THE DERIVATIVE OF A MATRIC FUNCTION

R. F. RINEHART

1. In the literature on functions of matrices the only definition of derivative which has appeared[1] is

$$\lim_{h\to 0} \frac{f(A + hI) - f(A)}{h} \tag{1}$$

where A is a square matrix over the complex field and h is a scalar complex variable. This definition is not in keeping with the spirit of the definition of derivative of a scalar function of a complex variable, in that the mode of approach to zero by the incremental matrix is severely restricted, namely via scalar matrices only.

It is natural to ask to what extent and under what circumstances a limit of the form of (1), with hI replaced by H, is independent of the mode of approach to zero of the incremental matrix H.

2. One approach to this question is to frame a more suitable definition of derivative and to investigate the conditions under which such derivative coincides with (1). We shall state such a definition in a form which does not require H to be nonsingular. On the other hand it seems necessary to require that H be commutative with A.

DEFINITION. *Let A be square matrix over the complex field C, and let $f(z)$ be a single-valued scalar function of a complex variable, "defined" for the matrix A, according to the definition of Giorgi, or any of its equivalents.*[2] *Let S denote the set of all matrices over C commutative with A, and let S_δ denote the subset of S whose matrices fulfill the condition $|h_{rs}| < \delta$, δ a positive real number. If there exists a $\delta > 0$ such that*

(a) *$f(A+H)$ is defined for $H \in S_\delta$,*

(b) *$f(A+H) - f(A)$ can be written as HQ for all H of S_δ,*

(c) *$\lim_{H\to 0} Q$ exists and is independent of the mode of approach to zero by $H \in S_\delta$*

then $\lim_{H\to 0} Q$ is defined to be the derivative of the matric function $f(Z)$ at $Z = A$.

Presented to the Society, December 28, 1954; received by the editors April 29, 1955.

[1] See, for instance, *Finite matrices*, by W. L. Ferrar, Oxford, Clarendon Press, 1951, p. 110.

[2] See *The equivalence of definitions of a matric function*, R. F. Rinehart, Amer. Math. Monthly, June, 1955.

By $H \to 0$ we mean of course that every element h_{rs} of H approaches zero. We shall denote the derivative as defined above by $f^I(A)$ in order to distinguish it from $f'(A)$, which will mean the scalar function $f'(z)$ formed for the matrix A.

One answer to the question raised in §1 is now provided by the following:

THEOREM 1. *A necessary and sufficient condition that $f^I(A)$ exist is that the scalar (single-valued) function $f(z)$ be analytic at the characteristic roots of A. Further,*

$$f^I(A) = f'(A) = \frac{1}{2\pi i}\int_{(\sigma)} \frac{f(z)dz}{(zI - A)^2}$$

where σ is a set of simple closed curves σ_i, each of which encloses just one characteristic root of A, and within and on which $f(z)$ is analytic.

SUFFICIENCY. Since $f(z)$ is analytic at the characteristic roots of A, the Cauchy integral formula holds for $f(A)$, viz.[2]

$$f(A) = \frac{1}{2\pi i}\int_{(\sigma)} \frac{f(z)dz}{zI - A}$$

where σ is a set of simple closed curves σ_i (which may be taken as circles) each of which encloses just one characteristic root of A. The integral here means the integral of each element of the integrand matrix.

Let S be as described in the definition. Let $\lambda_1, \cdots, \lambda_n$ be the characteristic roots of A. If $H \in S$, then by the well-known theorem of Frobenius the characteristic roots of H can be so ordered, $\mu_1, \cdots, \mu_n$, that the characteristic roots of $A+H$ will be $\lambda_i + \mu_i$, $i = 1, \cdots, n$. Hence if the absolute values of the μ_i are small enough, $\lambda_i + \mu_i$ can be made to lie within the circle σ_i, for each i.

Now the characteristic polynomial of H has coefficients which are homogeneous polynomials in the elements h_{rs} of H, leading coefficient being 1. Since the roots of a polynomial equation over C are continuous functions of the coefficients, except where leading coefficient vanishes, it follows that the characteristic roots of H can be constrained to be as small in absolute value as desired by making H sufficiently close to zero, i.e., by making $|h_{rs}|$ sufficiently small. If r denotes the smallest of the radii of the σ_i, we may therefore so choose a subset S_δ of S that for all H of S_δ, the characteristic roots of H are in absolute value less than r, and hence $\lambda_i + \mu_i$ will lie within σ_i for each i for every H of S_δ.

Since $f(z)$ is analytic in and on each σ_i, $f(A+H)$ is defined for each H of S_δ and the Cauchy integral formula holds for $f(A+H)$ with the same σ_i as occur for $f(A)$. Hence

$$f(A + H) - f(A) = \frac{1}{2\pi i}\int_{(\sigma)} [(zI - A - H)^{-1} - (zI - A)^{-1}]f(z)dz$$
$$= \frac{H}{2\pi i}\int_{(\sigma)} (zI - A - H)^{-1}(zI - A)^{-1}f(z)dz.$$

Consider any one of the circles σ_i. The elements of $(zI-A-H)^{-1} \cdot (zI-A)^{-1}$ are rational functions of z and the h_{rs} which have no singularities for z on σ_i and for $|h_{rs}| \leqq \delta_1 < \delta$. Hence the elements of the integrand are continuous functions of z and the h_{rs} in a closed domain of (h_{rs}, z) space, and therefore uniformly continuous on this domain. Hence the convergence of the integrand to $(zI-A)^{-2}f(z)$ as $h_{rs}\to 0$, is uniform for any element and from this it follows readily that

$$\lim_{H\to 0}\int_{(\sigma_i)} (zI - A - H)^{-1}(zI - A)^{-1}f(z)dz$$
$$= \int_{(\sigma_i)} \lim_{H\to 0} (zI - A - H)^{-1}(zI - A)^{-1}f(z)dz$$
$$= \int_{(\sigma_i)} (zI - A)^{-2}f(z)dz$$

regardless of the manner in which $H\to 0$ in S_δ. The same result holds, therefore, for the collection of the σ_i. Thus the existence of a matrix Q with the properties required by the definition is demonstrated, and $f^I(A)$ exists.

The proof has also established the second of the two asserted equivalents of $f^I(A)$, which is the matrix extension of the generalized Cauchy integral formula for $f^{(i)}(z)$ with $i=1$.

NECESSITY. Since the derivative $f^I(A)$ exists, we can make any restriction in the choice of S_δ we wish. Accordingly let S_δ' consist of all scalar matrices hI, with $|h| < \delta$. Then, following the procedure of Ferrar on the Giorgi form of definition of a matric function, it follows directly that

$$f^I(A) = f'(A)$$

and that $f(z)$ is necessarily analytic at each of the characteristic roots of A. Note that the derivative of a matrix function as defined earlier is precisely the derived function formed for the matrix.

Mathematical induction and the same line of reasoning as employed in the proof of the sufficiency condition of Theorem 1 yield the generalized Cauchy integral theorem for matrices.

THEOREM 2. *If $f(z)$ is analytic at the characteristic roots of A, then*

$$f^{(M)}(A) = f^{(m)}(A) = \frac{m!}{2\pi i}\int_{(\sigma)} \frac{f(z)dz}{(zI - A)^{m+1}}, \qquad m = 0, 1, \cdots,$$

where σ is the set of closed curves of Theorem 1.

An example of the advantage of the above more general concept of derivative is provided by the following theorem, whose proof is quite awkward without the concept of the general incremental matric.

THEOREM 3. *Let $A(t)$ be a square matrix whose elements are analytic functions of t in some open domain D of the t-plane. Let $A(t_1)A(t_2) = A(t_2)A(t_1)$ for all t_1, t_2 of D. Let $f(z)$ be analytic at the characteristic roots of $A(t)$, $t \in D$. Then $f(A(t))$ is a differentiable function of t in D, and*

$$\frac{df(A(t))}{dt} = f'(A(t))A'(t).$$

PROOF. Since $A(t+\Delta t)$ and $A(t)$ are commutative for t and $t+\Delta t$ in D, and since $f(z)$ is analytic at the characteristic roots of $A(t+\Delta t)$ and $A(t)$ we have by Theorem 1, with $H = A(t+\Delta t) - A(t)$,

$$f(A(t + \Delta t)) - f(A) = [A(t + \Delta t) - A(t)]Q \tag{2}$$

where $\lim_{t\to 0} Q = f'(A(t))$. Dividing (2) by Δt and taking the limit as $\Delta t \to 0$ yields the result.

The case of this theorem of usual interest is:

COROLLARY 1. *If $g(t)$ is analytic in a domain D of the t-plane, and if $f(z)$ is analytic at the characteristic roots $g(t)\lambda_i$ of $g(t)A$, $t \in D$, A being a constant matrix, then $f(g(t)A)$ is a differentiable function of t in D and,*

$$\frac{d}{dt} f(g(t)A) = f'(g(t)A)g'(t)A.$$

CASE INSTITUTE OF TECHNOLOGY