EXTENSION OF THE DERIVATIVE CONCEPT FOR FUNCTIONS OF MATRICES

R. F. RINEHART

1. Introduction. Let M_R^n and M_C^n denote the set of all square matrices of order n over the real and complex fields, respectively. By a function $f(Z)$ of a matrix Z of M_R^n (or M_C^n) is meant a mapping of a subset of $M_R^n(M_C^n)$ into $M_R^n(M_C^n)$. The question with which this paper is concerned is the establishment of suitable concepts of differentiability and derivative for such functions.

A meaningful and useful definition of these concepts, should of course bear some noticeable resemblance to the analogous concepts for scalar functions. In addition the derivative should preserve some of the elementary properties of the derivative for scalar functions. A modest set of such desirable properties is:

I

- (a) If $f(Z)$ is a constant, then $D_Z f(Z) = 0$,
- (b) $D_Z [f(Z) + g(Z)] = D_Z f(Z) + D_Z g(Z)$,
- (c) $D_Z [f(Z)g(Z)] = [D_Z f(Z)]g(Z) + f(Z)D_Z g(Z)$.

An additional important desired attribute, perhaps more stringent, is

II

The definitions of differentiability and derivative shall be applicable and meaningful when applied to the special functions on M_C^n arising from scalar functions of a complex variable [2]. For example, it would be desirable that the function e^Z turn out to be differentiable, according to the general definition of differentiability of functions on M_C^n .

In the fairly extensive literature on functions defined on M_R^n or M_C^n , or more generally on linear algebras with unit element over R or C , no definition of derivative has been given which satisfactorily fulfills requirements I and II. Most of the approaches to the derivative of a function on a linear algebra, have been such as to require that the algebra be commutative.¹ An exception is a paper of Ward [6] which defines derivative for the case of a noncommutative algebra. However, the derivative as there defined has the property that it

Presented to the Society, February 25, 1956 under the title *Generalization of the concept of derivative of a matrix function*; received by the editors April 26, 1956.

¹ See J. A. Ward [6] for an extensive bibliography on the subject up to 1940. See also, for example, [1].

usually lies outside the algebra if the algebra is noncommutative. This property precludes its satisfactory realization of attribute II.

Ringleb [4] and Volovel'skaya [7] succeeded in developing a function theory for noncommutative linear algebras with unit over the real or complex fields,² by pursuing a definition of the differential of a function on such an algebra suggested by Hausdorff. These definitions make the *a priori* severe requirement that the coordinates of the function have continuous first derivatives with respect to the coordinates of the argument element. Wagner [5] gave a definition of differential for the case of functions on a linear algebra, in developing a generalized power series theory of such functions, which does not demand this requirement. In each of these cases, one could suitably define the derivative as the "detached coefficient," so to speak, of the hypercomplex increment. A derivative so defined could be shown to satisfy I, but it is not evident that any of the derivatives so defined will fulfill II, although such can be shown in the case of the derivative as extracted from the work of Wagner. It should be also noted that the concept of differential is not adequate for providing the link (II), with functions on M_C^n arising from scalar functions. More specifically, if $f(Z)$ is such a function, it is in general false that $df(Z) = f'(Z)dZ$, where $f'(Z)$ is the function arising from $f'(z)$. The situation with respect to derivatives will be found to be more satisfactory.

In [3] the author framed a definition of differentiability and derivative of a function on M_C^n , based on the concept of a generalized difference quotient, and the limit thereof. That definition satisfies I and II, but suffers from the restriction that the incremental argument matrix is confined to the set of matrices of M_C^n which are commutative with the argument matrix.

The purpose of the present paper is to generalize that difference quotient concept in such a way as to remove this restriction, and to demonstrate that the resulting differentiability and derivative definitions possess the desired attributes I and II.

2. Differentiability and derivative. M_C^n and M_R^n are normed linear vector spaces [5], with, for example, Norm $\|z_{rs}\| = 1/n \max |z_{rs}|$. The concepts of limit and neighborhood are thus well defined on M_C^n and M_R^n .

Let $f(Z)$ be a function on M_C^n (or M_R^n) defined in a neighborhood $|z_{rs} - a_{rs}| < \delta$ of the matrix $A = \|a_{rs}\|$. We seek to define a difference quotient for $f(Z)$ in such a way that no restrictions of nonsingularity, or commutativity with A , of the increment matrix are demanded,

² Actually Ringleb confined his development to algebras over the real field only.

and thence define derivative. To this end we lay down the following

DEFINITION. Let $f(Z)$ on M_C^n (or M_R^n) be defined in a neighborhood of $Z=A$. Then $f(Z)$ is said to be differentiable at $Z=A$, if, for all H in a sufficiently small neighborhood N of 0,

(1) the difference $f(A+H)-f(A)$ is expressible in the form

$$f(A+H)-f(A) = \sum_{i=1}^k P_i H Q_i$$

where P_i and Q_i are in $M_C^n(M_R^n)$

$$(2) \quad \lim_{H \rightarrow 0} \sum_{i=1}^k P_i Q_i$$

exists and is independent of the manner of approach of H to 0 in N .³

If (1) and (2) are fulfilled, then $\lim_{H \rightarrow 0} \sum P_i Q_i$ is called the derivative of $f(Z)$ at $Z=A$, and is denoted by $f^I(A)$.

It is evident that $\sum_{i=1}^k P_i Q_i$ generalizes the concept of difference quotient for scalar functions. It may be remarked that if one pursues what may appear to be a more natural course, one might seek to define a right and a left difference quotient, and hence a right and a left derivative, by the assumptions $f(A+H)-f(A)=PH$ and $f(A+H)-f(A)=HQ$, respectively. For such a definition such a simple function as Z^2 is neither right- nor left-differentiable. These over-simplified generalizations are essentially of the same character as have forced earlier definitions into the consideration of commutative algebras only.

The essential feature of the scalar functional difference which the above definition generalizes is its expressibility as a homogeneous linear form in the increment H .

It may also be observed that the above definition is interpretable on any linear algebra with unit over a field with valuation. In this paper, however, we shall be concerned only with its application to M_C^n and M_R^n .

From the above definition follows,

THEOREM 2.1. The derivative $f^I(A)$, when it exists, is unique.

Suppose

$$f(A+H)-f(A) = \sum_{i=1}^k P_i H Q_i = \sum_{i=1}^m R_i H S_i$$

³ This last clause is of course redundant and is inserted only to preserve the parallel with the usual phrasing in complex function theory.

and that $f_1^I(A) = \lim_{H \rightarrow 0} \sum P_i Q_i$, and $f_2^I(A) = \lim_{H \rightarrow 0} \sum R_i S_i$ both exist independently of the manner of approach of H to 0. Choose H to be the scalar matrix hI , then $\sum P_i Q_i = \sum R_i S_i$ for all h sufficiently small. Hence

$$f_1^I(A) = \lim_{hI \rightarrow 0} \sum P_i Q_i = \lim_{hI \rightarrow 0} \sum R_i S_i = f_2^I(A).$$

3. Properties I. It is obvious that if $f(Z) = C$, a constant, then $f^I(Z) = 0$ at all Z . The remaining desired properties of I are also fulfilled:

THEOREM 3.1. *If $f(Z)$ and $g(Z)$ are differentiable functions on $M_C^n(M_R^n)$ at $Z = A$, then so is $s(Z) = f(Z) + g(Z)$, and, if one of $f(Z)$, $g(Z)$ is continuous at $Z = A$, so is $p(Z) = f(Z)g(Z)$, and*

$$\begin{aligned} s^I(A) &= f^I(A) + g^I(A), \\ p^I(A) &= f^I(A)g(A) + f(A)g^I(A). \end{aligned}$$

PROOF. $s(Z)$ and $p(Z)$ are both defined in a neighborhood of A , since $f(Z)$ and $g(Z)$ are. Hence,

$$\begin{aligned} s(A + H) - s(A) &= f(A + H) + g(A + H) - f(A) - g(A) \\ &= \sum_{i=1}^k P_i H Q_i + \sum_{i=1}^m S_i H T_i \end{aligned}$$

which is a linear homogeneous polynomial $\sum_{i=1}^{k+m} U_i H V_i$ in H , fulfilling (1) of the definition. Further, $\lim_{H \rightarrow 0} \sum_{i=1}^{k+m} U_i V_i$ exists, and equals $\lim_{H \rightarrow 0} \sum_{i=1}^k P_i Q_i + \lim_{H \rightarrow 0} \sum_{i=1}^m S_i T_i = f^I(A) + g^I(A)$, since the individual limits exist.

For the product, suppose $f(Z)$ is continuous at $Z = A$; then

$$\begin{aligned} p(A + H) - p(A) &= f(A + H)g(A + H) - f(A)g(A) \\ &= f(A + H)g(A + H) - f(A + H)g(A) + f(A + H)g(A) - f(A)g(A) \\ &= f(A + H) \sum_{i=1}^m S_i H T_i + \sum_{i=1}^k P_i H Q_i g(A) \end{aligned}$$

which is a linear homogeneous polynomial in H , fulfilling (I) of the definition.

Now under the assumption that $f(Z)$ is continuous at $Z = A$, i.e. that $\lim_{H \rightarrow 0} f(A + H) = f(A)$, it follows that

$$p^I(A) = f^I(A)g(A) + f(A)g^I(A).$$

The hypothesis that one of $f(Z)$, $g(Z)$ is continuous at $Z = A$ seems essential; in other words differentiability of a function on M_C^n (or M_R^n)

at a matrix A , does not appear to imply continuity of the function there.

4. Functions arising from scalar functions of a complex variable.

Let us now confine our attention to functions on M_C^n . As is well known, corresponding to a scalar function of a complex variable $f(z)$ there is a corresponding function $f(Z)$ on M_C^n [2]. Such a function $f(Z)$ is defined at a matrix A , if $f(z)$ is defined at the characteristic roots of A , and is analytic at those characteristic roots of multiplicity greater than one in the minimum equation of A . These functions constitute a subclass of the general functions on M_C^n .

For such functions, the derivative according to the present definition satisfies requirement II in the following way:

THEOREM 4.1. *Let $f(z)$ be a single valued scalar function of a complex variable, and let $f(Z)$ be the corresponding matrix function. Then a necessary and sufficient condition that $f^I(A)$ exist is that the scalar function $f(z)$ be analytic at each of the characteristic roots of A . Further, if $f^I(A)$ exists, then $f^I(A) = f'(A)$, where $f'(Z)$ is the matrix function corresponding to $f'(z)$.*

As in [3], the necessity of the condition follows immediately by using the Giorgi canonical form of $f(A)$ and choosing H to be a scalar matrix hI . This also demonstrates that $f^I(A)$, when it exists, is equal to $f'(A)$.

Sufficiency. Let the distinct characteristic roots of A be $\lambda_1, \lambda_2, \dots, \lambda_t$. Let each λ_i be enclosed in a circle σ_i centered at λ_i , and let the radii be sufficiently small that no two circles have a point in common. Since $f(z)$ is analytic at each λ_i , the radii may be taken sufficiently small that $f(z)$ is analytic on the boundary and in the interior of $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_t)$.

Since the characteristic roots of a matrix are continuous functions of the elements of the matrix, there exists a $\delta > 0$ such that for all H with Norm $H < \delta$, the characteristic roots of $A + H$ lie within σ . Let H be henceforth so restricted. Since $f(z)$ is analytic in and on σ , the Cauchy integral formula for $f(A + H)$ and $f(A)$ is valid [2]; hence, letting σ'_i denote the boundary of σ_i ,

$$\begin{aligned}
 f(A + H) - f(A) &= \frac{1}{2\pi i} \sum_{\alpha=1}^t \int_{(\sigma'_\alpha)} [(zI - A - H)^{-1} - (zI - A)^{-1}] f(z) dz \\
 &= \frac{1}{2\pi i} \sum_{\alpha=1}^t \int_{(\sigma'_\alpha)} (zI - A - H)^{-1} H (zI - A)^{-1} f(z) dz.
 \end{aligned}$$

4.1

The elements of $(zI - A - H)^{-1}$ and $(zI - A)^{-1}$ are rational functions of z whose denominators are, or can be taken to be, $\det(zI - A - H)$ and $\det(zI - A)$, respectively. Let $\lambda_\alpha^{(1)}, \lambda_\alpha^{(2)}, \dots, \lambda_\alpha^{(p)}$ be the distinct zeros, different from λ_α , of $\det(zI - A - H)$ within σ_α , and let their multiplicities be $m_1, \dots, m_p$. Let m_{p+1} and m be the multiplicities of λ_α as zeros of $\det(zI - A - H)$ and $\det(zI - A)$, respectively. Now the typical integral about σ_α in 4.1 is the sum of the integrals around a set of $p+1$ simple, closed, nonintersecting curves within σ_α , each enclosing just one of $\lambda_\alpha^{(1)}, \lambda_\alpha^{(2)}, \dots, \lambda_\alpha^{(p)}, \lambda_\alpha$ in its interior. By the extended Cauchy integral formula for scalar functions, and the well known rules for differentiating matrices whose elements are functions of a scalar variable,

$$\begin{aligned}
 & \frac{1}{2\pi i} \int_{\sigma'_\alpha} (zI - A - H)^{-1} H (zI - A)^{-1} f(z) dz \\
 4.2 \quad &= \sum_{h=1}^p \frac{1}{(m_h - 1)!} \frac{d^{m_h-1}}{dz^{m_h-1}} \\
 & \cdot [(z - \lambda_\alpha^{(h)})^{m_h} (zI - A - H)^{-1} H (zI - A)^{-1} f(z)]_{z=\lambda_\alpha^{(h)}} \\
 & + \frac{1}{(m' - 1)!} \frac{d^{m'-1}}{dz^{m'-1}} \\
 & \cdot [(z - \lambda_\alpha)^{m_{p+1}} (zI - A - H)^{-1} H (z - \lambda_\alpha)^m (zI - A)^{-1} f(z)]_{z=\lambda_\alpha}
 \end{aligned}$$

where $m' = m_{p+1} + m$.

From the right member of 4.2, recalling again the rules for differentiating matrices whose elements are functions of a scalar variable, it is evident that each of the integrals in 4.1 is expressible in the form

$$\sum P_d H Q_d$$

and hence requirement (1) of the definition is fulfilled.

Further, the right member of 4.2 shows that, since H is independent of z , the "coefficient" of H in $f(A + H) - f(A)$ is equal to

$$\int_{(\sigma)} (zI - A - H)^{-1} (zI - A)^{-1} f(z) dz.$$

But by precisely the same proof as given in [3], which did not depend on the commutativity of H with A , but merely on the fact that the characteristic roots of $(zI - A - H)$ are continuous functions of the elements of H , it follows that

$$\begin{aligned}
\lim_{H \rightarrow 0} \frac{1}{2\pi i} \int_{(\sigma)} (zI - A - H)^{-1} (zI - A)^{-1} f(z) dz \\
= \frac{1}{2\pi i} \int_{(\sigma)} \lim_{H \rightarrow 0} (zI - A - H)^{-1} (zI - A)^{-1} f(z) dz \\
= \frac{1}{2\pi i} \int_{(\sigma)} (zI - A)^{-2} f(z) dz.
\end{aligned}$$

Hence requirement (2) of the definition is fulfilled.

It is evident that the extended Cauchy integral theorem holds for these functions of matrices arising from scalar functions, and hence, for such functions, the existence of the first derivative at a matrix according to the definition, insures the existence of derivatives of all orders at that matrix.

REFERENCES

1. M. Frechét, *Les surfaces dérivables relativement à une règle de multiplication*, Ann. École Norm. vol. 71 (1954) pp. 29-85.
2. R. F. Rinehart, *The equivalence of definitions of a matrix function*, Amer. Math. Monthly vol. 52 (1953) pp. 395-413.
3. ———, *The derivative of a matrix function*, Proc. Amer. Math. Soc. vol. 7 (1955) pp. 2-5.
4. F. Ringleb, *Beiträge zur Funktionentheorie in hyperkomplexen Systemen*, Rend. Circ. Mat. Palermo vol. 57 (1933) pp. 311-340.
5. R. W. Wagner, *Differentials and analytic continuation in noncommutative algebras*, Duke Math. J. vol. 9 (1942) pp. 677-691.
6. J. A. Ward, *A theory of functions in linear associative algebras*, Duke Math. J. vol. 7 (1940) pp. 233-248.
7. S. N. Volovel'skaya, *Analytic functions in non-semisimple associative linear algebras*, Zapiski Naučno-Issledovatel'skogo Instituta Matematiki i Mehaniki i Har'kovskogo Matematičeskogo Obščestv (4) vol. 19 (1948) pp. 153-159.

CASE INSTITUTE OF TECHNOLOGY