On Differentiating a Matrix.

By Professor H. W. TURNBULL.

(*Received* 15*th September* 1927. *Read* 4*th June* 1927.)

INTRODUCTION.

The theorem $dx^r/dx = rx^{r-1}$ is well known. So also is the theorem that if $\Delta = | a_{ij} |$, $\Delta^{-1} = | A_{ji}/\Delta |$ concerning a determinant Δ and its reciprocal expressed by means of cofactors A_{ij} of a_{ij}. Not quite so well known is the Cayley Hamilton theorem that a matrix $X = [x_{ij}]$ satisfies its own characteristic equation

$$\phi(\lambda) \equiv | \lambda - x_{ij} | = 0.$$

Unlike as these three results are, they nevertheless can be looked upon as particular phases of a general theorem concerning a matrix differential operator $\Omega \equiv \left[\frac{\partial}{\partial x_{ji}}\right]$ acting upon a function of a matrix X or its transposed.

The chief properties are summed up in various theorems I-VII. Speaking generally, any function $f(X)$ of a single matrix is expressible as $\Omega\,\phi$ where ϕ is a determinant *scalar* function of the latent roots of X. The simple result, Theorem III,

$$\Omega\, s^r = rX^{r-1},$$

where s_r is the sum of the r^{th} powers of the latent roots of the matrix X is here established, but it requires a rather intricate Lemma (§ 7) concerning the principal minors of a determinant.

Although the finite matrix has been treated, the work is adaptable to infinite matrices.

The enquiry suggested itself as a natural continuation of Cayley's discovery that the determinant operator $| \Omega |$ has the property

$$| \Omega | \, | X |^r = \mu \, | X |^{r-1}$$

where μ is the numerical constant $r(r+1)(r+2)\ldots(r+n-1)$.

https://doi.org/10.1017/S0013091500007434 Published online by Cambridge University Press

§ 1. Let

$$X = [x_{ij}] = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nn} \end{bmatrix} \quad \text{............} \quad \text{(1)}$$

be an n rowed matrix whose n^2 elements are treated as independent variables. Further let capital letters A, B, C denote constant matrices, and Y, Z dependent variable matrices, wherein the elements y_{ij}, z_{ij} are functions of the n^2 variables x_{ij}.

Then I propose to develop the theory of matrix differentiation on the following basis. From X, let a matrix differential operator

$$\Omega = \left[\frac{\partial}{\partial x_{ji}}\right] = \begin{bmatrix} \frac{\partial}{\partial x_{11}} & \cdots & \frac{\partial}{\partial x_{n1}} \\ \frac{\partial}{\partial x_{12}} & \cdots & \frac{\partial}{\partial x_{n2}} \\ \cdots & \cdots & \cdots \\ \frac{\partial}{\partial x_{1n}} & \cdots & \frac{\partial}{\partial x_{nn}} \end{bmatrix} \quad \text{..................} \quad \text{(2)}$$

be formed by placing the n^2 differential operators $\frac{\partial}{\partial x_{ji}}$ in matrix array with the order of suffixes *transposed* from that of X.

In particular if $n = 1$ and there is only one variable x, this operator Ω becomes the ordinary $\frac{d}{dx}$. So we may regard Ω itself as a generalization of differentiation of a scalar number: and indeed it will be seen in what follows that many of the features of the differential calculus appear in this matrix calculus often in a very unexpected setting.

First we must give the law of transposition full play by defining the transposed matrices and operator, as indicated by an accent,

$$X' = [x_{ji}], \; A' = [a_{ji}], \; \Omega' = \left[\frac{\partial}{\partial x_{ij}}\right]. \quad \text{............} \quad \text{(3)}$$

In these the rows and columns of the corresponding unaccented matrices have been interchanged, as indicated by the reversal of suffix order. Also, for such accented symbols we have the fundamental laws as shown by

$$(A+B)'=A'+B',\ (AB)'=B'A', \qquad \ldots\ldots\ldots\ldots(4)$$

the latter illustrating what may be called the *reversal law* which also holds for the reciprocal operation, namely

$$(AB)^{-1}=B^{-1}A^{-1}. \qquad \ldots\ldots\ldots\ldots\ldots\ldots(5)$$

Next we define the effect of the operators Ω, Ω' by the ordinary multiplication law of matrices. So

$$\Omega\, Y=\left[\frac{\partial}{\partial x_{ji}}\right][y_{ij}]=[z_{ij}] \qquad \ldots\ldots\ldots\ldots\ldots\ldots(6)$$

where the typical element is given by

$$z_{ij}=\sum_{k=1}^{n}\frac{\partial y_{kj}}{\partial x_{ki}}=\frac{\partial y_{1j}}{\partial x_{1i}}+\frac{\partial y_{2j}}{\partial x_{2i}}+\ldots+\frac{\partial y_{nj}}{\partial x_{ni}}. \qquad \ldots\ldots\ldots(7)$$

and similarly

$$\Omega'\, Y=\left[\frac{\partial}{\partial x_{ij}}\right][y_{ij}]=\left[\sum_{k=1}^{n}\frac{\partial y_{kj}}{\partial x_{ik}}\right] \qquad \ldots\ldots\ldots\ldots\ldots(8)$$

An important special case occurs when Ω operates on a scalar expression $f(x_{ij})$, involving any of the n^2 variables x_{ij}. Using the ordinary multiplication law once again we obtain

$$\Omega f=\begin{bmatrix}\frac{\partial}{\partial x_{11}} & \cdots & \frac{\partial}{\partial x_{n1}}\\ \cdots & \cdots & \cdots\\ \frac{\partial}{\partial x_{1n}} & \cdots & \frac{\partial}{\partial x_{nn}}\end{bmatrix} f=\begin{bmatrix}\frac{\partial f}{\partial x_{11}}, & \frac{\partial f}{\partial x_{21}}, & \cdots & \frac{\partial f}{\partial x_{n1}}\\ \cdots & \cdots & \cdots & \cdots\\ \frac{\partial f}{\partial x_{1n}}, & \frac{\partial f}{\partial x_{2n}} & \cdots & \frac{\partial f}{\partial x_{nn}}\end{bmatrix} \qquad \ldots\ldots\ldots(9)$$

which is the matrix of the n^2 first partial differential coefficients of the function f.

There is no difficulty in proving immediately that

$$\Omega(Y+Z)=\Omega\, Y+\Omega\, Z,\ \Omega'(Y+Z)=\Omega'\, Y+\Omega' Z, \qquad \ldots(10)$$

so that at present the operator behaves as ordinary differentiation. Rather a different state of things holds for operation on a product YZ, which does not reproduce the ordinary formula

$$\frac{d}{dx}yz=\frac{dy}{dx}z+y\frac{dz}{dx}. \qquad \ldots\ldots\ldots\ldots\ldots\ldots(11)$$

But let a suffix c be provisionally attached to a matrix to indicate

https://doi.org/10.1017/S0013091500007434 Published online by Cambridge University Press

that for the purpose of this operation, its elements are to be regarded as constants. Then the product formula for Ω differentiation is

$$\Omega(YZ) = \Omega(YZ_c) + \Omega(Y_c Z). \qquad \text{............(12)}$$

In this first term Z_c is constant and Y undergoes operation in the second, Y is constant. And although we can write

$$\Omega(YZ_c) = (\Omega Y) Z_c = \Omega Y Z_c, \qquad \text{..............(13)}$$

we cannot assume $\Omega Y_c Z = Y_c \Omega Z$ because the algebra is non-commutative.

Formula (12) is proved by straightforward application of to each element of the matrix product ΩYZ. There is no need to detail the steps. The next (13) is true because of the associative law of multiplication for matrices.

§ 2. We now have the basis of a matrix calculus, so that it is possible to build up some elementary results. Thus if A is constant we have, by § 1 (6) and (8),

$$\Omega A = \Omega' A = 0. \qquad \text{....................(1)}$$

Next, by using the same formulae, we infer

$$\begin{array}{ll} \Omega X = n, & \Omega' X = 1, \\ \Omega X' = 1, & \Omega' X' = n, \end{array} \qquad \text{..........(2)}$$

where the scalar numbers on the right, each stand for a scalar matrix. For example if $n = 3$,

$$\Omega X = \begin{bmatrix} \frac{\partial}{\partial x_{11}} & \frac{\partial}{\partial x_{21}} & \frac{\partial}{\partial x_{31}} \\ \frac{\partial}{\partial x_{12}} & \frac{\partial}{\partial x_{22}} & \frac{\partial}{\partial x_{32}} \\ \frac{\partial}{\partial x_{13}} & \frac{\partial}{\partial x_{23}} & \frac{\partial}{\partial x_{33}} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix} = 3. \quad \text{..(3)}$$

Let a_1 denote the sum of the leading diagonal terms of $A = [a_{ij}]$ a constant matrix. Then in the same way we find

$$\begin{array}{ll} \Omega(AX) = \Sigma a_{ii} = a_1, & \Omega'(AX) = A', \\ \Omega(XA) = nA, & \Omega'(XA) = A, \\ \Omega(AX') = A', & \Omega'(AX') = a_1, \\ \Omega(X'A) = A, & \Omega'(X'A) = nA. \end{array} \qquad \text{......(4)}$$

The second column of these eight results is of course deducible from the first, but is given for completeness. We note the interesting fact that the relation

$$\Omega\,(AX') = A'$$

shews that *the fundamental process of transposing a matrix A can be effected by a differential operator.*

Operation on integral powers of X and X'.

§ 3. In the present calculus we next seek the analogue of $dx^r/dx = rx^{r-1}$. This leads first to the following result:

Theorem I

$$\Omega\,X'^r = X'^{r-1} + XX'^{r-2} + \ldots + X^i\,X'^{r-i-1} + \ldots + X^{r-1}. \qquad \ldots .(1)$$

Proof. We have by § 1 (11)

$$\Omega\,X'^2 = \Omega\,(X'X') = \Omega\,X'X'_c + \Omega\,X'_c X'.$$

But $\Omega\,X'A = A$, and $\Omega\,AX' = A'$. Substituting $A = X'_c$, we have $\Omega\,X'^2 = X' + X$.

The formula is now true by induction; for if $Y = X'^r$, then

$$\Omega\,YX' = (\Omega\,Y)\,X' + \Omega\,Y_c X' = (\Omega\,X'^r)\,X' + Y'$$

by § 2 (4). But $Y' = X^r$, since $Y = X'^r$. Hence if (1) is assumed to be true we immediately have

$$\Omega\,X'^{r+1} = \Omega\,YX' = X'^r + XX'^{r-1} + \ldots + X^r,$$

which reproduces the same law as (1). This proves the theorem.

Corollary

$$\Omega'\,X^r = X^{r-1} + X'X^{r-2} + \ldots + X'^{r-1}, \qquad \ldots\ldots\ldots.(2)$$

Further if $n = 1$, we have $X = X' = x$, $Y = x^r$ and both formulae (1) and (2) revert to the familiar $dx^r/dx = rx^{r-1}$.

§ 4. Curiously enough the corresponding formula for $\Omega\,X^r$, when the operand is not X' but X, leads to an entirely different type of result involving the sums of powers of the latent roots of the characteristic equation, which we must therefore first consider.

For the square matrix of order n

$$X = [x_{ij}] = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{bmatrix} \quad \dots\dots\dots\dots (1)$$

the characteristic equation is given determinantally by

$$|\lambda I - X| \equiv \begin{vmatrix} \lambda - x_{11} & -x_{12} & \dots & -x_{1n} \\ -x_{21} & \lambda - x_{22} & \dots & -x_{2n} \\ \dots & \dots & \dots & \dots \\ -x_{n1} & -x_{n2} & \dots & \lambda - x_{nn} \end{vmatrix} = 0. \quad \dots\dots\dots\dots (2)$$

We write the expanded form of this determinant as

$$\phi(\lambda) = \lambda^n + p_1 \lambda^{n-1} + p_2 \lambda^{n-2} + \dots + p_n \quad \dots\dots\dots (3)$$

so that the coefficients p are polynomials in the n^2 arguments x_{ij}. Then the well-known Cayley-Hamilton theorem, that the matrix X itself satisfies the characteristic equation can be expressed as

$$\phi(X) = X^n + p_1 X^{n-1} + p_2 X^{n-2} + \dots + p_n = 0 \quad \dots\dots\dots (4)$$

identically.

With the usual notation s_r for the sum of the r^{th} powers of the n roots $\lambda_1, \lambda_2, \dots \lambda_n$ of $\phi(\lambda) = 0$, and h_r for the sum of the homogeneous products, r at a time, of these roots, we have the following relations

$$\left.\begin{aligned} & s_1 + p_1 = 0, \\ & s_2 + s_1 p_1 + 2p_2 = 0, \\ & \dots\dots\dots\dots\dots\dots \\ & s_r + s_{r-1} p_1 + s_{r-2} p_2 + \dots + s_1 p_{r-1} + r p_r = 0, \\ & \text{if } 0 < r < n+1, \text{ and} \\ & s_r + s_{r-1} p_1 + s_{r-2} p_2 + \dots + s_1 p_{r-1} = 0, \\ & \text{if } r > n. \end{aligned}\right\} \quad \dots\dots (5)$$

Also

$$\left.\begin{aligned} & h_2 + p_1 = 0, \\ & h_2 + h_2 p_1 + p_2 = 0, \\ & \dots\dots\dots\dots\dots\dots \\ & h_r + h_{r-1} p_1 + h_{r-2} p_2 + \dots + h_1 p_{r-1} + p_r = 0, \\ & \text{if } 0 < r < n+1, \text{ and} \\ & h_r + h_{r-1} p_1 + \dots + h_{r-n} p_n = 0 \end{aligned}\right\} \quad \dots\dots\dots (6)$$

if $r > n$. The relations between s and p are due to Newton, and those between h and p to Wronski, being proved by equating coefficients of powers of t on both sides of the identity

$$1 = (1 + p_1 t + p_2 t^2 + \ldots + p_n t^n)(1 + h_1 t + h_2 t^2 + \ldots + h_r t^r + \ldots)$$

since this last is

$$(1 - \lambda_1 t)(1 - \lambda_2 t) \ldots (1 - \lambda_n t)\{(1 - \lambda_1 t) \ldots (1 - \lambda_n t)\}^{-1}.$$

Sylvester gave a theorem whereby the latent roots of related matrices could be derived, namely if $f(X)$ is a rational function of the matrix X, then the latent roots of the matrix $f(x)$ are $f(\lambda_i)$, $i = 1, 2, \ldots . n$.

Such a function is

$$f(X) = \frac{a_0 X^p + a_1 X^{p-1} + \ldots . + a_p}{b_0 X^q + b_1 X^{q-1} + \ldots . + b_q} = \frac{N(X)}{D(X)}, \quad \ldots\ldots\ldots(7)$$

where the coefficients a_i, b_i are scalar, so that the order of division, forwards or afterwards, is immaterial since

$$N(X) . \{D(X)\}^{-1} = \{D(X)\}^{-1} N(X). \quad \ldots\ldots\ldots\ldots(8)$$

In particular the latent roots of the function X^r are

$$\lambda_1^r, \lambda_2^r, \ldots \lambda_n^r,$$

from which it follows by forming the result corresponding to the first of set (5), that if X^r is written as a matrix, the sum of its leading diagonal elements is

$$\Sigma \lambda_i^r. \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(9)$$

§ 5. We can now establish the following theorem:

THEOREM II. *If r is an integer, positive or negative*

$$\Omega X^r = \sum_{i=1}^{n} \frac{X^r - \lambda_i^r}{X - \lambda_i} \quad \ldots\ldots\ldots\ldots\ldots\ldots(1)$$

where the fraction is interpreted as an abbreviation for the series

$$X^{r-1} + \lambda_i X^{r-2} + \ldots . + \lambda_i^{r-1}.$$

Such a formula can also be written

$$\Omega X^r = nX^{r-1} + s_1 X^{r-2} + \ldots + s_{j-1} X^{r-j} + \ldots + s_{r-1} \quad \ldots\ldots(2)$$

since $s_{j-1} = \Sigma \lambda^{j-1}$.

Proof. We have already proved this if $r = 1$, since $\Omega X = n$. If $r = 2$, we have

$$\Omega X^2 = \Omega (XX) = \Omega (X X_c) + \Omega (X_c X)$$
$$= nX + s_1$$

by § 2 (4). So we prove the case for $r + 1$ by induction from that of r. In fact,

$$\Omega X^{r+1} = \Omega (X^r X) = (\Omega X^r) X + \Omega (X_c{}^r X).$$

And since $\Omega (AX) = \Sigma a_{ii}$, therefore $\Omega (X_c{}^r X) = \Sigma \lambda_i{}^r = s_r$ by Sylvester's Theorem. Accordingly

$$\Omega X^{r+1} = \sum_i \frac{X^r - \lambda_i{}^r}{X - \lambda_i} . X + \Sigma \lambda_i{}^r$$
$$= \sum_i \frac{X^{r+1} - \lambda_i{}^{r+1}}{X - \lambda_i}$$

treating, as we may, the right hand side by the rules of ordinary algebra. This proves the theorem if r is a positive integer.

The negative case proceeds similarly, starting with

$$0 = \Omega X^0 = \Omega (X^{-1} X) = (\Omega X^{-1}) X + \Omega (X_c{}^{-1} X);$$

whence

$$(\Omega X^{-1}) X = - \Sigma \lambda^{-1},$$
$$\Omega X^{-1} = - \Sigma \lambda^{-1} X^{-1},$$
$$= \Sigma \frac{X^{-1} - \lambda^{-1}}{X - \lambda},$$

giving the case when $r = -1$. Then if s is any positive integer we can prove the formula for X^{-s} assuming it for X^{-s+1}. Thus

$$\Omega X^{-s+1} = \Omega (X^{-s} X) = (\Omega X^{-s}) X + \Omega X_c{}^{-s} X$$
$$= (\Omega X^{-s}) X + \Sigma \lambda_i{}^{-s}.$$

Substituting for ΩX^{-s+1} this leads to the desired result.

Corollary I. *If $f(X)$ is a scalar polynomial function of the matrix X,*

$$\Omega f(X) = \sum_{i=1}^{n} \frac{f(X) - f(\lambda_i)}{X - \lambda_i} \qquad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

Corollary II. *If $f(X)$ is a rational function of X the same is true.*

For we prove the results term by term after developing the function $f(X)$ in ascending or descending powers of X, starting with the formula (1). How (3) also holds for an analytic function $f(X)$ will be briefly considered in § 12.

OPERATION ON A SCALAR FUNCTION.

§ 6. We return to the formula (9) of the first paragraph,

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

and seek its principal applications. A simple instance is given when we take

$$f = \lambda_1 + \lambda_2 + \ldots + \lambda_n = -p_1 = \Sigma\, x_{ii} \quad \ldots\ldots\ldots\ldots (2)$$

which gives the unit matrix as result:

$$\Omega f = \begin{bmatrix} 1 & 0 & 0 & \ldots \\ 0 & 1 & 0 & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots 1 \end{bmatrix} = 1. \quad \ldots\ldots\ldots\ldots\ldots (3)$$

This leads to the question, what happens if the other elementary functions, p_i, of the latent roots are operated upon? The answer is given by the following equations

THEOREM III.
$$\left.\begin{aligned} &\Omega\, s_1 = 1, \\ &\Omega\, s_2 = 2X, \\ &\Omega\, s_3 = 3X^2, \\ &\ldots\ldots\ldots\ldots \\ &\Omega\, s_r = rX^{r-1}, \end{aligned}\right\}$$

THEOREM IV.
$$\left.\begin{aligned} &1 + \Omega\, p_1 = 0, \\ &X + p_1 + \Omega\, p_2 = 0, \\ &X^2 + p_1 X + p_2 + \Omega\, p_3 = 0, \\ &\ldots\ldots\ldots\ldots\ldots \\ &X^{n-1} + p_1 X^{n-2} + \ldots + p_{n-1} + \Omega\, p_n = 0. \end{aligned}\right\}$$

THEOREM V.
$$\left.\begin{aligned} &1 - \Omega\, h_1 = 0, \\ &X + h_1 - \Omega\, h_2 = 0, \\ &X^2 + h_1 X + h_2 - \Omega\, h_3 = 0, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned}\right\}$$

Here we have sets of results which appear to be fundamental in the general theory of functions of a single variable matrix X. The first set, which is of surprising simplicity, is obviously true if $n = 1$, for it reverts to the formula

$$\Omega\, s_r = \frac{d}{dx} x^r = rx^{r-1}.$$

§ 7. The proof of Theorem III depends on that of IV which in turn needs a lemma concerning the principal minors of a general determinant, which must now be considered. A proof of V follows directly from IV and need not be given.

Let the determinant $|X| = |X_{ij}|$ be also written $(12\ldots n)_{12\ldots n}$ with two rows of n integers, the upper row referring to the columns and the lower to the rows of $|X|$, both in their correct order. Also let j_i denote the element x_{ij}, and $(ab)_{pq}$ the minor $a_p b_q - a_q b_p$ from the a^{th}, b^{th} columns and the p^{th} and q^{th} rows: and so on. Then the principal minors are typified by

$$i_i,\ (ij)_{ij},\ (ijk)_{ijk},\ \ldots \qquad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

with the same letters in both sets. Throughout, each letter denotes an integer 1, 2, ... n.

We now form the sum of all $\binom{n}{r}$ principal minors with r rows and columns, and take

$$p_r = (-)^r \Sigma (ij\ldots m)_{ij\ldots m}, \qquad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

where there are r letters $i, j, \ldots m$. This leads to n functions

$$\begin{aligned} p_1 &= -\Sigma i_i, \\ p_2 &= (-)^2 \Sigma (ij)_{ij}, \\ &\ldots\ldots\ldots\ldots \\ p_n &= (-)^n (ij\ldots)_{ij\ldots} = (-)^n |X|, \end{aligned} \qquad \ldots\ldots\ldots\ldots (3)$$

which by a well known theorem in determinants are the various coefficients in the characteristic equation

$$|\lambda - x_{ij}| \equiv \lambda^n + p_1\lambda^{n-1} + p_2\lambda^{n-2} + \ldots + p_n = 0. \qquad \ldots\ldots (4)$$

The lemma in question can now be enunciated.

Lemma. *If* $i \neq j$ *then*

$$x_{i1}\frac{\partial p_{r-1}}{\partial x_{j1}} + x_{i2}\frac{\partial p_{r-1}}{\partial x_{j2}} + \ldots + x_{in}\frac{\partial p_{r-1}}{\partial x_{jn}} = \frac{\partial p_r}{\partial x_{ji}},$$

and if $i = j$,

$$x_{i1}\frac{\partial p_{r-1}}{\partial x_{i1}} + x_{i2}\frac{\partial p_{r-1}}{\partial x_{i2}} + \ldots + x_{in}\frac{\partial p_{r-1}}{\partial x_{in}} - p_{r-1} = \frac{\partial p_r}{\partial x_{ii}}$$

where $r = 1, 2, \ldots n$. *If* $r = n + 1$, *the right members of these identities are replaced by zero.*

Proof. There is no difficulty in proving these if $n < 3$ or $r < 3$. We therefore take $n \geqslant r > 2$.

Consider $\frac{\partial p_{r-1}}{\partial x_{jk}}$, $j \neq k$. Since p_{r-1} is a sum of minors, and x_{jk} is an element of the original determinant, only those minors containing both indices j and k lead to a non zero term; also the result of differentiating each such minor gives a minor of order $r-2$. Thus

$$(-)^{r-1} \frac{\partial p_{r-1}}{\partial x_{jk}} = \Sigma \frac{\partial}{\partial x_{jk}} (jkab\,..\,c)_{jkab\,..\,c} \qquad \ldots\ldots\ldots\ldots(5)$$

Summed for $r-3$ integers $a, b, \ldots c$ chosen in $\binom{n-2}{r-3}$ ways from the integers $1, 2, \ldots n$, excluding j and k. But

$$\frac{\partial}{\partial x_{jk}} (jkab\,..\,c)_{jkab\,..\,c} = -\frac{\partial}{\partial x_{jk}} (kjab\,..\,c)_{jkab\,..\,c} = -(jab\,..\,c)_{kab\,..\,c} \qquad \ldots\ldots(6)$$

dropping the first entry in each index row, since x_{jk} is the element of row j and column k. Hence

$$(-)^{r} \frac{\partial p_{r-1}}{\partial x_{jk}} = \Sigma\, (jab\,..\,c)_{kab\,..\,c}; \qquad \ldots\ldots\ldots\ldots\ldots(7)$$

and as it is useful to exclude i as well as j and k from the values of $a, b, \ldots c$, we write this last as

$$\Sigma\, (jab\,..\,c)_{kab\,..\,c} + \Sigma\, (jib\,..\,c)_{kib\,..\,c}. \qquad \ldots\ldots\ldots\ldots(7')$$

Furthermore since $\frac{\partial}{\partial x_{jj}} (jda\,..\,c)_{jda\,..\,c} = (da\,..\,c)_{da\,..\,c}$, then

$$(-)^{r} \frac{\partial p_{r-1}}{dx_{jj}} = -\Sigma\, (dab\,..\,c)_{dab\,..\,c} - \Sigma\, (iab\,..\,c)_{iab\,..\,c} \qquad \ldots\ldots(8)$$

where d is any index unequal to i or j but included in the summation.

Multiplying each result (7) and (8) by its x_{ik} and summing for $k = 1, 2, \ldots n$ we obtain the following relation,

$$(-)^{r} \sum_{k} x_{ik} \frac{\partial p_{r-1}}{\partial x_{jk}} = \Sigma\, i_i\, (jab\,..\,c)_{iab\,..\,c} - \Sigma\, j_i\, (dab\,..\,c)_{dab\,..\,c}$$
$$- \Sigma\, j_i\, (iab\,..\,c)_{iab\,..\,c} + \Sigma\, d_i\, (jab\,..\,c)_{dab\,..\,c} + \Sigma\, d_i\, (jib\,..\,c)_{dib\,..\,c} \qquad \ldots(9)$$

the five sums on the right occurring by putting $k = i, j, d$ in turn and dropping obviously zero terms with repeated lower indices. These are summed for $a, b, \ldots c, d$ excluding i and j. But

$$\Sigma\, d_i\, (jib\,..\,c)_{dib\,..\,c} = -\Sigma\, a_i\, (jib\,..\,c)_{iab\,..\,c}.$$

Also for a fixed group of lower indices, the $r-2$ terms

$$i_i\, (jab\,..\,c)_{iab\,..\,c} - j_i\, (iab\,..\,c)_{iab\,..\,c} - \Sigma\, a_i\, (jib\,..\,c)_{iab\,..\,c}$$

vanish, for they equal the zero determinant

$$(ijab\,..\,c)_{iiab\,..\,c}.$$

https://doi.org/10.1017/S0013091500007434 Published online by Cambridge University Press

This disposes of three groups on the right of (9). For the other two we have

$$-\Sigma j_i (dab \,..\, c)_{dab \,..\, c} + \Sigma d_i (jab \,..\, c)_{dab \,..\, c} = -\Sigma (jdab \,..\, c)_{idab \,..\, c} \quad \ldots\ldots (10)$$

with $\binom{n-2}{r-2}$ terms on the right, due to combinations of $r-2$ letters $d, a, b, \ldots c$. But this last is $\Sigma \frac{\partial}{\partial x_{ji}} (ijdab \,..\, c)_{ijdab \,..\, c}$, which is $(-)^r \frac{\partial p_r}{\partial x_{ji}}$ Hence (9) can be written

$$\sum_k x_{ik} \frac{\partial p_{r-1}}{\partial x_{jk}} = \frac{\partial p_r}{\partial x_{ji}}$$

proving the first part of the lemma.

The second case, when $i = j$, leads likewise through a formula such as (9) to

$$(-)^r \sum_k x_{jk} \frac{\partial p_{r-1}}{\partial x_{jk}} = -\Sigma (jdab \,..\, c)_{jdab \,..\, c} \quad \ldots\ldots\ldots\ldots (11)$$

But

$$(-)^r \frac{\partial p_r}{\partial x_{jj}} = \Sigma \frac{\partial}{\partial x_{jj}} (jkdab \,..\, c)_{jkdab \,..\, c} \quad \ldots\ldots\ldots\ldots (12)$$

$$= \Sigma (kdab \,..\, c)_{kdab \,..\, c}$$

where j alone is excluded from the indices. Furthermore

$$\Sigma (jdab \,..\, c)_{jdab \,..\, c} + \Sigma (kdab \,..\, c)_{kdab \,..\, c} = (-)^{r-1} p_{r-1}. \quad \ldots\ldots (13)$$

Combining these last three results,

$$\sum_k x_{jk} \frac{\partial p_{r-1}}{\partial x_{jk}} - p_{r-1} = \frac{\partial p_r}{\partial x_{jj}}$$

which completes the proof of the lemma if $r < n+1$. If $r = n+1$ the term on the right of (10) is zero, and the desired result follows.

Proof of Theorem IV.

§ 8. By taking all values of i and j we can write the formulae of the lemma as a single matrix equation

$$X \,.\, \Omega\, p_{r-1} - p_{r-1} = \Omega\, p_r, \quad \ldots\ldots\ldots\ldots\ldots (1)$$

for $\Omega\, p_{r-1}$ is the matrix $\left[\frac{\partial p_{r-1}}{\partial x_{ji}}\right]$, and on multiplying this forwards by X and subtracting the scalar matrix p_{r-1}, we obtain the matrix whose n^2 elements are precisely those of $\Omega\, p_r$, as the lemma shews.

https://doi.org/10.1017/S0013091500007434 Published online by Cambridge University Press

Taking $r = 1, 2, 3 \ldots n$, in succession we immediately deduce the formulae of Theorem III from (1), ending with the Caley-Hamilton result, which answers to taking $r = n+1$, and $p_{n+1} = 0$. Thus we can write

$$X^r + p_1 X^{r-1} + \ldots + p_r + \Omega\, p_{r+1} = 0 \qquad \ldots\ldots\ldots\ldots(2)$$

$r = 0, 1, 2, \ldots\ldots$, where $p_{n+i} = 0,\ i > 0$.

Proof of Theorem III.

§ 9. We may prove this by induction, assuming the formula for $\Omega\, s_{r-1}$ to be true,

Since $s_1 = x_{11} + x_{22} + \ldots + x_{nn}$ it follows directly that the formula is true for s_1. We assume it true for s_{r-1} and proceed to deduce it for s_r. Also to effect this we operate with Ω upon a scalar function, and incidentally use these readily verified results:

$$\Omega\, c = 0,\ \ \Omega\, \theta\phi = \phi\, \Omega\, \theta + \theta\, \Omega\, \phi = \theta'\phi + \theta\phi', \qquad \ldots\ldots\ldots(1)$$

where c is constant, θ, ϕ are scalar, and the accent denotes the effect of the operator.

In fact by operating with Ω on the identity

$$s_r + p_1 s_{r-1} + p_2 s_{r-2} + \ldots\ldots + p_{r-1} s_1 + r p_r = 0 \qquad \ldots\ldots(2)$$

we get

$$\Omega\, s_r + p_1 (r-1) X^{r-2} + p_2 (r-2) X^{r-3} + \ldots + p_{r-1}$$
$$+ p_1' s_{r-1} + p_2' s_{r-2} + \ldots + p_{r-1}' s_1 + r p_r' = 0. \qquad \ldots\ldots\ldots\ldots(3)$$

Substituting the values of $p_1', p_2', \ldots p_r'$ from § 8 (2) and arranging the result in descending powers of X, we obtain $-r$ for the coefficient of X^{r-1}, due to the term $r p_r'$. The coefficient of X^{r-2} is zero, since

$$p_1 (r-1) - s_1 - r p_1 = 0,$$

and in general, if $i = 1, 2, \ldots r$, the coefficient of X^{r-i} is zero, since

$$p_{i-1} (r - i + 1) - s_1 p_{i-2} - s_2 p_{i-3} \ldots - s_{i-1} - r p_{i-1} = 0.$$

All that is left is

$$\Omega\, s_r - r X^{r-1} = 0$$

which proves the theorem if $r < n+1$.

Next to prove it if $r > n$, we operate on

$$s_r + p_1 s_{r-1} + \ldots + p_n\, s_{r-n} = 0 \qquad \ldots\ldots\ldots\ldots(4)$$

so that

$$\Omega\, s_r + p_1 (r-1) X^{r-2} + \ldots + p_n (r-n) X^{r-n-1}$$
$$+ p_1' s_{r-1} + \ldots + p_n' s_{r-n} = 0. \qquad \ldots\ldots\ldots\ldots\ldots(5)$$

But $$p_1 X^{r-2} + p_2 X^{r-3} + \ldots . + p_n X^{r-n-1} = - X^{r-1}$$

by the Cayley-Hamilton theorem. Hence, after rearranging terms,

$$\Omega s_r - rX^{r-1} - p_1 X^{r-2} - 2p_2 X^{r-3} \ldots - np_n X^{r-n-1} + p_1' s_{r-1} + \ldots + p_n' s_{r-n} = 0. \quad \ldots\ldots\ldots\ldots\ldots\ldots (6)$$

The theorem is therefore true provided

$$p_1 X^{r-2} + 2p_2 X^{r-3} \ldots + np_n X^{r-n-1} = p_1' s_{r-1} + \ldots + p_n' s_{r-n} \quad \ldots (7)$$

where $r > n$. But this last can be proved by induction if

(1) it is true when $r = n + 1$,

(2) $X(p_1' s_{r-1} + \ldots + p_n' s_{r-n}) = p_1' s_r + p_2' s_{r-1} + \ldots + p_n' s_{r-n+1}$,

since the left hand member of (7) is only altered by the presence of a new factor X when $r - 1$ is replaced by r.

Now these two conditions are readily verified. For (1) if $r = n + 1$, the formula becomes

$$p_1 X^{n-1} + 2p_2 X^{n-2} + \ldots + np_n = - s_n - (X + p_1) s_{n-1} \ldots - (X^{n-1} + \ldots + p_{n-1}) s_1$$

after using Theorem IV. Taking all terms to one side the coefficient of X^{n-i} is

$$ip_i + p_{i-1} s_1 + p_{i-2} s_2 + \ldots + s_i$$

which vanishes identically for all requisite values of i. And again for condition (2), we substitute for each p_i' and obtain

$$X\{s_{r-1} + (X + p_1) s_{r-2} + \ldots . + (X^{n-1} + \ldots + p_{n-1}) s_{r-n}\} = s_r + (X + p_1) s_{r-1} + \ldots + (X^{n-1} + \ldots + p_{n-1}) s_{r-n+1}$$

or, multiplying out by X and using § 8 (2) on the last term on the left

$$Xs_{r-1} + (X^2 + Xp_1) s_{r-2} + \ldots + (X^{n-1} + \ldots + p_{n-2} X) s_{r-n+1} - p_n s_{r-n} = s_r + (X + p_1) s_{r-1} + \ldots + (X^{n-1} + \ldots + p_{n-1}) s_{r-n+1}.$$

Equating coefficients of X^i, $i = 0, 1, 2, \ldots n - 1$, the results are identically equal, owing to (4)

This proves Theorem III, that *for positive integral values of* r.

$$\left[\frac{\partial}{\partial x_{ji}}\right] (\lambda_1{}^r + \lambda_2{}^r + \ldots + \lambda_n{}^r) \equiv \Omega s_r = rX^{r-1} = r\,[x_{ij}]^{r-1}.$$

§ 10. We may notice still another form of the result. For after using the Cayley-Hamilton theorem we can reverse the relations of Theorem III, writing

$$\begin{aligned} Xp_n' &= p_n \\ X^2 p_{n-1}' &= p_{n-1} X + p_n \\ X^3 p_{n-2}' &= p_{n-2} X^2 + p_{n-1} X + p_n \qquad \ldots\ldots\ldots\ldots(1) \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ X^n p_1' &= p_1 X^{n-1} + p_2 X^{n-2} + \ldots + p_n \end{aligned}$$

But $p_n = (-)^n |X|$ and $p_n' = (-)^n \left[\dfrac{\partial |X|}{\partial x_{ji}}\right]$ where the element of the matrix is therefore cofactor of x_{ji} in $|X|$. Then if $Xp_n' = p_n$ is written

$$\frac{1}{X} = \frac{p_n'}{p_n} = \left[\frac{\partial |X|}{\partial x_{ji}} / |X|\right] \qquad \ldots\ldots\ldots\ldots\ldots(2)$$

we come back to the well known result that the elements of the reciprocal matrix are given by those of the adjugate determinant divided by $|X|$. The other formulae generalize on this.

Generalization of Theorem III.

§ 11. Further the formula $\Omega s_r = rX^{r-1}$ holds for zero and negative integral values of r. For a similar argument applies to the reverse formulae

$$\begin{aligned} p_n s_{-1} + p_{n-1} &= 0 \\ p_n s_{-2}' + p_{n-1} s_{-1} + 2p_{n-2} &= 0, \text{ etc.} \qquad \ldots\ldots\ldots\ldots(1) \end{aligned}$$

where $s_{-i} = \Sigma \lambda^{-i}$. We can proceed by induction from s_{-r} to s_{-r-1} provided the formula is true if $r = 1$.

But $p_n s_{-1} + p_{n-1} = 0$, whence after operating with Ω

$$p_n' s_{-1} + p_n s_{-1}' + p_{n-1}' = 0.$$

Substituting from (1), we have

$$Xp_n s_{-1} + X^2 p_n s_{-1}' + Xp_{n-1} + p_n = 0$$

so that $s_{-1}' = -X^{-2}$, which is what we want to prove.

More generally, if $f(X)$ is an analytic function of X, with scalar coefficients, capable of development in a power series ascending or descending, or even both, as,

$$\sum_r (a_r X^r + b_r X^{-r})$$

https://doi.org/10.1017/S0013091500007434 Published online by Cambridge University Press

then we may apply the formula $\Omega\, s_r = rX^{r-1}$ and obtain the general result

$$\Omega \sum_{i=1}^{n} f(\lambda_i) = \sum_{i=1}^{n} \Omega f(\lambda_i) = f'(X), \qquad \text{..........(2)}$$

where $f'(\lambda)$ is the ordinary scalar derived function $\dfrac{df(\lambda)}{d\lambda}$.

Replacing $f'(X)$ by $f(X)$ and $f(\lambda)$ by $\int f(\lambda)\, d\lambda$ we deduce the alternative forms of the same theorem

$$f(X) = \Omega \sum_{i=1}^{n} \int f(\lambda_i)\, d\lambda_i. \qquad \text{...............(3)}$$

$$f([x_{pq}]) = \left[\frac{\partial}{\partial x_{qp}} \sum_{i=1}^{n} \int f(\lambda_i)\, d\lambda_i\right]. \qquad \text{...............(4)}$$

But again, writing λ short for λ_i and $\psi(\lambda) = \int f(\lambda)\, d\lambda$, or $f(\lambda) = \psi'(\lambda)$, we can evaluate $\Omega\, \psi(\lambda)$ as follows.
We have

$$\frac{\partial}{\partial x_{ji}} \psi(\lambda) = \frac{d\psi}{d\lambda}, \quad \frac{\partial \lambda}{\partial x_{ji}} = f(\lambda) \cdot \frac{\partial \lambda}{\partial x_{ji}}.$$

Hence

$$\left[\frac{\partial \psi(\lambda)}{\partial x_{ji}}\right] = f(\lambda) \left[\frac{\partial \lambda}{\partial x_{ji}}\right].$$

This gives by (3)

$$f(X) = f(\lambda_1)\Lambda_1 + f(\lambda_2)\Lambda_2 + \ldots + f(\lambda_n)\Lambda_n \qquad \text{........(5)}$$

where Λ_r is the matrix $\left[\dfrac{\partial \lambda_r}{\partial x_{ji}}\right] \equiv \Omega\, \lambda_r$. Hence we have the general result:—

THEOREM VI. *A function $f(X)$ of a matrix X can be expressed as a linear function of the n matrices Λ_r obtained by operating with Ω on the n latent roots λ_r of X.*

By taking n linearly independent particular functions, say $f(X) = 1,\ X,\ X^2, \ldots.\ X^{n-1}$ we obtain a system of n equations (5) which can be solved for the Λ's, provided the latent roots are all distinct. This expresses each Λ_r as a polynomial in X with coefficients rational in the λ's. Thereby any other function $f(X)$ can be evaluated. We obtain, in fact

$$\Lambda_r = \frac{\Pi(X - \lambda_i)}{\Pi(\lambda_r - \lambda_i)}, \quad \begin{matrix} i = 1, 2, \ldots n \\ r = 1, 2, \ldots n \end{matrix} \quad i \neq r. \qquad \text{........(6)}$$

Further it may be noted that the convergency of a matrix power series,

$$P(X) \equiv a_0 + a_1 X + a_2 X^2 + \ldots + a_r X^r + \ldots$$

with scalar coefficients, is guaranteed if all n latent roots lie within the circle of convergence of the series $P(z)$, z being a complex scalar number. This is manifest by (5).

It should be remarked that such a function as $\sqrt{\lambda}$, which cannot be expanded in a power series near $\lambda = 0$, admits of this treatment. For if $|\mu| < 1$ we have

$$(1+\mu)^{\frac{1}{2}} = 1 + \tfrac{1}{2}\mu - \tfrac{1}{8}\mu^2 + \ldots\ldots = g(\mu), \text{ say};$$

whence by the Theorem, if μ_r is a latent root of $[y_{ij}]$,

$$g([y_{ij}]) = \sum_r g(\mu_r)\left[\frac{\partial}{\partial y_{ji}}\right]\mu_r.$$

But let $[x_{ij}] = 1 + [y_{ij}]$, so that $\lambda_r = 1 + \mu_r$. Then

$$\frac{\partial}{\partial y_{ji}} = \frac{\partial}{\partial x_{ji}}:$$

whence

$$g(X-1) = \sum_r g(\mu_r)\,\Omega(\lambda_r - 1).$$

So

$$X^{\frac{1}{2}} = \sum_r \sqrt{\lambda_r}\,\Omega\,\lambda_r;$$

and indeed the argument is perfectly general. It covers for instance the case of the Corollary § 5 (3) for an analytic function:

$$\Omega f(X) = \sum_\lambda \frac{f(X) - f(\lambda)}{X - \lambda}.$$

This particular result is a kind of limit property, closely akin to Fermat's Formula

$$\frac{df(x)}{dx} = \lim \frac{f(x+h) - f(x)}{h}.$$

For although the fraction $\{f(X) - f(\lambda)\}/(X - \lambda)$ may formally be calculated as a scalar quantity, assuming that $X - \lambda \neq 0$, and obtaining say $F(X)$ as result, we are in fact dealing with a ratio of *singular* matrices, since $|X - \lambda| = 0$ and, by Sylvester's Theorem, $|f(X) - f(\lambda)| = 0$ also. Conversely it is interesting to note that if n

scalar numbers λ are fixed and X is an arbitrary variable matrix, these n fractions only yield a singular denominator when X assumes a value giving the λ's for its latent roots.

[*Note added January* 1928. A direct proof of Theorem III can be given by differentiating the formula

$$s_r = \Sigma\, x_{\alpha\beta}\, x_{\beta\gamma} \ldots x_{\lambda\alpha}$$

with regard to x_{ji}. This formula follows from Sylvester's Theorem, § 4. The summation has r suffixes $\alpha, \beta, \ldots \lambda$ each running from 1 to n, and the result is cyclically symmetrical in $\alpha\beta \ldots \lambda$. Thus differentiation leads to the ij^{th} term in the matrix rX^{r-1}, which yields the desired theorem.

Theorem IV now follows by reversing the steps throughout §§8, 9. A further result can be given as follows:

Theorem VII $\quad \Omega(\Omega X - X\Omega) f(X) = f'(X)$.

For by theorem II, it can be shewn that $(\Omega X - X\Omega) X^r = s_r$. Hence $\Omega(\Omega X - X\Omega) X^r = \Omega s_r = rX^{r-1}$. Provided $f(z)$ can be dealt with term by term, as above, the result follows. Hence

For a scalar function of a single matrix X, the operator $\Omega(\Omega X - X\Omega)$ behaves like an ordinary differential operator.]