Matrix Differentiation of the Characteristic Function

By H. W. TURNBULL, St Andrews University.

(*Received* 12*th January* 1931. *Read* 16*th January* 1931.)

The following work is a sequel to three previous communications,[1] and more particularly to the first. The present object is to shew the effect of repeated operation with the matrix differential operator $\Omega \equiv \left[\frac{\partial}{\partial x_{ji}}\right]$, when it acts upon a scalar matrix formed from an n rowed determinant $|x_{ij}|$, or sums of principal minors, the n^2 elements x_{ij} being treated as independent variables. Thus when z is a scalar quantity Ωz means the matrix $[\partial z/\partial x_{ji}]$, whose ij^{th} element is the derivative $z/\partial x_{ji}$.

§ 1. *Fundamental Formulae.*

From the square matrix

$$X = [x_{ij}] = \begin{bmatrix} x_{11} & \dots & x_{1n} \\ \dots & \dots & \dots \\ x_{n1} & \dots & x_{nn} \end{bmatrix} \tag{1}$$

there may be derived a determinant $|X|$ and a characteristic function $\phi(\lambda)$, given by

$$\phi(\lambda) \equiv |\lambda I - X| \equiv \begin{vmatrix} \lambda - x_{11} & \dots\dots & - x_{1n} \\ \dots & \dots & \dots \\ - x_{n1} & \dots & \lambda - x_{nn} \end{vmatrix} \tag{2}$$

$$= p_0 \lambda^n + p_1 \lambda^{n-1} + \dots + p_{n-1} \lambda + p_n. \tag{3}$$

Clearly p_n is equal to $(-)^n |X|$, while $p_0 = 1$. The reciprocal of this polynomial $\phi(\lambda)$ can be expanded in the form

$$\psi(\lambda) = \frac{1}{\phi(\lambda)} = \frac{h_0}{\lambda^n} + \frac{h_1}{\lambda^{n+1}} + \frac{h_2}{\lambda^{n+2}} + \dots \tag{4}$$

---

[1] I. H. W. Turnbull, On differentiating a matrix, *Proc. Edinburgh Math. Soc.* (2), 1 (1927), 111-128.

II. A matrix form of Taylor's Theorem (2), 2 (1929), 33-54.

III. The invariant theory of bilinear forms, *Proc. London Math. Soc.* (1931).

for suitably large values of the modulus of λ, where the coefficients h_r are homogeneous products of the n latent roots λ_i of X, defined by $\phi(\lambda_i) = 0$. The coefficients p and h satisfy the well known Wronskian relations

$$h_r p_0 + h_{r-1} p_1 + h_{r-2} p_2 + \ldots + h_1 p_{r-1} + h_0 p_r = 0, \tag{5}$$

where $r = 1, 2, \ldots$. The unit matrix is denoted by $I = [\delta_{ij}]$ in terms of the Kronecker delta; and an arbitrary constant matrix by $A = [a_{ij}]$. Both λ and the a_{ij} are independent of the x_{ij}, whereas the h_r and p_r are clearly functions of the x_{ij}. As usual s_r denotes the sum of the r^{th} powers of the n latent roots λ_i.

By $\Omega\,\theta$ is meant the matrix $[\partial\theta/\partial x_{ji}]$ whose ij^{th} element is $\partial\theta/\partial x_{ji}$, θ being a scalar quantity. Taking θ to be s, p and h in turn, the fundamental formulae of Ω differentiation (*Cf.* I, p. 119) are

$$\Omega\, s_r = rX^{r-1}, \tag{6}$$

$$P_r = X^r + p_1 X^{r-1} + \ldots + p_{r-1} X + p_r I = -\Omega\, p_{r+1}, \tag{7}$$

$$H_r = X^r + h_1 X^{r-1} + \ldots + h_{r-1} X + h_r I = \quad \Omega\, h_{r+1}. \tag{8}$$

It is useful to have a special notation P and H for these polynomial scalar functions of the matrix X, whose order is shewn by the suffix. Initially r is taken to be zero or a positive integer, so that $P_0 = H_0 = I$; when $r \geqslant n$, the right member of (7) disappears, p_r being zero, and the Cayley Hamilton equation

$$P_n \equiv \phi(X) \equiv X^n + p_1 X^{n-1} + \ldots + P_{n-1} X + P_n I = 0 \tag{9}$$

is put in evidence.

The reciprocal properties (7) and (8) are brought out very clearly by the following new proof, which is based on the inverse of the λ-matrix $\lambda I - X$.

Letting X_{ij} denote the cofactor of x_{ij} in the determinant $|X|$, we may write the reciprocal of the non-singular matrix X in the form

$$X^{-1} = [X_{ji}]/|X|.$$

But we have

$$X_{ji} = \frac{\partial}{\partial x_{ji}} |X|;$$

hence

$$[X_{ji}] = \Omega|X|, \quad X^{-1} = \Omega|X|/|X|. \tag{12}$$

Let each x_{ij} be replaced by $x_{ij} - \lambda a_{ij}$, where λ and a_{ij} are constants. This leaves $\partial/\partial x_{ji}$, and therefore Ω unaltered, but replaces the matrix X by $X - \lambda A$. Accordingly we have the relation

$$\frac{1}{X - \lambda A} = \frac{\Omega | X - \lambda A |}{| X - \lambda A |}, \tag{13}$$

identically for all values of λ and a, a result which can also be exhibited as

$$\frac{1}{X - \lambda A} = \Omega \log | X - \lambda A |. \tag{14}$$

In particular let A be replaced by the unit matrix I. Then

$$- \log | X - \lambda I | = - \log (\lambda_1 - \lambda)(\lambda_2 - \lambda) \ldots . (\lambda_n - \lambda)$$
$$= - \log (-\lambda)^n + \frac{s_1}{\lambda} + \frac{s_2}{2\lambda^2} + \frac{s_3}{3\lambda^3} + \ldots .$$

for large enough values of the modulus of λ, while

$$-(X - \lambda I)^{-1} = \frac{I}{\lambda} + \frac{X}{\lambda^2} + \frac{X}{\lambda^3} + \ldots .$$

Result (6) follows at once by substituting these values in (14) and comparing coefficients of corresponding negative powers of λ. More generally, if $A^{-1} = C$, the same procedure leads to the relation

$$\Omega_s (CX)^r = r (CX)^{r-1} C \tag{15}$$

in the notation of II, p. 37.

To obtain the relation (7), let (13) be written in the form

$$\frac{| \lambda A - X |}{\lambda A - X} = - \Omega | \lambda A - X |. \tag{16}$$

Treating numerator and denominator of the left member as a polynomial and a linear function of λ, we may perform ordinary long division in every case when A commutes with X. This is so when $A = I$, making the left member $\phi(\lambda) \div (I\lambda - X)$. The polynomial $\phi(\lambda)$ is given by (3); on carrying out the long division the result is

$$\frac{\phi(\lambda)}{I\lambda - X} = I\lambda^{n-1} + (X + p_1 I)\lambda^{n-2} + \ldots . + (X^{n-1} + \ldots . + p_{n-1} I)$$
$$+ \frac{\phi(X)}{I\lambda - X}$$
$$= P_0 \lambda^{n-1} + P_1 \lambda^{n-2} + \ldots . + P_{n-1} + \frac{P_n}{I\lambda - X}. \tag{17}$$

https://doi.org/10.1017/S0013091500007811 Published online by Cambridge University Press

Again from the right member of (16), with $A = I$, we obtain

$$-(\lambda^{n-1}\Omega p_1 + \lambda^{n-2}\Omega p_2 + \ldots + \Omega p_n),$$

since $\Omega \lambda^n = 0$. On multiplying throughout, here and in (17), by $I\lambda - X$, expanding, and equating coefficients of powers of λ, we obtain the relations (7), and also the Cayley Hamilton theorem implied by $\phi(X) = 0$.

Reciprocally, since $\phi(\lambda)\psi(\lambda) = 1$, it follows that

$$\{\Omega\phi(\lambda)\}\psi(\lambda) + \phi(\lambda)\Omega\psi(\lambda) = 0;$$

but, since

$$\phi(\lambda) = |\lambda I - X|,$$

we have

$$\frac{\Omega\phi(\lambda)}{\phi(\lambda)} = \frac{1}{X - \lambda I} = -\frac{\Omega\psi(\lambda)}{\psi(\lambda)}. \tag{18}$$

Again by ordinary long division of the series (4) by $I - X\lambda^{-1}$, arranged in descending powers (all negative) of λ, we have

$$\frac{\psi(\lambda)}{I - X\lambda^{-1}} = \lambda^{-n} + (X + h_1 I)\lambda^{-n-1} + (X^2 + h_1 X + h_2 I)\lambda^{-n-2} + \ldots$$
$$= H_0\lambda^{-n} + H_1\lambda^{-n-1} + H_2\lambda^{-n-2} + \ldots$$

Also

$$\Omega\psi(\lambda) = h_1\Omega\lambda^{-n-1} + h_2\Omega\lambda^{-n-2} + \ldots$$

On substituting in (18), clearing of fractions, and comparing coefficients as before, the relations (8) follow. Incidentally we have the result[1]

$$\frac{1}{I\lambda - X} = \frac{P_0\lambda^{n-1} + P_1\lambda^{n-2} + \ldots + P_{n-2}\lambda + P_{n-1}}{(\lambda-\lambda_1)(\lambda-\lambda_2)\ldots(\lambda-\lambda_n)} \tag{19}$$
$$= (H_0 + H_1\lambda^{-1} + H_2\lambda^{-2} + \ldots)(\lambda-\lambda_1)\ldots(\lambda-\lambda_n)\lambda^{-n-1}.$$

These coefficients P and H are matrices which commute with X and with each other, since they are polynomials in X. From this relation each P_r can be deduced as a linear function of the H_s with $s \leqslant r$, the coefficients being polynomial expressions in the p's. Correlatively for H in terms of P. Also if the r^{th} Wronskian relation (5) is written $w_r(h, p) = 0$, it follows that

$$w_r(H, p) = w_r(h, P). \tag{20}$$

For example $H_2 p_0 + H_1 p_1 + H_0 p_2 = h_2 P_0 + h_1 P_1 + h_0 P_2$.

---

[1] *Cf.* L. E. Dickson, *Modern Algebraic Theories* (Chicago, 1926), 48, after replacing P_r by C_{n-1-r}.

§ 2. *The Converse Problem.*

By solving the recurrence relations (7) and (8) for successive powers of X we obtain the following equations, in which an accent denotes the effect of the Ω operation:

$$- p_1' = I \;\; = h_1',$$
$$- h_1 p_1' - p_2' = X = h_2' + h_1' p_1,$$
$$- h_2 p_1' - h_1 p_2' - p_3' = X^2 = h_3' + h_2' p_1 + h_1' p_2,$$

and in general (since $p_0' = h_0' = 0$),

$$- w_r (h, p') = X^{r-1} = w_r (h', p). \tag{21}$$

These follow at once from (18), on multiplying throughout by $\phi(\lambda)\psi(\lambda)$ (which is unity), then expanding each of the three expressions in descending powers of λ, and again equating coefficients. These alternative expressions for a power of X lead to the theorem:

The $(r-1)^{\text{th}}$ *power of a matrix* X *is obtained by* Ω *differentiation from the* r^{th} *Wronksian relation, either by treating the* p*'s as constants, or else by treating the* h*'s as constants and affixing a negative sign to the result.*

§ 3. *Successive* Ω *differentiation.*

THEOREM I. *Any two consecutive coefficients* p_r, p_{r+1} *of the characteristic function* $\phi(\lambda)$ *satisfy the matrix differential equation*

$$\Omega^2 p_{r+1} = (n - r)\, \Omega p_r. \tag{22}$$

Proof. The left member of this equation denotes the effect of Ω operating upon Ωp_{r+1}, and is therefore equal to

$$- \Omega (X^r + p_1 X^{r-1} + \dots + p_{r-1} X + p_r I).$$

Now, by I, p. 117 (2),

$$\Omega X^\nu = s_0 X^{\nu-1} + s_1 X^{\nu-2} + \dots + s_{\nu-1} I, \tag{23}$$

where $s_0 = n$. Also $\Omega p_{r-\nu} X^\nu = p_{r-\nu} \Omega X^\nu + (\Omega p_{r-\nu}) X^\nu$. Let this last be simplified, by use of (7) and (23), and arranged in descending powers of X. On summing the results for $\nu = 0, 1, 2, \dots, r$ we have

$$\Omega^2 p_{r+1} = q_0 X^{r-1} + q_1 X^{r-2} + \dots + q_{r-2} X + q_{r-1} I,$$

where

$$q_m = - (s_m + p_1 s_{m-1} + \dots + p_m s_0) + (r - m) p_m.$$

After using the Newtonian relation

$$s_m + p_1 s_{m-1} + \ldots + p_{m-1} s_1 + m p_m = 0$$

q_m becomes $(r - n) p_m$. Hence

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

which proves the theorem.

THEOREM II. *Correlatively, consecutive coefficients h_{r+1}, h_r satisfy the equation*

$$\Omega^2 h_{r+1} = (n+r)\,\Omega\, h_r. \tag{24}$$

Proof. The proof is analogous to that of Theorem I, but utilizes the relation

$$s_m + h_1 s_{m-1} + \ldots + h_{m-1} s_1 = m h_m.$$

As a consequence of these two theorems we may express each matrix P_ν and H_ν as a matrix derivative of p_μ and h_μ, respectively, provided that the suffix μ exceeds ν. For example,

$$\Omega^3 p_{r+1} = (n-r)\,\Omega^2 p_r = (n-r)(n-r-1)\,\Omega\, p_{r-1}.$$

This leads straightforwardly to the relations

$$\begin{aligned}\Omega^{n-r} p_n &= \Omega^{n-r-1} p_{n-1} = 2!\,\Omega^{n-r-2} p_{n-2} = \ldots \\ &= (n-r-1)!\,\Omega\, p_{r+1} = -(n-r-1)!\,P_r,\end{aligned} \tag{25}$$

where $r = 0, 1, \ldots, n-1$. In particular, when $r = 0$, the result may be written

$$\Omega^m p_m = -(n-1)!/(n-m)!\,, \qquad 0 < m \leqslant n, \tag{26}$$

so that the effect of m operations with Ω upon the coefficient p_m in the characteristic function, yields a negative integer.

Similarly from Theorem II,

$$H_r = \Omega\, h_{r+1} = \frac{(n+r)!}{(n+r+1)!}\,\Omega^2 h_{r+2} = \frac{(n+r)!}{(n+r+2)!}\,\Omega^3 h_{r+3} = \ldots \tag{27}$$

THEOREM III. *Any power series*

$$f(X) = a_0 I + a_1 X + a_2 X^2 + \ldots \tag{28}$$

with scalar coefficients a_i can be derived from the scalar matrix $|X|\,I$ by means of a matrix operator $g(\Omega)$ which is a scalar polynomial, of order n, or less, in Ω.

Proof. On substituting for powers of X from (21) we have

$$f(X) = \beta_0 p_1' + \beta_1 p_2' + \dots + \beta_{n-1} p_n',$$

where

$$\beta_m = -a_m - a_{m+1} h_1 - a_{m+2} h_2 - \dots, \quad (m = 0, 1, 2, \dots).$$

Also by (25),

$$p_1' = \Omega p_1 = \frac{1}{(n-1)} \Omega^2 p_2 = \dots = \frac{1}{(n-1)!} \Omega^n p_n,$$

$$p_2' = \Omega p_2 = \frac{1}{(n-2)!} \Omega^{n-1} p_n;$$

and so on. Hence we have

$$f(X) = \left(\frac{\beta_0 \Omega^n}{(n-1)!} + \frac{\beta_1 \Omega^{n-1}}{(n-2)!} + \dots + \frac{\beta_{n-2} \Omega^2}{1!} + \beta_{n-1} \Omega \right) p_n.$$

$$= (-)^n g(\Omega) p_n, \text{ say.}$$

The theorem follows since $p_n = (-)^n |X|$.

COROLLARY. *Any polynomial $f(X)$ of order r less than n can be derived from an earlier coefficient p_m by an analogous operator $g_m(\Omega)$, whenever $m > r$.*

A similar theorem holds for the derivation of a polynomial $f(X)$ from a coefficient h_m of higher order. For example

$$X^2 = -\left(1 + \frac{h_1}{n-2} \Omega + \frac{h_2}{(n-1)(n-2)} \Omega^2 \right) \Omega p_3$$

$$= \left(1 + \frac{p_1}{n+2} \Omega + \frac{p_2}{(n+1)(n+2)} \Omega^2 \right) \Omega h_3.$$

THEOREM IV. *The operator $\Omega e^{\lambda \Omega}$ has the same effect upon $p_n = \phi(0)$, that Ω has upon the characteristic function $\phi(\lambda)$.*

Proof. We have

$$\Omega e^{\lambda \Omega} p_n = \left(\Omega + \lambda \Omega^2 + \frac{\lambda^2 \Omega^3}{3!} + \dots \right) p_n$$

$$= \Omega (\lambda^n + p_1 \lambda^{n-1} + \dots + p_n), \text{ by (25)},$$

$$= \Omega \phi(\lambda),$$

which proves the theorem.

We are not however entitled to deduce the equality of $e^{\lambda \Omega} p_n$ and $\phi(\lambda)$, by operating with Ω^{-1}, since it by no means follows that when $\Omega Y = 0$, Y itself is zero.

§ 4. *Connection with invariant theory.*

As has been pointed out in III (see Introduction), the Ω process is equivalent to polarization by use of a sum of symbolic operators

$$\left(u \left| \frac{\partial}{\partial a}\right.\right)\left(x \left| \frac{\partial}{\partial a}\right.\right) \equiv \left(\sum_{i=1}^{n} u_i \frac{\partial}{\partial a_i}\right)\left(\sum_{i=1}^{n} x_j \frac{\partial}{\partial a_j}\right),$$

where the matrix $[x_{ij}]$ is expressed in symbolic notation by various equivalents

$$[x_{ij}] = [\alpha_i a_j] = [\beta_i b_j] = [\gamma_i c_j] = \text{etc.} \tag{29}$$

In fact Ω is given by

$$u\,\Omega\,\xi = \Sigma\left(u \left| \frac{\partial}{\partial a}\right.\right)\left(\xi \left| \frac{\partial}{\partial \alpha}\right.\right), \tag{30}$$

where the summation runs through the equivalent symbols, one term for each pair a, α. The ij^{th} element of Ω is given by the coefficient of $u_i \xi_j$ in this expression (30); and $u\,\Omega\,x$ denotes the bilinear differential form $\Sigma\, u_i \dfrac{\partial}{\partial x_{ji}} \xi_j$ in the usual matrix product notation. The quantities p_r are now invariants of the bilinear form $\Sigma\, u_i x_{ij} \xi_j$; namely

$$p_1 = -(a \,|\, \alpha), \quad p_2 = \frac{1}{2!}(ab \,|\, \alpha\beta), \quad p_3 = -\frac{1}{3!}(abc \,|\, \alpha\beta\gamma), \; \ldots . \tag{31}$$

Since the effect of the right hand operation in (30) is to replace each pair a, α in the operand by u, x, formulae (7) are now almost intuitive. For example

$$u\,\Omega\,\xi\, p_2 = \frac{1}{2!}\left(\,(ub \,|\, \xi\beta) + (au \,|\, \alpha\xi)\,\right) = a_\alpha u_\xi - a_\xi u_\alpha,$$

since the symbols a, α are equivalent to b, β.

Translated back into the original notation this becomes

$$u\,(\Omega\, p_2)\,\xi = -\,u\,(X + p_1 I)\,\xi$$

identically for all u and ξ. Whence

$$X + p_1 I + \Omega\, p_2 = 0$$

and similarly for all relations (7).

Repeated Ω operation now appears as repeated polarization. For example

$$u\,\Omega^2\,\xi = \Sigma\left(u\,\middle|\,\frac{\partial}{\partial a}\right)\left(\frac{\partial}{\partial \alpha}\,\middle|\,\frac{\partial}{\partial b}\right)\left(\frac{\partial}{\partial \beta}\,\middle|\,\xi\right),$$

summed for all pairs of distinct equivalent symbols a, α and b, β. When this acts, for example, upon p_3 it strikes out two symbols a, b and two symbols α, β in every possible way, replacing them by the single u and ξ. This leads to the result

$$\Omega^2\,p_3 = (n-2)\,\Omega\,p_2 = -\,(n-2)\,(X + p_1\,I),$$

and similarly for other cases.

https://doi.org/10.1017/S0013091500007811 Published online by Cambridge University Press