Can. J. Math., Vol. XXV, No. 3, 1973, pp. 554–559

THE FRECHET DIFFERENTIAL OF A PRIMARY MATRIX FUNCTION

DAVID L. POWERS

Let X be a given complex matrix of order n. If $f(z)$ is analytic at the eigenvalues of X, one may define the primary matrix function $f(X)$ with stem function $f(z)$ by using any of several well-known methods: for instance, canonical forms, power series, or interpolating polynomials [**9**]. The Frechet differential of the primary matrix function f at X, evaluated for the "increment" H, is written $df(X; H)$ and is defined [**1**, p. 143] to be the unique linear function of H which satisfies

$$||f(X + H) - f(X) - df(X; H)|| = o(||H||).$$

If H commutes with X it is known [**6**] that $df(X; H) = Hf^{(1)}(X)$. (Throughout this paper, $f^{(k)}(X)$ denotes the primary matrix function with stem function $f^{(k)}(z)$.) Several attempts have been made to find a similarly satisfying expression for the differential under less restrictive conditions [**7**; **11**], but these have had only partial success because of the reliance on canonical forms to represent $f(X)$. The results of these investigations and the important role of commutativity motivate the following.

Definition. The successive commutes, or inner derivatives, of H with respect to X are defined by: $H_{(0)} = H$ and $H_{(k)} = H_{(k-1)}X - XH_{(k-1)}$ for $k \geqq 1$.

In [**12**] Roth investigated commutes in detail, and some additional information will be found in [**3**].

As we shall be dealing with linear functions on matrices, it is convenient to use some notation and results of Neudecker [**4**]. Let the columns of the $n \times n$ matrix H be $h_1, h_2, \ldots, h_n$. We define an isomorphism between the $n \times n$ matrices and the $n^2 \times 1$ matrices by $\operatorname{vec} H = [h_1^T, h_2^T, \ldots, h_n^T]^T$. It can be shown that

$$\operatorname{vec}(XH) = (I \times X) \operatorname{vec} H, \quad \operatorname{vec}(HX) = (X^T \times I) \operatorname{vec} H$$

where $\times$ denotes the Kronecker product. By induction, we find that $\operatorname{vec} H_{(k)} = (X^T \times I - I \times X)^k \operatorname{vec} H$ with the convention that the zeroth power of a square matrix is the identity.

In this notation, it is easy to prove that

$$(1) \qquad e^{-X}He^X = \sum_{n=0}^{\infty} \frac{1}{n!} H_{(n)}.$$

Received February 14, 1972 and in revised form, June 12, 1972.

Proof. $\operatorname{vec}(e^{-X}He^{X}) = [(e^{X})^{T} \times e^{-X}] \operatorname{vec} H$. But

$$(e^{X})^{T} \times e^{-X} = \exp(X^{T} \times I - I \times X).$$

(See [**4**].) Thus,

$$\operatorname{vec}(e^{-X}He^{X}) = \sum_{n=0}^{\infty} \frac{1}{n!} (X^{T} \times I - I \times X)^{n} \operatorname{vec} H$$
$$= \sum_{n=0}^{\infty} \frac{1}{n!} \operatorname{vec} H_{(n)}.$$

A second identity which will be needed is

$$X^{r-1}H + X^{r-2}HX + \ldots + HX^{r-1} = \sum_{s=1}^{r} \binom{r}{s} X^{r-s}H_{(s-1)} \tag{2}$$

The proof of (2) will be found in reference [**3**].

This latter formula, in fact, is the Frechet differential of the function $f(X) = X^{r}$. For, if $(X + H)^{r} - X^{r}$ is expanded, the terms which are linear in H are precisely the left-hand side of (2). From this point, we may pass to the first representation for a Frechet differential.

THEOREM 1. *Let* $f(z) = \sum_{r=0}^{\infty} a_{r}(z - c)^{r}$ *be analytic in a disk of centre* c *and radius* R. *Let* $\rho + \sigma < R$, *where* ρ *and* σ *are respectively the spectral radii of* $X - cI$ *and* $X^{T} \times I - I \times X$. *Then*

$$df(X;H) = \sum_{s=1}^{\infty} \frac{1}{s!} f^{(s)}(X)H_{(s-1)}. \tag{3}$$

Proof. We may safely assume that $c = 0$ and that a norm has been chosen to make $||X|| = \rho + \epsilon < R - \sigma$. Define $g(x) = \sum_{r=1}^{\infty} |a_{r}|\xi^{r}$, and take H so small that this series converges at $\xi_{1} = ||X|| + ||H||$. Then the series for $f(X + H)$ converges absolutely, and

$$f(X + H) - f(X) = \sum_{r=1}^{\infty} a_{r}[(X + H)^{r} - X^{r}]$$
$$= \sum_{r=1}^{\infty} a_{r}[X^{r-1}H + X^{r-2}HX + \ldots + HX^{r-1}] + o(||H||).$$

The remainder on the right, containing terms of order two or higher in H, is bounded in norm by $||H||^{2}g''(\xi_{1})$. The differential is transformed formally, using (2), as follows:

$$df(X;H) = \sum_{r=1}^{\infty} a_{r}[X^{r-1}H + X^{r-2}HX + \ldots + HX^{r-1}]$$
$$= \sum_{r=1}^{\infty} a_{r} \sum_{s=1}^{r} \binom{r}{s} X^{r-s}H_{(s-1)}$$
$$(*) \qquad = \sum_{s=1}^{\infty} \sum_{r=s}^{\infty} a_{r} \binom{r}{s} X^{r-s}H_{(s-1)}$$
$$= \sum_{s=1}^{\infty} \frac{1}{s!} f^{(s)}(X)H_{(s-1)}.$$

The rearrangement of the series is justified because, in the starred line, the inner sum is bounded in norm by $(s!)^{-1}g^{(s)}(\rho + \epsilon)$, and

$$\limsup \|H_{(\nu-1)}\|^{1/\nu} \leqq \sigma < R - \rho - \epsilon.$$

A special case of (3) was discovered and used by Fer [**2**] in the iterative solution of the matrix equation: $\dot{U} = H(t)U$, $U(0) = I$. He sets $U = e^X V$, where

$$X(t) = \int_0^t H(\tau)d\tau.$$

Since

$$(d/dt)f(X) = df(X; (d/dt)X) = df(X, H)$$

(see [**1**]), the time derivative of e^X is found to be, using (1),

$$(d/dt)e^X = e^X \sum_{s=1}^{\infty} \frac{1}{s!} H_{(s-1)}$$

$$= He^X - e^X \sum_{s=1}^{\infty} \frac{s}{(s+1)!} H_{(s)}.$$

Thus the matrix $V(t)$ satisfies

$$\dot{V} = \left[\sum_{s=1}^{\infty} \frac{s}{(s+1)!} H_{(s)} \right] V, \; V(0) = I$$

in which the coefficient of V vanishes with t at $t = 0$.

When H is required to satisfy some restriction relative to X, the hypotheses on the function can be relaxed, as follows.

COROLLARY 1 (Roth [**12**]). *Let P be a nonsingular matrix such that $P^{-1}XP = (\lambda_1 I + N_1) \oplus \ldots \oplus (\lambda_m I + N_m)$, where $\lambda_i \neq \lambda_j$ if $i \neq j$, and each N_i is nilpotent. If H is such that $P^{-1}HP = H_1 \oplus \ldots \oplus H_m$, where H_i is of the same order as $\lambda_i I + N_i$, then $H_{(k)} = 0$ for some k. If f is analytic at the eigenvalues of X, then*

$$df(X; H) = \sum_{s=1}^{k} \frac{1}{s!} f^{(s)}(X) H_{(s-1)}.$$

It is easily shown that $k \leqq 2\mu - 1$, μ being the greatest of the multiplicities of roots of the minimum equation of X. Thus for a diagonalizable X, $k = 1$ and the assumption is that H commutes with X.

A different sort of restriction on H can also provide a simplified representation for the differential.

LEMMA 1. *Let f be analytic at the eigenvalues of X, and let $H = MX - XM$ for some M. Then*

$$df(X; H) = Mf(X) - f(X)M.$$

Proof. Assume first that M is specified by

$$\text{vec}\, M = (X^T \times I - I \times X)^+ \,\text{vec}\, H$$

https://doi.org/10.4153/CJM-1973-056-1 Published online by Cambridge University Press

where $+$ denotes the Moore-Penrose generalized inverse [**5**], so that H and M are of the same order as H vanishes. Using the power series expansion for matrix exponentials, we find that

$$\begin{aligned} e^{-M}(X+H)e^{M} &= (I - M + - \ldots)(X+H)(I+M+\ldots) \\ &= X + H - MX + XM + o(\|H\|) \\ &= X + o(\|H\|). \end{aligned}$$

As $f(X)$ has a Frechet derivative [**10**], it satisfies a Lipschitz condition at X [**1**, p. 155]. Hence

$$\begin{aligned} f(X+H) &= e^{M}f(e^{-M}(X+H)e^{M})e^{-M} \\ &= e^{M}f(X + o(\|H\|))e^{-M} \\ &= e^{M}f(X)e^{-M} + o(\|H\|). \end{aligned}$$

Again employing the power series expansion for the exponential, and isolating the first order terms, yields

$$\begin{aligned} f(X+H) - f(X) &= f(X) + Mf(X) - f(X)M - f(X) + o(\|H\|) \\ &= Mf(X) - f(X)M + o(\|H\|). \end{aligned}$$

Thus the lemma is proved for the special choice of M. In general, M may be expressed as $M = M_1 + M_2$, where M_1 is as above and M_2 commutes with X; but then M_2 also commutes with $f(X)$. Hence the lemma holds in general.

Roth proves a similar theorem under more restrictive hypotheses and by different means [**12**]. For stem functions analytic in a sufficiently large region, application of (3) gives an alternate proof.

An immediate application of the Lemma can be made to the approximation of $f(Z)$ if $Z = [z_{ij}]$ has distinct diagonal entries. Set $X = \operatorname{diag}\{z_{11}, \ldots, z_{nn}\}$, so $f(Z) \simeq f(X) + df(X; Z - X) = [\phi_{ij}]$. The elements ϕ_{ij} are given by $\phi_{ii} = f(z_{ii})$ and, for $i \neq j$,

$$\phi_{ij} = z_{ij}\frac{f(z_{ii}) - f(z_{jj})}{z_{ii} - z_{jj}}.$$

This approximation would have been useful in [**13**] where Z is a Ritz matrix and $f(z) = (z)^{-\frac{1}{2}} \tanh(z^{\frac{1}{2}})$.

We shall now show that the two different restrictions on H mentioned in the Corollary and Lemma 1 are complementary.

LEMMA 2. *For any matrix H of order n, there exist matrices L and M satisfying: $H = L + MX - XM$, and $L_{(k)} = 0$ for some k.*

Proof. Let $P^{-1}XP = (\lambda_1 I + N_1) \oplus \ldots \oplus (\lambda_m I + N_m)$ as in Corollary 1. Let $P^{-1}HP = [H_{ij}]$, $P^{-1}LP = [L_{ij}]$, $P^{-1}MP = [M_{ij}]$ be partitioned conformally with $P^{-1}XP$. In order that $L_{(k)} = 0$ for some k, $L_{ij} = 0$ $(i \neq j)$ is necessary and sufficient [**12**]. The equations to be satisfied are thus:

(4) $$M_{ij}(\lambda_j I + N_j) - (\lambda_i I + N_i)M_{ij} = H_{ij} \qquad (i \neq j),$$

(5) $$L_{ii} = H_{ii} - M_{ii}N_i + N_i M_{ii}.$$

https://doi.org/10.4153/CJM-1973-056-1 Published online by Cambridge University Press

The submatrices M_{ii} are, in fact, arbitrary but we shall take them to be 0. Equation (4) determines M_{ij} uniquely, since $\lambda_i \neq \lambda_j$ for $i \neq j$ [**4**]. We also note that M is of the same order as H when H vanishes.

By combining the Lemmas and the Corollary, and using the linearity of the Frechet differential, one easily proves the principal representation below.

THEOREM 2. *Let $f(z)$ be analytic at the eigenvalues of X. Let $H = L + MX - XM$, where $L_{(k)} = 0$ for some k. Then*

$$df(X;H) = Mf(X) - f(X)M + \sum_{s=1}^{k} \frac{1}{s!} f^{(s)}(X)L_{(s-1)}.$$

Several derivatives for functions of matrices have been defined by limit processes [**8**; **10**]. It is also possible to obtain the Frechet differential via a limit.

THEOREM 3. *For sufficiently small $\epsilon \neq 0$, let $S = S(\epsilon)$ be defined by $SX - XS + \epsilon S = H$, and let $f(z)$ be analytic at the eigenvalues of X. Then*

$$df(X;H) = \lim_{\epsilon \to 0} \left[Sf(X) - f(X)S + \epsilon \sum_{r=1}^{2n-1} \frac{1}{r!} f^{(r)}(X)S_{(r-1)} \right]. \tag{6}$$

Proof. For $\epsilon \neq 0$ and sufficiently small, S exists and is unique. By substitution one verifies that

$$S(\epsilon) = M - \sum_{m=1}^{2n-1} (-\epsilon)^{-m} L_{(m-1)} + O(\epsilon)$$

where L and M are as in Lemma 2, and that the quantity in brackets in (6) is $df(X;H) + O(\epsilon)$. The upper limit in the sums may be reduced if more is known about the matrix X.

The Frechet derivative, that is, the linear function $H \to df(X;H)$, may be realized as a matrix of order n^2 if H and $df(X;H)$ are replaced by their images under "vec". We replace $X^T \times I - I \times X$ by Δ for brevity.

THEOREM 4. *Under the conditions of Theorem* 1, 2, *or* 3, *the Frechet derivative of f at X is*

$$\sum_{r=1}^{\infty} \frac{1}{r!} [I \times f^{(r)}(X)]\Delta^{r-1}, \tag{Th 1}$$

$$[f(X^T) \times I - I \times f(X)]\Gamma + \sum_{r=1}^{k} \frac{1}{r!} [I \times f^{(\nu)}(X)]\Delta^{\nu-1}\Pi \tag{Th 2}$$

where Γ is any matrix having the property that $\Delta^{k+1}\Gamma = \Delta^k$ for some integer k, and $\Pi \equiv I - \Delta\Gamma$, or

$$\lim_{\epsilon \to 0} \left\{ \left[f(X^T) \times I - I \times f(X) + \epsilon \sum_{r=1}^{2n-1} \frac{1}{r!} (I \times f^{(r)}(X))\Delta^{r-1} \right] (\Delta + \epsilon I)^{-1} \right\}. \tag{Th 3}$$

References

1. J. Diendonné, *Foundations of modern analysis* (Academic Press, New York, 1960).

2. F. Fer, *Résolution de l'équation matricielle $dV/dt = pV$ par produit infini d'exponentielles matricielles,* Bull. de l'Acad. R. de Belg., Classe des Sciences *44* (1958), 818–829.

3. N. Jacobson, *Abstract derivation and Lie algebras,* Trans. Amer. Math. Soc. *42* (1937), 206–224.

4. H. Neudecker, *A note on Kronecker matrix products and matrix equation systems,* SIAM J. Appl. Math. *17* (1969), 603–606.

5. R. Penrose, *On best approximate solutions of linear matrix equations,* Proc. Cambridge Philos. Soc. *52* (1956), 17–19.

6. H. B. Phillips, *Functions of matrices,* Amer. J. Math. *14* (1919), 266–278.

7. D. L. Powers, *On the differentials of certain matrix functions,* Can. J. Math. *23* (1971), 282–286.

8. R. F. Rinehart, *The derivative of a matric function,* Proc. Amer. Math. Soc. *7* (1955), 2–5.

9. ——— *The equivalence of definitions of a matric function,* Amer. Math. Monthly *62* (1955), 395–414.

10. ——— *Extension of the derivative concept for functions of matrices,* Proc. Amer. Math. Soc. *8* (1957), 329–335.

11. ——— *The differential of a primary matrix function,* Rend. Circ. Mat. Palermo *15* (1966), 209–215.

12. W. E. Roth, *On k-commutative matrices,* Trans. Amer. Math. Soc. *39* (1936), 483–495.

13. A. Z. Szeri and D. L. Powers, *Pivoted plane pad bearings: a variational solution.* Trans. ASME(F) *3* (1970), 466–471.

Clarkson College of Technology,
Potsdam, New York