PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 120, Number 4, April 1994

A SYMMETRY PROPERTY OF THE FRÉCHET DERIVATIVE

ROY MATHIAS

(Communicated by Lance W. Small)

ABSTRACT. Let A and B be $n \times n$ matrices. We show that the matrix representing the linear transformation

$$X \mapsto (AXB + BXA)^T$$

(which is from the space of $n \times n$ matrices to itself) with respect to the usual basis is symmetric and show a similar symmetry property for the Fréchet derivative of a function $f(X) = \sum_{i=0}^{\infty} a_i X^i$ defined on the space of $n \times n$ matrices.

Let M_n denote the space of complex $n \times n$ matrices. Given $X \in M_n$ we define $\mathrm{vec}(X)$ to be the vector in C^{n^2} obtained by stacking the columns of X, i.e.,

$$\mathrm{vec}(X)_{(j-1)n+i} = x_{ij}, \qquad i, j = 1, \dots, n.$$

Let $E_{ij} \in M_n$ denote the matrix with i, j entry 1 and all other entries 0. When we refer to the *matrix representation* of a linear transformation L on M_n we mean the representation with respect to the basis $\{E_{1,1}, E_{2,1}, \dots, E_{n,1}, E_{1,2}, \dots, E_{n,n}\}$. With this notation the matrix that represents L is the matrix M such that $M\,\mathrm{vec}(X) = \mathrm{vex}(L(X))$ for all $X \in M_n$. Let $A \otimes B$ denote the Kronecker (or tensor) product of A and B. Let T_n denote the $n^2 \times n^2$ permutation matrix such that $T_n \mathrm{vec}(X) = \mathrm{vec}(X^T)$ for all $X \in M_n$. Since $(X^T)^T = X$, it follows that $T_n = T_n^{-1}$. But because T_n is a permutation matrix, we must also have $T_n^{-1} = T_n^T$, so $T_n = T_n^{-1} = T_n^T$.

We will make M_n an inner product space with inner product $\langle A, B\rangle \equiv \mathrm{tr}\, AB^*$. We say that a linear transformation L on M_n is Hermitian if

$$\langle X, L(X)\rangle = \mathrm{tr}[XL(X)^*] \in \mathbb{R} \quad \forall X \in M_n. \tag{1}$$

It is easy to check that if M, the matrix representing L, is real then L is Hermitian if and only if $M = M^T$.

Let $f(z) = \sum_{i=0}^{\infty} a_i z^i$, where the series has radius of convergence R. Then for any $X \in M_n$ with spectral radius less than R the Frechet derivative of f at X applied to $Z \in M_n$ can be shown to be

$$L_f(X, Z) = \sum_{m=1}^{\infty} a_i \sum_{k=1}^{m} X^{k-1} Z X^{m-k}. \tag{2}$$

Received by the editors July 29, 1992.
1991 *Mathematics Subject Classification*. Primary 15A04, 15A24, 15A69.
Key words and phrases. Kronecker product, tensor product, Fréchet derivative, complex symmetric matrix.

A derivation of this can be found at the beginning of §2 in [2].

We will show that the matrix representing the linear transformation $Z \mapsto L_f(X, Z)^T$ is (possibly complex) symmetric, and we will show other results of this nature. This symmetry is exploited in an algorithm in [4].

First we will state some basic Kronecker product identities for $n \times n$ matrices A, B, X:

$$\text{vec}(AXB) = B^T \otimes A \, \text{vec}(X), \tag{3}$$

$$(A \otimes B)^T = A^T \otimes B^T, \tag{4}$$

$$T_n(A \otimes B) = (B \otimes A) T_n. \tag{5}$$

These are Lemma 4.3.1, equation 4.2.2, and Corollary 4.3.10 of [1] respectively. Note that in [1] the matrix T_n is denoted by $P(n, n)$.

We now combine these identities to obtain a simple result upon which our subsequent results are based.

Lemma 1. *Let $A, B \in M_n$ be given. The matrix representation of the linear transformation*

$$L(X) = (AXB + BXA)^T \tag{6}$$

is (possibly complex) symmetric. If in addition A and B are real then the linear transformation L is Hermitian.

Proof. By (3) the matrix representation of $X \mapsto (AXB + BXA)$ is $B^T \otimes A + A^T \otimes B$, so the matrix representation of L is $M \equiv T_n(B^T \otimes A + A^T \otimes B)$. Let us show that $M^T = M$:

$$\begin{aligned} M^T &= [T_n(B^T \otimes A + A^T \otimes B)]^T = (B^T \otimes A + A^T \otimes B)^T T_n^T \\ &= (B^T \otimes A)^T T_n + (A^T \otimes B)^T T_n = (B \otimes A^T) T_n + (A \otimes B^T) T_n \\ &= T_n(A^T \otimes B) + T_n(B^T \otimes A) = M. \end{aligned}$$

We have used $T_n^T = T_n$ for the third equality, (4) for the fourth, and (5) for the fifth.

If A and B are real then so is M. Since M is real and symmetric, it follows that L is Hermitian. □

One can prove that if A and B are real then the linear transformation $L(X) = (AXB + BXA)^T$ is Hermitian without using Kronecker products.

$$\begin{aligned} \langle X, (AXB + BXA)^T \rangle &= \operatorname{tr} X[(AXB + BXA)^T]^* \\ &= \operatorname{tr} X\overline{(AXB)} + \operatorname{tr} X\overline{(BXA)} = \operatorname{tr} XA\overline{X}B + \operatorname{tr} XB\overline{X}A \\ &= \operatorname{tr} XA\overline{X}B + \operatorname{tr} \overline{X}AXB = \operatorname{tr} XA\overline{X}B + \overline{\operatorname{tr} XA\overline{X}B}, \end{aligned}$$

which is real for all $X \in M_n$.

Theorem 2. *Let $f(z) = \sum_{i=0}^{\infty} a_i z^i$, where the series has radius of convergence R. Let $X \in M_n$ have spectral radius less than R. Then the matrix representation of the linear transformation*

$$Z \mapsto [L_f(X, Z)]^T \tag{7}$$

is (possibly complex) symmetric. If in addition X is real and a_i, $i = 1, 2, \ldots$, are real then the linear transformation in (7) is Hermitian.

This result is true for a more general class of functions. We will discuss this generalization after Theorem 3.

Proof. The linear transformation in (7) can be expressed as

$$Z \mapsto \sum_{m=1}^{\infty} a_i \sum_{k=1}^{m} (X^{k-1} Z X^{m-k} + X^{m-k} Z X^{k-1})^T / 2$$

which is a sum of terms of the form $Z \mapsto (AZB + BZA)^T$. The result now follows from Lemma 1. □

One can show that for the exponential function

$$L_{\exp}(X, Z) = \int_0^1 e^{tX} Z e^{(1-t)X} \, dt,$$

for example, by substituting $e^Y = \sum_0^\infty Y^k/k!$, integrating term by term, and then comparing the result with (2). One way to estimate $L_{\exp}(X, Z)$ is to approximate the integral by the composite trapezoidal rule

$$L_{\exp, T} \equiv \frac{1}{2^{k+1}} \left\{ Z e^X + 2 \sum_{j-1}^{2^k - 1} e^{jX/2^k} Z e^{(2^k - j)X/2^k} + e^X Z \right\}$$

or the composite Simpson's rule

$$\begin{aligned} L_{\exp, S} \equiv \frac{1}{6 \cdot 2^k} \Bigg\{ & Z e^X + 2 \sum_{j=1}^{2^k - 1} e^{2jX/2^k} Z e^{(2^k - 2j)X/2^k} \\ & + 4 \sum_{j=1}^{2^k - 1} e^{(2j-1)X/2^k} Z e^{(2^k - 2j + 1)X/2^k} + e^X Z \Bigg\}. \end{aligned}$$

It is useful to know that these approximations have the same symmetry property as $L_{\exp}(X, Z)$.

Theorem 3. *Let $X \in M_n$. The matrix representations of the linear transformations*

$$Z \mapsto [L_{\exp, T}(X, Z)]^T \quad \textit{and} \quad Z \mapsto [L_{\exp, S}(X, Z)]^T \tag{8}$$

are (possibly) symmetric. If X is real then the linear transformations in (8) are Hermitian.

Proof. The result for $L_{\exp, T}$ follows from the formula

$$\begin{aligned} L_{\exp, T}(X, Z) = \frac{1}{2^{k+1}} \{ & (Z e^X + e^X Z)/2 \\ & + 2 \sum_{j=1}^{2^k - 1} (e^{jX/2^k} Z e^{(2^k - j)X/2^k} + e^{(2^k - j)X/2^k} Z e^{jX/2^k})/2 \\ & + (e^X Z + Z e^X)/2 \} \end{aligned}$$

and Lemma 1. The result for $L_{\exp, S}$ follows similarly. □

Let D be a domain in the complex plane, and let f be analytic on D. The *primary matrix function* associated with f is defined on the set of matrices with spectrum contained in D as follows:

(a) if $A = S \operatorname{diag}(\lambda_1, \ldots, \lambda_n) S^{-1}$ then

$$f(A) = S \operatorname{diag}(f(\lambda_1), \ldots, f(\lambda_n)) S^{-1},$$

(b) if A is not diagonalizable then define $f(A)$ by continuity.

The analyticity of f ensures that $f(A)$ is well defined. One can also define $f(A)$ via the Jordan form; this definition gives an explicit form even when A is not diagonalizable. See [1, §6.6] for this and a further discussion of primary matrix functions.

One can show that a primary matrix function is Fréchet differentiable at any X with spectrum contained in D, so the Fréchet derivative is equal to the directional derivative. So by [1, Theorem 6.6.14(3)]

$$L_f(X, Z) = \frac{d}{dt} f(X + tZ)|_{t=0} = \frac{d}{dt} p(X + tZ)|_{t=0}$$

where p is any polynomial such that if λ is an eigenvalue of $X \oplus X$ of algebraic multiplicity m then $p^{(i)}(\lambda) = f^{(i)}(\lambda)$, $i = 0, 1, \ldots, m-1$. (Note that the restriction in [1, Theorem 6.6.14(3)] that D be simply connected is not necessary.) Thus $L_f(X, Z) = L_p(X, Z)$, and since p is just a polynomial it follows from Theorem 2 that the matrix representation of

$$Z \mapsto [L_f(X, Z)]^T = [L_p(X, Z)]^T$$

is (possibly complex) symmetric.

A primary matrix function that cannot be represented as a power series and for which one wants to compute the Fréchet derivative is the matrix sign function $\operatorname{sgn}(A)$ [3]. It corresponds to the function $f(z) = \operatorname{sign}(\operatorname{Re}(z))$ on $D = \{z\colon \operatorname{Re}(z) \neq 0\}$. An equivalent definition that is often used is

$$\operatorname{sgn}(A) = S \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} S^{-1},$$

where $A = S\left(\begin{smallmatrix} P & 0 \\ 0 & N \end{smallmatrix}\right)S^{-1}$ and P and $-N$ have spectrum in the open right-half plane.

Acknowledgment

Professor R. Horn pointed out that Theorem 2 is valid for primary matrix functions.

References

1. R. A. Horn and C. R. Johnson, *Topics in matrix analysis*, Cambridge Univ. Press, New York, 1991.
2. C. Kenney and A. J. Laub, *Conditions estimates for matrix functions*, SIAM J. Matrix Anal. Appl. **10** (1989), 191–209.
3. ——, *Polar decomposition and matrix sign function condition estimates*, SIAM J. Sci. Statist. Comput. **12** (1991), 488–504.
4. R. Mathias, *Evaluating the Fréchet derivative of the matrix exponential*, Numer. Math. **63** (1992), 213–226.

Department of Mathematics, College of William and Mary, Williamsburg, Virginia 23187

E-mail address: na.mathias@na-net.ornl.gov