously extend the above notation to include functions of several matrix arguments, such as $f(X, Y)$, $F(X, Y)$, $\sigma(X, Y)$. More generally there could be functions $F(X, Y \dots, x_{ij}, \dots)$ where the scalar variables x_{ij} occur among the scalar coefficients as

well as implicitly in the matrix arguments. Such cases however are hereafter left aside.

It may be observed that type I is a special case of II—when all matrix constants are scalar multiples of the unit matrix.

We now come to certain fundamental theorems.

THEOREM I. *The matrix derivate of a scalar function $f(X)$ is the ordinary derived function $f'(X)$, which is also the derivate of $f(X')$. Or*

$$\Omega_s f(X) = \Omega_s f(X') = f'(X). \quad \dots\dots\dots(3)$$

Proof for the case of a polynomial. Since $\Omega(Y + Z) = \Omega Y + \Omega Z$ and $\Omega_s X^r = rX^{r-1}$, the result follows by operating with Ω on $f(X)$ in § 1 (10). Thus

$$\begin{aligned} f(X) &= a_0 + a_1 X + \dots + a_p X^p, \\ \Omega_s f(X) &= a_1 + 2a_2 X + \dots + pa_p X^{p-1} = f'(X) \end{aligned}$$

in the ordinary sense of $f'(X)$.

The second part of the theorem follows by the cyclic reversal law of § 2 (10); or simply because the spur of $f(X')$ is the same as that of $f(X)$.

THEOREM II. *The matrix derivate of a matrix polynomial function $F(X)$ of order p is a matrix polynomial of order $p - 1$.*

This follows immediately from the law of cyclic differentiation. For example, the derivate of the *general linear polynomial*

$$\Sigma AXB + C$$

is ΣBA . That of the *general quadratic polynomial*

$$\Sigma AXBXC + \Sigma DXE + F$$

is the linear polynomial

$$\Sigma BXCA + \Sigma CAXB + \Sigma ED.$$

THEOREM III. *The matrix derivate of a matrix polynomial $F(X, X')$ of order p in the two variables X and X' is another such polynomial of order $p - 1$.*

The proof again follows from the cyclic laws. For example

$$\begin{aligned} \Omega_s (AXX' + X'BX) &= X'A + (AX)' + (BX)' + X'B \\ &= X'(A + A' + B + B'). \end{aligned}$$

THEOREM IV. *The special matrix geometrical progression*

$$AX^{r-1} + XAX^{r-2} + \dots + X^{r-i}AX^{i-1} + \dots + X^{r-1}A$$

is the matrix derivate of AX^r .

This is a particular application of Theorem II, presenting no difficulties, once the cyclic law of differentiation is grasped. We have

$$\left. \begin{aligned} \Omega_s AX &= A, \\ \Omega_s AX^2 &= AX + XA, \\ \Omega_s AX^3 &= AX^2 + XAX + X^2A, \\ \dots\dots\dots \\ \Omega_s AX^r &= AX^{r-1} + XAX^{r-2} + \dots + X^{r-1}A \end{aligned} \right\} \dots\dots\dots(4)$$

Further results. A few more examples of this operation may now be quoted, before proceeding with the general theory. There is no difficulty in establishing the following formulae, where r is a positive integer:

$$\left. \begin{aligned} \Omega_s AX' &= A', \\ \Omega_s AX'^2 &= XA' + A'X, \\ \dots\dots\dots \\ \Omega_s AX'^r &= X^{r-1}A' + X^{r-2}A'X + \dots + A'X^{r-1} \end{aligned} \right\} (5)$$

$$\left. \begin{aligned} \Omega_s (XX') &= \Omega_s (X'X) = 2X', \\ \Omega_s (X^r X'^r) &= 2(X^{r-1}X'^r + X^{r-2}X'^rX + \dots + X'^rX^{r-1}) = \Omega_s (X'^r X^r), \\ \Omega_s (XX')^r &= 2rX'(XX')^{r-1} = 2r(X'X)^{r-1}X' = \Omega_s (X'X)^r \end{aligned} \right\} (6)$$

Formulae involving $\Omega' = \left[\frac{\partial}{\partial x_{ij}} \right]$:

$$\Omega_s' AX = A', \quad \Omega_s' AX^2 = X'A + AX', \text{ etc. } \dots\dots\dots(7)$$

Ω -matrices. All these matrices on the left are of type $\Omega\sigma(X, \dots)$, where Ω or Ω' acts on a scalar of X, X', A . Let such matrices be called Ω -matrices. The formulae suggest that a large class of functions $F(X, X')$ can be expressed as Ω matrices. In particular the last result of (6) shews that the recurring product

$$XX'XX' \dots X$$

of an odd number of factors can be so expressed.

§ 4. THE SPECIAL GEOMETRICAL PROGRESSION.

The matrix polynomial

$$G_r(X, A, X) = AX^{r-1} + XAX^{r-2} + \dots + X^i AX^{r-i-1} + \dots + X^{r-1}A \dots(1)$$

is called a special geometrical progression, whose fore and after ratios are equal to X and X^{-1} . It is obviously allied to a more general function

$$A + XAY + X^2AY^2 + \dots + X^iAY^i + \dots \dots\dots(2)$$

which admits of satisfactory treatment analogous to the ordinary elementary theory. To derive the special series we replace A by AX^{r-1} and Y by X^{-1} , taking the first r terms of the general series. It will be seen that this special progression constantly occurs in the theory of Ω differentiation. An example of the general progression, proved in the original communication, was

$$\Omega' X^r = X^{r-1} + X' X^{r-2} + \dots + X'^{r-1} \dots\dots\dots(3)$$

THEOREM V. *Provided that X satisfies certain conditions to be specified, the only value of A which satisfies the equation*

$$AX^{r-1} + XAX^{r-2} + \dots + X^{r-1}A = I$$

is $A = \frac{1}{r} X^{1-r}$.

Proof. Multiply the given equation, fore and aft in turn, by X , and subtract. Hence

$$AX^r - X^rA = IX - XI = 0.$$

So A commutes with X^r and is therefore a scalar function of X^r , provided that X^r itself satisfies no rational integral equation of order lower than n .¹ Consequently A is a scalar function of X and commutes with all integral powers of X . The equation can now be written $AX^{r-1} + AX^{r-2} + \dots + AX^{r-1} = I$, or $rAX^{r-1} = I$. Hence, if X is non-singular, $A = \frac{1}{r} X^{1-r}$.

These limitations for the value of X may be stated otherwise: X may have repeated, but not zero, latent roots: and if a root λ occurs exactly ν times, its corresponding principal minor L in the canonical form of X must contain $\nu - 1$ units in the over-principal diagonal. For example

$$X = H \begin{bmatrix} L & & & \\ & M & & \\ & & N & \\ & & & \end{bmatrix} H^{-1}, \quad L = \begin{bmatrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \lambda & 1 \\ & & & \lambda \end{bmatrix}.$$

§ 5. RATIONAL FUNCTIONS OF X .

The formula $\Omega_s X^r = rX^{r-1}$ has been proved when r is a negative integer², but we need a more general result embodying the cyclic

¹ Cullis. *Matrices and Determinoids*, Vol. 3 (Cambridge, 1925) 483.

² *Proc. Edin. Math. Soc.*, loc. cit. 125.

laws of § 1. In § 2 (5) let $Y = BX^{-1}$ where B is constant. Then

$$\Omega_s XBX^{-1} = BX^{-1} + \left[x_{pq} \frac{\partial y_{qp}}{\partial x_{ji}} \right]. \dots\dots\dots(1)$$

But $s(XBX^{-1}) = s(BX^{-1}X) = s(B)$; and hence the left hand side $\Omega_s XBX^{-1}$ vanishes. On the right, in (1), the last term can be written $\left[\frac{\partial}{\partial x_{ji}} c_{pq} y_{qp} \right]$, where $[c_{pq}]$ is a constant matrix C , provided that, after differentiation, C is replaced by X . Also

$$c_{pq} y_{qp} = s(CY) = s(CBX^{-1}).$$

Hence we replace (1) by the result

$$0 = BX^{-1} + \Omega_s(CBX^{-1}), \quad C = X. \dots\dots\dots(2)$$

But CB is a constant (for the operator Ω_s) and can be written as A , so that $B = C^{-1}A = X^{-1}A$ outside the operation. Hence

$$\Omega_s AX^{-1} = -X^{-1}AX^{-1}, \dots\dots\dots(3)$$

where A is a constant.¹ In particular, if A is the unit matrix, $\Omega_s X^{-1} = -X^{-2}$. We may now adjoin the following theorem to theorem IV:

THEOREM VI

$$\left. \begin{aligned} \Omega_s AX^{-1} &= -X^{-1}AX^{-1}, \\ \Omega_s AX^{-2} &= -X^{-2}AX^{-1} - X^{-1}AX^{-2}, \\ \dots\dots\dots \\ \Omega_s AX^{-r} &= -X^{-r}AX^{-1} - X^{-r+1}AX^{-2} - \dots \\ &\quad - X^{-r+i}AX^{-i-1} - \dots - X^{-1}AX^{-r} \end{aligned} \right\} \dots\dots\dots(4)$$

Proof. The first of these has just been proved: the rest follow by induction. Case r follows from case $r - 1$ by substituting $Y = BX^{-r}$ in § 2 (5). Proceeding as before we have

$$\Omega_s XBX^{-r} = BX^{-r} + \Omega_s CBX^{-r}, \quad C = X.$$

But $s(XBX^{-r}) = s(BX^{-r+1})$ by the cyclic law. Hence $\Omega_s XBX^{-r}$ is $-X^{-r+1}BX^{-1} + \dots - X^{-1}BX^{-r+1}$ by the formula for case $r - 1$. Putting $A = CB$ as before the result follows.

In particular if $A = I$, then

$$\Omega_s X^{-r} = -rX^{-r+1} \dots\dots\dots(5)$$

¹ Cf. Weierstrass, *Monatsb. d. Berliner Acad.* (1858), p. 214. Frobenius, *Crelle* 84 (1878), 17, where a similar formula occurs, reached from another angle.

COROLLARY. *If X' replaces X throughout on the left of (4), then X remains unaltered on the right, but A becomes A' . Also*

$$\Omega_s X'^{-r} = -rX^{-r+1} \dots\dots\dots(6)$$

The proofs are straightforward.

These results establish THEOREM I for the case when $f(X)$ is a rational scalar function $\phi(X)/\psi(X)$, provided that the function can be developed as an ascending series

$$a_{-r}X^{-r} + \dots + a_0 + a_1X + \dots + a_iX^i + \dots$$

which is finite or, when infinite, is convergent.

§ 6. FRACTIONAL INDICES. CHANGE OF VARIABLE.

Let r be a positive integer and $Y = X^r$, or $[y_{ij}] = [x_{ij}]^r$. With the double suffix summation convention this gives

$$y_{ij} = x_{ia_1} x_{a_1a_2} \dots x_{a_{r-2}a_{r-1}} x_{a_{r-1},j} \dots\dots\dots(1)$$

The n^2 quantities y_{ij} may be taken to be n^2 new independent variables, because their Jacobian $\partial(y)/\partial(x)$ —a determinant of n^2 rows and columns—will be shewn not to vanish identically. We may take Y to be a new independent variable in place of X , and write

$$X = [x_{ij}], \quad Y = [y_{ij}], \quad \Omega = \left[\frac{\partial}{\partial x_{ji}} \right], \quad \omega = \left[\frac{\partial}{\partial y_{ji}} \right]. \dots\dots\dots(2)$$

THEOREM VII. *The following relation holds*

$$\Omega = X^{r-1}\omega + X^{r-2}\omega X + \dots + X^{r-i}\omega X^{i-1} + \dots + \omega X^{r-1} \dots(3)$$

where both right and left members are matrix operators, and X is to be treated as constant, in the right hand summation, as far as the operator ω is concerned.

Proof. Let $f(x)$ be any ordinary regular scalar function of the n^2 variables x_{ij} . Then the partial differential coefficients satisfy the equation

$$\frac{\partial f}{\partial x_{ji}} = \frac{\partial f}{\partial y_{qp}} \frac{\partial y_{qp}}{\partial x_{ji}}$$

with n^2 terms implied by p, q on the right. But $\partial y_{qp}/\partial x_{ji}$ consists of r terms, due to differentiating the r successive factors in (1): and these lead to r instalments for $\partial f/\partial x_{ji}$, say

$$\frac{\partial f}{\partial x_{ji}} = T_1 + T_2 + \dots + T_r.$$

In T_1 , we have by (1)

$$T_1 = \frac{\partial f}{\partial y_{qp}} \frac{\partial x_{qa_1}}{\partial x_{ji}} x_{a_1 a_2} \dots x_{a_{r-1} p}$$

which vanishes unless $q = j, a_1 = i$. Hence

$$T_1 = x_{ia_2} x_{a_2 a_3} \dots x_{a_{r-1} p} \frac{\partial f}{\partial y_{jp}}$$

which is the ij th element in the matrix

$$[T_1] = X^{r-1} \omega f$$

as is at once apparent on writing this element out in full for this last matrix. Similarly

$$T_2 = \frac{\partial f}{\partial y_{pq}} x_{qa_1} \frac{\partial x_{a_1 a_2}}{\partial x_{ji}} x_{a_2 a_3} \dots x_{a_{r-1} p}$$

which requires $a_1 = j, a_2 = i$, so that T_2 is the ij th element in a matrix $X^{r-2} \omega X f$. This goes on until T_r is reached, giving

$$\begin{aligned} \Omega f(x) &= (X^{r-1} \omega f + X^{r-2} \omega X f + \dots + \omega X^{r-1} f) \\ &= (X^{r-1} \omega + X^{r-2} \omega X + \dots + \omega X^{r-1}) f. \end{aligned} \dots (4)$$

Since $f(x)$ is an arbitrary function this proves the theorem.

The function f on the right is of course to be expressed explicitly as a function of the n^2 new variables y . It will be seen that the formula reverts to the elementary result

$$\frac{df}{dx} = rx^{r-1} \frac{df}{dy}, \quad y = x^r$$

when $n = 1$.

If δ_{ij} and $(x^\nu)_{ij}$ denote the typical elements in the unit matrix and the matrix X^ν , the above reduction gives

$$\frac{\partial y_{qp}}{\partial x_{ji}} = \delta_{qj} (x^{r-1})_{ip} + \dots + (x^\nu)^{-1}_{qj} (x^{r-\nu})_{ip} + \dots + (x^{r-1})_{qj} \delta_{ip}.$$

This series becomes equal to r if X is the unit matrix and if $q = j, i = p$, because each factor is unity. The series is zero if either $q \neq j$ or $i \neq p$. Hence the Jacobian $|\partial y_{qp} / \partial x_{ji}|$ of the n^2 elements y with regard to x does not vanish identically, because in this case it is equal to r^{n^2} .

THEOREM VIII. *If ν is a positive integer, then*

$$\Omega_s X^{\frac{1}{\nu}} = \frac{1}{\nu} X^{\frac{1}{\nu}-1}.$$

This proves the fundamental formula § 2 (9) for a fractional case.

Proof. Let $X^{\frac{1}{\nu}} = Y$, so that $X = Y^\nu$. The expression $s(Y)$, which is $s(X^{\frac{1}{\nu}})$, is a scalar quantity, and $\Omega s(X^{\frac{1}{\nu}})$ is a matrix, Z say. Let $\omega = \left[\frac{\partial}{\partial y_{ji}} \right]$. Then, by (4) above,

$$\omega s(Y) = (Y^{\nu-1} \Omega + Y^{\nu-2} \Omega Y + \dots + \Omega Y^{\nu-1}) s(X^{\frac{1}{\nu}})$$

where, on the right Y is constant for the operation Ω , which acts merely on the scalar operand $s(X^{\frac{1}{\nu}})$. But a scalar commutes with Y . Hence

$$\begin{aligned} \omega s(Y) &= Y^{\nu-1} \Omega s(X^{\frac{1}{\nu}}) + Y^{\nu-2} \Omega s(X^{\frac{1}{\nu}}) Y + \dots + \Omega s(X^{\frac{1}{\nu}}) Y^{\nu-1} \\ &= Y^{\nu-1} Z + Y^{\nu-2} ZY + \dots + ZY^{\nu-1} \end{aligned}$$

by the associative law. But $\omega s(Y) = 1$, hence

$$Y^{\nu-1} Z + Y^{\nu-2} ZY + \dots + ZY^{\nu-1} = 1$$

and by Theorem IV, this may be solved for Z , giving $Z = \frac{1}{\nu} Y^{1-\nu}$.

But $Y = X^{\frac{1}{\nu}}$, so that finally,

$$\Omega s X^{\frac{1}{\nu}} = Z = \frac{1}{\nu}, \quad Y^{1-\nu} = \frac{1}{\nu} X^{\frac{1}{\nu}-1}$$

subject to the conditions stated in Theorem IV, and now governing the matrix $X^{\frac{1}{\nu}}$.

THEOREM IX. *If μ, ν are positive integers*

$$\Omega_s X^{\frac{\mu}{\nu}} = \frac{\mu}{\nu} X^{\frac{\mu}{\nu}-1}$$

Proof. Let $X^{\frac{\mu}{\nu}} = Y$, and $X^\mu = Y^\nu = U$. This gives three different matrices each of n^2 independent variables, and three operators

$$\Omega = \left[\frac{\partial}{\partial x_{ji}} \right], \quad \omega = \left[\frac{\partial}{\partial y_{ji}} \right], \quad \theta = \left[\frac{\partial}{\partial u_{ji}} \right]. \quad \dots\dots\dots(5)$$

We form the spur $s(Y)$, so that

$$s(X^{\frac{\mu}{\nu}}) = s(Y) = s(U^{\frac{1}{\nu}});$$

and we consider the matrix

$$Z = \Omega s(X^{\frac{\mu}{\nu}})$$

In this case, by Theorem VII,

$$X^{\mu-1} \theta + \dots + \theta X^{\mu-1} = \Omega, \quad Y^{\nu-1} \theta + \dots + \theta Y^{\nu-1} = \omega.$$

By Theorem VIII, $\theta s(U^{\frac{1}{\nu}}) = \frac{1}{\nu} U^{\frac{1}{\nu}-1} = \frac{1}{\nu} X^{\mu-\mu}$. Hence

$$\begin{aligned} Z &= X^{\mu-1}(\theta s(U^{\frac{1}{\nu}})) + \dots + (\theta s(U^{\frac{1}{\nu}})) X^{\mu-1} \\ &= \mu X^{\mu-1} \frac{1}{\nu} X^{\mu-\mu} \\ &= \frac{\mu}{\nu} X^{\mu-1} \end{aligned}$$

which proves the theorem.

THEOREM X. For all rational values of r , $\Omega_s X^r = r X^{r-1}$.

Proof. This follows from what has gone before by shewing that it holds for $r = -1$.

Let x^{ji} denote the ij^{th} element in the reciprocal matrix $1/X$. We take

$$XY = 1, \quad Y = [y_{ij}] = [x^{ji}] = X^{-1}, \quad \dots\dots(6)$$

and $y_{pa} x_{aq} = \delta_{pq} = 0$ or 1 , according as $p \neq q$ or $p = q$.

By differentiation of $y_{pa} x_{aq}$ we have

$$\begin{aligned} \frac{\partial y_{pa}}{\partial x_{ji}} x_{aq} + y_{pa} \frac{\partial x_{aq}}{\partial x_{ji}} &= 0, \text{ or} \\ \left. \begin{aligned} \frac{\partial y_{pa}}{\partial x_{ji}} x_{aq} &= 0, \\ q \neq i, \end{aligned} \right\} \text{ and } \left. \begin{aligned} \frac{\partial y_{pa}}{\partial x_{ji}} x_{aq} + y_{pj} &= 0, \\ q = i. \end{aligned} \right\}. \end{aligned}$$

Multiply these last by $x_{\gamma q}$ and sum for $q = 1, 2, \dots, n$. Then

$$\frac{\partial y_{pa}}{\partial x_{ji}} x_{aq} x_{\gamma q} + y_{pj} x_{\gamma} = 0,$$

since the second (single) term only appears for the value i of q . But $x_{aq} x_{\gamma q} = \delta_{a\gamma}$, which only is non-zero if $\gamma = a$, in which case it is unity. Hence

$$\frac{\partial y_{pa}}{\partial x_{ji}} + y_{pj} x^{ai} = 0.$$

But $\frac{\partial f}{\partial x_{ji}} = \frac{\partial f}{\partial y_{,a}} \frac{\partial y_{pa}}{\partial x_{ji}}$.

Hence $\frac{\partial f}{\partial x_{ji}} + \frac{\partial f}{\partial y_{pa}} y_{pj} x^{ai} = 0$.

Rearranging the second term we find it to be the ij^{th} element of a

product matrix. Thus

$$\Omega f = \left[\frac{\partial f}{\partial x_{ji}} \right] = - \left[x^{ai} \frac{\partial f}{\partial y_{pa}} y_{pj} \right] = - X^{-1} \omega Y f.$$

Finally this can be written in the form

$$X \Omega + \omega Y = 0 \quad \dots\dots\dots(7)$$

or, equally well, since $XY = 1$,

$$\omega = - X \Omega X, \quad \Omega = - X^{-1} \omega X^{-1}. \quad \dots\dots\dots(8)$$

In particular if the operand is $s(Y^r)$ or $s(X^{-r})$, we have

$$\Omega_s X^{-r} = - X^{-1} (\omega_s Y^r) X^{-1} = - X^{-1} r Y^{r-1} X^{-1} = - r X^{-r-1}$$

for all positive rational values of r , by Theorem IX. This demonstrates Theorem X.

COROLLARY. *The law of transformation from Ω to ω , when Y is any negative integral power of X , follows the rule of Theorem VI.*

THEOREM XI. *If $Y = f(X)$ is a scalar function of X , capable of expression as a series $\sum_r a_r X^r$, where r is an integer and each a_r is a scalar constant, then the operators ω_s and Ω_s are connected by the identity*

$$\Omega_s = \Omega_s (Y \omega_s)$$

where, on the right, Ω_s acts solely on Y .

Proof. This follows directly from the results of Theorems VII and X.

Also if we write $\frac{d}{dX}$ for Ω_s and $\frac{d}{dY}$ for ω_s , the result can be expressed as:

$$\frac{d}{dX} = \frac{d}{dX} \left(Y \frac{d}{dY} \right),$$

which may be regarded as the matrix generalization of the ordinary formula $\frac{d}{dx} = \frac{dy}{dx} \frac{d}{dy}$. But the operator $\frac{d}{dX}$ is not associative with

Y and $\frac{d}{dY}$: it is not true to write $\frac{d}{dX} = \frac{dY}{dX} \frac{d}{dY}$. Other results which follow without difficulty are as follows:—

$$\text{If } Y = AX, \text{ then } \Omega = \omega A. \quad \dots\dots\dots(9)$$

$$\text{If } Y = A + X, \text{ then } \Omega = \omega. \quad \dots\dots\dots(10)$$

$$\text{If } Y = f(X), \text{ then } \Omega X - X \Omega = \omega Y - Y \omega, \quad \dots\dots\dots(11)$$

provided that, for purposes of operation, these factors X and Y are maintained as constants.

This last follows from § 6 (4) by multiplying fore and aft in turn by X and subtracting. The formula is true when $Y = X^r$, and therefore also when $Y = \sum \alpha_r X^r$. It gives an interesting example of a *matrix differential invariant* for a general scalar functional transformation.

If $Y = AXBXC$, then $\Omega = BXC\omega A + C\omega AXB$ (12)

If $Y = AX^{\frac{1}{2}}$, then $\omega = A^{-1} Y \Omega A^{-1} + \Omega A^{-1} Y A^{-1}$ (13)

In this case $X = (A^{-1} Y)^2 = A^{-1} Y A^{-1} Y$ and the result (13) follows straightforwardly. If operation takes place on the spur of Y , it leads to the further result

$$1 = \omega_s(Y) = A^{-1} Y Z A^{-1} + Z A^{-1} Y A^{-1},$$

where $Z = \Omega_s AX^{\frac{1}{2}}$. Hence, by a slight reduction,

if $Z = \Omega_s AX^{\frac{1}{2}}$, then $\sqrt{X} \cdot Z + Z \sqrt{X} = A$ (14)

Similarly, if ν is a positive integer and $X = Y^\nu$, $Z = \Omega_s AX^{\frac{1}{\nu}}$, then

$$\left. \begin{aligned} & Y^{\nu-1} \Omega + \dots + Y^{\nu-i} \Omega Y^{i-1} + \dots + \Omega Y^{\nu-1} = \omega, \\ \text{and} & Y^{\nu-1} Z + \dots + Z Y^{\nu-1} = A, \\ \text{and} & Y^\nu Z - Z Y^\nu = YA - AY. \end{aligned} \right\} \dots \dots (15)$$

The corresponding results for $Z = \Omega_s AX^{\frac{\mu}{\nu}}$, when $X = Y^\nu$ are as follows:

or $\left. \begin{aligned} & Y^{\nu-1} Z + \dots + Z Y^{\nu-1} = Y^{\mu-1} A + \dots + A Y^{\mu-1}, \\ & Y^\nu Z - Z Y^\nu = Y^\mu A - A Y^\mu. \end{aligned} \right\} \dots \dots (16)$

§ 7. THE BINOMIAL THEOREM.

The operator Ω_s plays the part of ordinary differentiation in developing a scalar function $f(X + A)$ as a power series in A , so giving a matrix form of Taylor's and Maclaurin's Theorems. We first consider the case when $f(X) = X^r$, and obtain the *binomial theorem*:

$$(X + A)^r = X^r + \Omega_s AX^r + \frac{1}{2!} (\Omega_s A)^2 X^r + \dots + \frac{1}{\nu!} (\Omega_s A)^\nu X^r + \dots (1)$$

If r is a positive integer this series ends, with $\nu = r$, in $\nu + 1$ terms. For example, if $r = 3$, by actual multiplication

$$(X + A)^3 = X^3 + (X^2A + XAX + AX^2) + (XA^2 + AXA + A^2X) + A^3,$$

with its four groups of terms on the right, arranged in ascending

degree of A . By § 3 (4) the group linear in A is $\Omega_s AX^3$, and the next group is $\frac{1}{2} \Omega_s A(X^2A + XAX + AX^2)$, as may be verified. Finally $\frac{1}{6} (\Omega_s A)^3 X^3 = A^3$ and the theorem is true if $r = 3$.

By actual multiplication

$$(X + A)^r = X^r + \sum X^{\alpha_1} AX^{\alpha_2} + \sum X^{\beta_1} AX^{\beta_2} AX^{\beta_3} + \dots + A^r \} \dots (2)$$

where $\alpha_1 + \alpha_2 + 1 = \beta_1 + \beta_2 + \beta_3 + 2 = \dots = r$,

and the summation extends to all different positive and zero values of the integers $\alpha_1, \alpha_2, \beta_1, \dots$. By § 3 (4) the second term in the expansion (2) agrees with $\Omega_s AX^r$ in (1). Again by § 3 (4),

$$\Omega_s AX^{\alpha_1} AX^{\alpha_2} = \sum_{i=1}^{\alpha_1} X^{\alpha_1-i} AX^{\alpha_2} AX^{i-1} + \sum_{i=1}^{\alpha_2} X^{\alpha_2-i} AX^{\alpha_1} AX^{i-1}.$$

But $\alpha_1 + \alpha_2 = r - 1$: hence

$$\begin{aligned} \sum_{\alpha_1=0}^{r-1} \Omega_s AX^{\alpha_1} AX^{r-\alpha_1-1} &= 2 \sum_{\alpha_1=0}^{r-1} \sum_{i=0}^{\alpha_1} X^{\alpha_1-i} AX^{r-2-\alpha_1} AX^i \\ &= 2 \sum X^{\beta_1} AX^{\beta_2} AX^{\beta_3}. \end{aligned}$$

Hence the third or β term on the right of (2) is $\frac{1}{2} (\Omega_s A) \Omega_s AX^r$, which agrees with the third term in (1). Similarly by induction we justify each step in the series (1). For the operation $\Omega_s A$, performed on $X^{p_1} AX^{p_2} A \dots AX^{p_\nu}$, by the cyclic rule of differentiation, gives ν series—one for each p_μ —such as

$$\sum_{i=1}^{p_\mu} X^{p_\mu-i} AX^{p_\mu+1} A \dots AX^{i-1}, \quad \mu = 1, 2, \dots, \nu;$$

and each of these ν series gives the requisite terms of order $\nu + 1$ in A , when the p 's run through their values. Hence the operation $\Omega_s A$, performed on terms of order $\nu - 1$ in A , produce the terms of order ν , multiplied by ν . This proves (1) when r is a positive integer.

Negative Exponent. For the negative exponent a similar proof holds, provided that the infinite series, which necessarily arise, are uniformly convergent. If the moduli of the latent roots of the matrix AX^{-1} are all less than unity, we have

$$\frac{1}{X + A} = X^{-1} - X^{-1}AX^{-1} + X^{-1}AX^{-1}AX^{-1} - X^{-1}AX^{-1}AX^{-1}AX^{-1} + \dots (3)$$

as is easy to verify. By operating on both sides of this identity with

$\left[\frac{\partial}{\partial a_{ji}} \right]_s$, treating X as a constant, we deduce

$$\frac{1}{(X + A)^2} = X^{-2} - X^{-2}AX^{-1} - X^{-1}AX^{-2} + X^{-2}AX^{-1}AX^{-1} + \dots \quad (4)$$

This operation is performed on the right hand member of the identity by means of Theorem VI and the cyclic law. On the left it requires an intermediate step, involving a new variable $B = X + A$. Thus (§ 6 (10))

$$\left[\frac{\partial}{\partial b_{ji}} \right] = \left[\frac{\partial}{\partial a_{ji}} \right]; \text{ also } \left[\frac{\partial}{\partial b_{ji}} \right]_s B^{-1} = -B^{-2};$$

whence $\left[\frac{\partial}{\partial a_{ji}} \right]_s (X + A)^{-1} = -(X + A)^{-2}$. The process may be repeated, and the terms of the same degree in A may be grouped in ascending order; and the result is the theorem

$$(X + A)^{-r} = X^{-r} + \Omega_s AX^{-r} + \frac{1}{2!} (\Omega_s A)^2 X^{-r} + \dots \dots \dots (5)$$

where r is a positive integer.

Fractional Exponent. For values less than unity, of the moduli of the latent roots of X_1 , we have

$$(1 + X)^r = a_0 + a_1 X + a_2 X^2 + \dots + a_\nu X^\nu + \dots \dots \dots (6)$$

where the coefficients a_ν are the ordinary binomial coefficients—whether r is integral or fractional. Let

$$e^{\Omega_s A} = 1 + \Omega_s A + \frac{1}{2!} (\Omega_s A)^2 + \frac{1}{3!} (\Omega_s A)^3 + \dots \dots \dots (7)$$

denote the full operation implied in (1), so that $(X + A)^\nu = e^{\Omega_s A} X^\nu$ at any rate when ν is an integer. Changing X to $X + A$ in (6) we have

$$\begin{aligned} (1 + X + A)^r &= a_0 + a_1 (X + A) + \dots + a_\nu e^{\Omega_s A} X^\nu + \dots \\ &= e^{\Omega_s A} (a_0 + a_1 X + \dots + a_\nu X^\nu + \dots) = e^{\Omega_s A} (1 + X)^r \end{aligned}$$

provided that this alteration in the order of terms in the double series is justified. If $Y = 1 + X$, and $\omega = \left[\frac{\partial}{\partial y_{ji}} \right]$, then by § 6 (10),

$\omega_s = \Omega_s$. Hence by substitution,

$$(Y + A)^r = e^{\omega_s A} Y^r = Y^r + \omega_s A Y^r + \frac{1}{2!} (\omega_s A)^2 Y^r + \dots,$$

where r may be fractional. In this way the binomial theorem may be extended to cover all rational exponents.

§ 8. TAYLOR'S THEOREM.

If $f(X)$ is a scalar function of X , which can be expanded in a finite series of terms $\Sigma a_r X^r$, where r is rational, or is an infinite series which converges and admits of rearrangement of its terms, then

$$f(X + A) = e^{\Omega_s A} f(X) = f(X) + \Omega_s A f(X) + \frac{1}{2!} (\Omega_s A)^2 f(X) + \dots \dots (1)$$

provided also that A and X are restricted to a suitable domain of values.

Proof. The presence of a scalar coefficient a_r in the term $a_r X^r$ leaves the binomial theorem unaffected; a_r acts as a constant multiplier throughout. If the binomial expansion of $a_r (X + A)^r$ is applied to each term of $f(X)$, and the terms are regrouped in ascending degree for A , the result is formula (1).

This can only be regarded as a very crude proof, in the absence of a more precise general theorem analogous to the mean value theorem of the differential calculus.

COROLLARY I.

$$f(A) = f(0) + \Omega_s A f(X)_{X=0} + \frac{1}{2!} (\Omega_s A)^2 f(X)_{X=0} + \dots \dots (2)$$

This gives a method of expanding a function as a power series, corresponding to Maclaurin's expansion.

COROLLARY II. If A is small, the approximate value of $f(X + A)$ is given by the relation

$$f(X + A) = f(X) + \Omega_s A f(X) + O(A^2). \dots \dots (3)$$

Here the notation $O(A^2)$ means that every element in the matrix $f(X + A) - f(X) - \Omega_s A f(X)$ consists of second or higher orders of small quantities a_{ij} . The formula is a direct consequence of Taylor's Theorem; and it leads to a definition of the differential $df(X)$ of a scalar function of a matrix.

Definition of dX and $df(X)$. By dX is meant the matrix of n^2 independent differentials $[dx_{ij}]$. By $df(X)$ is meant the matrix formed of the terms of degree 1 in dx_{ij} in

$$f(X + dX) - f(X).$$

It follows that $\Omega_s dX f(X)$ is the differential of $f(X)$. $\dots \dots (4)$

Example. $f(X) = X^{\frac{1}{2}}$. Here, when A is small,

$$(X + A)^{\frac{1}{2}} = X^{\frac{1}{2}} + Z + O(A^2), \dots\dots\dots(5)$$

where $Z = \Omega_s AX^{\frac{1}{2}}$, and it may be verified directly that the value of Z satisfying (5) is given, as it should be, by § 6 (14):

$$A = \sqrt{X} \cdot Z + Z\sqrt{X}. \dots\dots\dots(6)$$

In fact if $(X + A)^{\frac{1}{2}} = P$, and $X^{\frac{1}{2}} = Y$, then $P = Y + E$ where E is small. Hence $A = P^2 - Y^2 = EY + YE + E^2$. If E is of the same order as A , this last agrees with (6) by taking $Z = E$, which in turn agrees with (5). In general there is only one Z satisfying the equation

$$EY + YE + E^2 = YZ + ZY,$$

and this value is E , if the term E^2 is neglected.

By taking $f(X) = X^{\frac{1}{\nu}}$ or $X^{\frac{\mu}{\nu}}$, the results of § 6 (15), (16) may be utilized to illustrate (3) above.

§ 9. THE OPERATORS $(\Omega X - X\Omega)$ AND $\Omega(\Omega X - X\Omega)$.

Certain other results may conveniently be added to what has gone before. As was shewn in the original enquiry,¹ the operator $\Psi \equiv \Omega(\Omega X - X\Omega)$ yields the result, analogous to that of Ω_s ,

$$\Psi f(X) = f'(X). \dots\dots\dots(1)$$

But, unlike Ω_s , the Ψ operator produces an essentially different result when it acts on $f(X')$. In fact, if $f(X') \equiv X'^r$, then

$$\begin{aligned} \Psi X'^r &= (n + 1) \{X'^{r-1} + XX'^{r-2} + \dots + X^{r-1}\} \\ &\quad - nX^{r-1} - s(X) X^{r-2} - s(X^2) X^{r-3} - \dots - s(X^{r-1}). \dots\dots(2) \end{aligned}$$

In this operation $\Psi \equiv \Omega(\Omega X - X\Omega)$, the two X 's which enter are treated as variables and are affected by the Ω factors standing on their left. There are, in fact, two subsidiary operators Φ and Φ_c , defined as

$$\Phi = \Omega X - X\Omega, \quad \Phi_c = \Omega X_c - X_c\Omega. \dots\dots\dots(3)$$

In Φ , X is variable, in Φ_c the X , written X_c , is treated as a constant for the Ω operation, as was the case in § 6. So Φ_c is the invariant

¹ *Proc. Edinburgh Math. Soc., loc. cit.* p. 128.

operator, § 6 (11), for transformations of the type $Y = f(X)$. Typical results are as follows:

$$\left. \begin{aligned} \Phi X' &= (\Omega X - X \Omega) X' = (n + 1) X' - X, \\ \Phi X'^r &= (\Omega X - X \Omega) X'^r = (n + 1) X'^r - X^r, \end{aligned} \right\} \dots\dots\dots(4)$$

$$\Phi f(X') = (n + 1)f(X') - f(X), \dots\dots\dots(5)$$

$$\Phi_s(X^r) = ns(X^r), \quad \Phi_s(f(X)) = ns(f(X)). \dots\dots\dots(6)$$

$$\left. \begin{aligned} \Phi_c X'^r &= X'^r - X^r, \\ \Phi_c f(X') &= f(X') - f(X), \end{aligned} \right\} \dots\dots\dots(7)$$

$$\Phi_c s(f(X)) = 0, \dots\dots\dots(8)$$

$$\Phi AX^r = (nA - s(A)) X^r + s(AX^r), \quad \Phi_c AX^r = s(AX^r) - s(A) X^r \dots(9)$$

$$\Psi AX^r = AX^{r-1} + XAX^{r-2} + \dots + X^{r-1}A + \left| \begin{matrix} s(1) & s(A) \\ s(X) & s(AX) \end{matrix} \right| X^{r-2} + \left| \begin{matrix} s(1) & s(A) \\ s(X^2) & s(AX^2) \end{matrix} \right| X^{r-3} + \dots + \left| \begin{matrix} s(1) & s(A) \\ s(X^{r-1}) & s(AX^{r-1}) \end{matrix} \right| \dots\dots\dots(10)$$

In this last $s(1)$ is n . If $A = I$, the unit matrix, all these two-rowed determinants vanish, and the formula reverts to $\Psi X^r = rX^{r-1}$.

§ 10. CONNEXION WITH QUANTUM ALGEBRA.

In the *Proceedings of the Cambridge Philosophical Society*, 23 (1926) 412-417, Dr P. A. M. Dirac developed a calculus of q -numbers, involving the ideas of functionality and differentiation. Put into the preceding notation, Y is defined as a function $f(X)$ of a single variable X , if every value of B satisfying the condition $XB = BX$ also satisfies $YB = BY$. If X_0, Y_0 are particular values of X, Y such that $Y_0 = f(X_0)$, then every value of C satisfying $X_0 = CXC^{-1}$ must also satisfy $Y_0 = CYC^{-1}$. These, which I call shortly the conditions B and C , were shewn to characterize a particular function f quite adequately. A further condition, here called D , was added to define differentiation: $Z = f'(X)$ is the derivative of Y with regard to X , if every value of P satisfying the condition

$$XP - PX = 1 \dots\dots\dots(1)$$

also satisfies the condition

$$YP - PY = Z = f'(X). \dots\dots\dots(2)$$

This was shewn to yield the ordinary results of differentiation adequately. For instance if $Y = X^2$ then Z , as defined by $YP - PY$, is $2X$. Unfortunately this remarkable algebraic theory of q -numbers

breaks down in the case of matrices. For if X and P are matrices of order n , then $s(XP - PX) = s(XP) - s(PX) = 0$ by the cyclic law; whereas $s(1) = n$. Condition (1) above is accordingly impossible for *finite* matrices, although it holds for certain matrices of infinite order.

Nevertheless a close analogy can be revealed between this q -number calculus of Dr Dirac and the preceding calculus of finite matrices. It is easy to verify that the scalar function $Y = f(X)$ of the latter calculus satisfies conditions B and C of Dirac. As for the further condition D , suppose that A is a matrix of order n , such that a value of P exists which satisfies the condition

$$XP - PX = A, \quad \dots\dots\dots(3)$$

then it follows that

$$YP - PY = \Omega_3 AY, \quad \dots\dots\dots(4)$$

provided that A is treated as constant under the Ω operation. This is manifestly analogous to the relations (1) and (2) above. Granted (3), the proof of (4) follows by an interesting use of the formulae in §3 (4) and §5 (4). It is also true for fractional indices such as occur in $Y = X^{\frac{1}{2}}$. If for instance

$$X^{\frac{1}{2}}P - PX^{\frac{1}{2}} = B,$$

then $XP - PX = BX^{\frac{1}{2}} + X^{\frac{1}{2}}B$; and this expression is equal to A , when $B = \Omega_3 AX^{\frac{1}{2}}$ (§6 (14)). The details of a formal proof are not difficult to supply.

It must however be added that, for a given X , condition (3) imposes restrictions on the value of A —which has hitherto been taken to be an arbitrary constant matrix. If, for example, X is a quasi scalar matrix, with n unequal diagonal elements λ_i all differing from zero, and with zeroes everywhere else, then A must have zeroes throughout its diagonal but is otherwise arbitrary. It would in fact be difficult to find a matrix A which more completely contradicted the requirements of the right hand member in condition (1)!