Operators
and
Matrices
Volume 7, Number 1 (2013), 71–90
doi:10.7153/oam-07-03

DIFFERENTIATING MATRIX FUNCTIONS

KELLY BICKEL

(*Communicated by K. Veselić*)

Abstract. Real-valued functions on $\mathbb{R}^d$ induce matrix-valued functions defined on the space of d-tuples of $n \times n$ pairwise-commuting self-adjoint matrices. We examine the geometry of this space of matrices and conclude that a suitable notation of differentiation of these matrix functions is differentiation along curves. We prove that continuously differentiable real-valued functions induce continuously differentiable matrix functions and give a formula for the derivative. We also show that real-valued m-times continuously differentiable functions defined on open rectangles in $\mathbb{R}^2$ induce matrix functions that can be m-times continuously differentiated along m-times continuously differentiable curves.

1. Introduction

Every real-valued function defined on $\mathbb{R}$ induces a matrix-valued function on the space of $n \times n$ self-adjoint matrices by acting on the spectrum of each matrix. Likewise, each real-valued function f defined on an open set $\Omega \subseteq \mathbb{R}^d$ induces a matrix-valued function F on the space of d-tuples of $n \times n$ pairwise-commuting self-adjoint matrices with joint spectrum in Ω. Let $S = (S^1, \dots, S^d)$ be such a d-tuple and let U be a unitary matrix diagonalizing S as follows:

$$S^r = U \begin{pmatrix} x_1^r & & \\ & \ddots & \\ & & x_n^r \end{pmatrix} U^*,$$

for $1 \leqslant r \leqslant d$. Denote the joint spectrum of S by $\sigma(S) := \left\{ x_i = (x_i^1, \dots, x_i^d) : 1 \leqslant i \leqslant n \right\}$ and define

$$F(S) := U \begin{pmatrix} f(x_1) & & \\ & \ddots & \\ & & f(x_n) \end{pmatrix} U^*, \tag{1}$$

where $F(S)$ is independent of the choice of U.

This paper will show that certain differentiability properties of the original function pass to the matrix function. Even for a one-variable function, this is nontrivial. Let $f \in C^1(\mathbb{R}, \mathbb{R})$ and consider the simple case of differentiating the associated matrix

Mathematics subject classification (2010): 26B05; 15A60.
Keywords and phrases: Matrix functions; Differentiability; Derivative formulas.
This research was partially supported by National Science Foundation grant DMS-0966845.

function F along a C^1 curve $S(t)$ of $n\times n$ self-adjoint matrices. At first glance, it seems reasonable to write $S(t)=U(t)D(t)U^*(t)$, for $U(t)$ unitary and $D(t)$ diagonal. Then $F(S(t))=U(t)F(D(t))U^*(t)$ and we can differentiate using the product rule.

However, there is no guarantee that we can decompose $S(t)$ into its eigenvector and eigenvalue matrices so that the eigenvectors are even continuous. As demonstrated by the following example from [9], eigenvector behavior at points where distinct eigenvalues coalesce can be unpredictable. Specifically, let

$$S(t)=e^{-\frac{1}{t^2}}\begin{pmatrix}\cos\left(\frac{2}{t}\right) & \sin\left(\frac{2}{t}\right)\\ \sin\left(\frac{2}{t}\right) & -\cos\left(\frac{2}{t}\right)\end{pmatrix} \text{ for t } \neq 0, \text{ and } S(0)=0.$$

For $t\neq 0$, the eigenvalues of $S(t)$ are $\pm\, e^{-\frac{1}{t^2}}$ and their associated eigenvectors are

$$\pm\begin{pmatrix}\cos\left(\frac{1}{t}\right)\\ \sin(\frac{1}{t})\end{pmatrix} \text{ and } \pm\begin{pmatrix}\sin(\frac{1}{t})\\ -\cos(\frac{1}{t})\end{pmatrix}.$$

Thus, even an infinitely differentiable curve can have singularities in its eigenvectors.

The differentiability of matrix functions defined from one-variable functions is discussed frequently in the literature (see [2], [4], [6]). The most comprehensive result is by Brown and Vasudeva in [3], who prove that an m-times continuously differentiable real-valued function induces an m-times continuously Fréchet differentiable matrix-valued function.

If a matrix-valued function is defined using a real-valued function on $\mathbb{R}^d$ as in (1), its domain is the space of d-tuples of pairwise-commuting $n\times n$ self-adjoint matrices, denoted CS_n^d. For $d>1$, the space of d-tuples of $n\times n$ self-adjoint matrices is denoted S_n^d and for $d=1$, is denoted S_n.

It should be noted that there is an alternate approach for inducing a matrix function from a multivariate function; the d matrices $S^1,\dots,S^d$ are viewed as operators on Hilbert spaces $H^1,\dots,H^d$ and $F(S)$ is viewed as an operator on $H^1\otimes\dots\otimes H^d$. Brown and Vasudeva generalize their one-variable result to these matrix functions in [3].

In this paper, we focus on matrix functions defined as in (1). Specifically, in Section 2, we analyze the geometry of CS_n^d and conclude that a suitable notion of differentiability for functions on this space is differentiation along curves. If we fix S in CS_n^d, Theorem 1 characterizes the directions Δ in S_n^d such that there is a C^1 curve $S(t)$ in CS_n^d with $S(0)=S$ and $S'(0)=\Delta$. In Theorem 2, we show that the joint eigenvalues of locally Lipschitz curves in CS_n^d can be represented by locally Lipschitz functions.

In Section 3, we examine the differentiability properties of induced matrix functions. Specifically, in Theorem 3, we show that a C^1 function induces a matrix function that can be continuously differentiated along C^1 curves. We then calculate a formula for the derivative along curves and in Theorem 4, prove that it is continuous.

In Section 4, we consider higher-order differentiation. With additional domain restrictions, in Theorem 6, we show that C^m functions induce matrix functions that can be m-times continuously differentiated along C^m curves. We also calculate a formula for the derivatives and in Theorem 7, show they are continuous. In Section 5, we discuss several applications of the differentiability results.

Before proceeding, I would like to thank John McCarthy for his guidance during this research and the referees for their many useful suggestions.

2. The Geometry of CS_n^d

Let $S = (S^1, \dots, S^d)$ be in CS_n^d (or S_n^d) and let $x_i = (x_i^1, \dots, x_i^d)$ be in $\sigma(S)$. Define

$$\|S\| := \max_{1 \leqslant r \leqslant d} \|S^r\| \text{ and } \|x_i\| := \max_{1 \leqslant r \leqslant d} |x_i^r|, \tag{2}$$

where $\|S^r\|$ is the usual operator norm. As each $S \in S_n$ is determined by its upper triangular part, which has n^2 degrees of freedom, S_n can be equated with $\mathbb{R}^{n^2}$. Then, CS_n^d can be viewed as a subset of $\mathbb{R}^m$, where $m = dn^2$. It follows from basic facts about self-adjoint matrices that the norm on CS_n^d inherited from Euclidean space and the one defined in (2) are equivalent norms. Now, observe that CS_n^d is not a linear space; if A and B are pairwise-commuting d-tuples, the sum $A+B$ need not pairwise commute. Thus, neither the Fréchet nor Gâteaux derivatives can be defined for functions on CS_n^d because both require the function to be defined on linear sets around each point.

Recall that CS_n^d is the set of elements $S \in S_n^d$ with $[S^r, S^s] = 0$ for all $1 \leqslant r, s \leqslant d$, where $[\cdot,\cdot]$ denotes Lie bracket. Thus, CS_n^d is the zero set of the polynomials associated with $d(d-1)/2$ commutator operations and so is a real algebraic variety. A result by Whitney in [11] and discussed by Kaloshin in [7] says every algebraic variety defined by polynomials on m real variables can be decomposed into smooth submanifolds of $\mathbb{R}^m$ that fit together 'regularly' and whose tangent spaces fit together 'regularly.' For a manifold N, let TN denote the tangent space of N and let T_SN denote the tangent space based at a point S in N. For a closed subset X of $\mathbb{R}^m$, we can define

DEFINITION 1. A *stratification* of X is a locally finite partition Z of X into locally closed pieces $\{M_\alpha\}$ such that

- (i) Each piece $M_\alpha \in Z$ is a smooth submanifold of $\mathbb{R}^m$.
- (ii) (*Condition of frontier*) If $M_\alpha \cap \bar{M}_\beta \neq \emptyset$ for pieces M_α, M_β, then $M_\alpha \subset \bar{M}_\beta$.

EXAMPLE 1. Consider CS_2^2, the space of pairs of self-adjoint, commuting 2×2 matrices. In the following definitions, $a,b,c,d \in \mathbb{R}$. Define

$$M_1 := \left\{ \left(U \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} U^*, U \begin{pmatrix} c & 0 \\ 0 & d \end{pmatrix} U^* \right) : U \text{ is } 2\times 2 \text{ unitary}, a \neq b, c \neq d \right\},$$

$$M_2 := \left\{ \left(\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}, U \begin{pmatrix} c & 0 \\ 0 & d \end{pmatrix} U^* \right) : U \text{ is } 2\times 2 \text{ unitary}, c \neq d \right\},$$

$$M_3 := \left\{ \left(U \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} U^*, \begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix} \right) : U \text{ is } 2\times 2 \text{ unitary}, a \neq b \right\},$$

$$M_4 := \left\{ \left(\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}, \begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix} \right) \right\}.$$

It is easy to see that $CS_2^2 = \cup M_i$ and each M_i is locally closed. With a little work, one can show each M_i is a smooth submanifold of $\mathbb{R}^8$. As this example clearly satisfies the condition of frontier, this partition $\{M_i\}$ is a stratification of CS_2^2.

In general, one should expect a stratification of CS_n^d into pieces to be related to the number and multiplicity of the repeated eigenvalues of the elements of CS_n^d.

Whitney's result says CS_n^d has a specific decomposition Z into smooth submanifolds of $\mathbb{R}^m$ where $m = dn^2$, called a Whitney stratification. This stratification has further regularity involving the tangent spaces of the pieces of Z, but as we do not need those details here, see [7] for the specifics. We let $\{M_\alpha\}$ denote the pieces of Z and define $TCS_n^d := \cup TM_\alpha$. Given a function $F : CS_n^d \to S_n$, one type of derivative is a map $DF : TCS_n^d \to TS_n$ such that

$$DF|_{TM_\alpha} : TM_\alpha \to TS_n$$

is the usual differential map for each M_α. In Theorem 5, we analyze such maps. However, these differential maps cannot be easily generalized to analyze higher-order differentiation. Furthermore, for each $S \in CS_n^d$ and piece M_α containing S, the tangent space $T_S M_\alpha$ might only contain a subset of the vectors tangent to CS_n^d at S. Example 2 will show that strict containment often occurs.

To retain information about all tangent vectors, we will mostly study differentiation along differentiable curves. We first determine which $\Delta \in S_n^d$ are vectors tangent to CS_n^d at a given point S. For any $\Delta \in S_n^d$ and $S \in CS_n^d$, we ask

Is there a C^1 curve $S(t)$ in CS_n^d with $S(0) = S$ and $S'(0) = \Delta$?

For an element $S \in CS_n^d$ with distinct joint eigenvalues, Agler, McCarthy, and Young in [1] gave necessary and sufficient conditions on S and Δ for such a C^1 curve to exist. We extend their result to an arbitrary element S. Fix $S \in CS_n^d$ and $\Delta \in S_n^d$. Let U be a unitary matrix diagonalizing each component of S such that the repeated joint eigenvalues of S appear consecutively. Numbering the x_i's appropriately, define

$$D^r := U^* S^r U = \begin{pmatrix} x_1^r & & \\ & \ddots & \\ & & x_n^r \end{pmatrix}, \tag{3}$$

for each $1 \leqslant r \leqslant d$. Then, for each r, define the two matrices

$$\Gamma^r := U^* \Delta^r U$$
$$\tilde{\Gamma}^r_{ij} := \begin{cases} \Gamma^r_{ij} & \text{if } x_i = x_j \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Then $\tilde{\Gamma}^r$ is a block diagonal matrix. Each block corresponds to a distinct joint eigenvalue of S and has dimension equal to the multiplicity of that eigenvalue.

THEOREM 1. *Let $S \in CS_n^d$ and $\Delta \in S_n^d$. Then there exists a C^1 curve $S(t)$ in CS_n^d with $S(0) = S$ and $S'(0) = \Delta$ if and only if for all $1 \leqslant s, r \leqslant d$,*

$$[D^r, \Gamma^s] = [D^s, \Gamma^r] \text{ and } \left[\tilde{\Gamma}^r, \tilde{\Gamma}^s\right] = 0.$$

Proof. ($\Rightarrow$) Assume $S(t)$ is a C^1 curve in CS_n^d with $S(0)=S$ and $S'(0)=\Delta$. Define

$$R(t) := U^* S(t) U,$$

where U diagonalizes S as in (3). Then $R(t)$ is a C^1 curve in CS_n^d with $R(0)=D$ and $R'(0)=\Gamma$. We will first prove that

$$[D^r,\Gamma^s]=[D^s,\Gamma^r] \text{ and } [\Gamma^r,\Gamma^s]_{ij}=0,$$

for all pairs $1\leqslant r,s\leqslant d$ and (i,j) such that $x_i=x_j$. We will use those commutativity results to conclude

$$\left[\tilde{\Gamma}^r,\tilde{\Gamma}^s\right]=0,$$

for each pair $1\leqslant r,s\leqslant d$. Since $R(t)$ is C^1 in a neighborhood of $t=0$, we can write

$$R^r(t)=D^r+\Gamma^r t+h^r(t),$$

for each $1\leqslant r\leqslant d$, where $|h^r(t)_{ij}|=o(|t|)$ for $1\leqslant i,j\leqslant n$. For each pair r and s, the pairwise-commutativity of $R(t)$ implies

$$\begin{aligned}
0 &= [R^r(t),R^s(t)]\\
&= [D^r+\Gamma^r t+h^r(t),D^s+\Gamma^s t+h^s(t)]\\
&= \big([D^r,h^s(t)]+[h^r(t),D^s]+[h^r(t),h^s(t)]\big)\\
&\quad+\big([D^r,\Gamma^s]+[\Gamma^r,D^s]+[\Gamma^r,h^s(t)]+[h^r(t),\Gamma^s]\big)t\\
&\quad+[\Gamma^r,\Gamma^s]t^2,
\end{aligned}\tag{5}$$

where the term $[D^r,D^s]$ was omitted because it vanishes. Fix $t\neq 0$ and divide each term in (5) by t. Letting t tend towards zero yields

$$0=[D^r,\Gamma^s]-[D^s,\Gamma^r].\tag{6}$$

Choose i and j such that $x_i=x_j$. Then, the ij^{th} entry of (5) reduces to

$$0=[h^r(t),h^s(t)]_{ij}+\big(\,[\Gamma^r,h^s(t)]_{ij}-[\Gamma^s,h^r(t)]_{ij}\,\big)t+[\Gamma^r,\Gamma^s]_{ij}t^2.$$

Fix $t\neq 0$ and divide both sides by t^2. Letting t tend towards zero yields

$$0=[\Gamma^r,\Gamma^s]_{ij}.\tag{7}$$

Fix r and s with $1\leqslant r,s\leqslant d$. Since $\tilde{\Gamma}^r$ and $\tilde{\Gamma}^s$ are block diagonal matrices with blocks corresponding to the distinct joint eigenvalues of S, it follows that $\tilde{\Gamma}^r\tilde{\Gamma}^s$ and $\tilde{\Gamma}^s\tilde{\Gamma}^r$ are also such block diagonal matrices. Thus, if i and j are such that $x_i\neq x_j$,

$$\left[\,\tilde{\Gamma}^r,\,\tilde{\Gamma}^s\right]_{ij}=\left(\,\tilde{\Gamma}^r\tilde{\Gamma}^s-\tilde{\Gamma}^s\tilde{\Gamma}^r\,\right)_{ij}=0.$$

Now, fix i and j such that $x_i = x_j$. By the definition of $\tilde{\Gamma}$,

$$\begin{aligned}\left[\tilde{\Gamma}^r, \tilde{\Gamma}^s\right]_{ij} &= \sum_{k=1}^{n} \tilde{\Gamma}^r_{ik}\tilde{\Gamma}^s_{kj} - \tilde{\Gamma}^s_{ik}\tilde{\Gamma}^r_{kj} \\ &= \sum_{\{k: x_k = x_i\}} \Gamma^r_{ik}\Gamma^s_{kj} - \Gamma^s_{ik}\Gamma^r_{kj} \\ &= -\sum_{\{k: x_k \neq x_i\}} \Gamma^r_{ik}\Gamma^s_{kj} - \Gamma^s_{ik}\Gamma^r_{kj},\end{aligned}$$

where the last equality uses (7). Thus, it suffices to show that if $x_k \neq x_i$,

$$\Gamma^r_{ik}\Gamma^s_{kj} - \Gamma^s_{ik}\Gamma^r_{kj} = 0.$$

Assume $x_k \neq x_i$, and fix q with $x^q_k \neq x^q_i$. Apply (6) to pairs r, q and s, q to get

$$[D^q, \Gamma^r] = [D^r, \Gamma^q] \quad \text{and} \quad [D^q, \Gamma^s] = [D^s, \Gamma^q].$$

Restricting to the ik^{th} and kj^{th} entries of the previous two equations yields

$$\begin{aligned}\Gamma^r_{ik}(x^q_i - x^q_k) &= \Gamma^q_{ik}(x^r_i - x^r_k), \\ \Gamma^r_{kj}(x^q_k - x^q_j) &= \Gamma^q_{kj}(x^r_k - x^r_j), \\ \Gamma^s_{ik}(x^q_i - x^q_k) &= \Gamma^q_{ik}(x^s_i - x^s_k), \\ \Gamma^s_{kj}(x^q_k - x^q_j) &= \Gamma^q_{kj}(x^s_k - x^s_j).\end{aligned}$$

Since $x_i = x_j$ and $x^q_k \neq x^q_i$, we can replace all the x_j's with x_i's in the above set of equations and solve for the Γ^r and Γ^s entries. Using these relations gives

$$\Gamma^r_{ik}\Gamma^s_{kj} - \Gamma^s_{ik}\Gamma^r_{kj} = \frac{\Gamma^q_{ik}(x^r_i - x^r_k)\Gamma^q_{kj}(x^s_i - x^s_k)}{(x^q_i - x^q_k)^2} - \frac{\Gamma^q_{ik}(x^s_i - x^s_k)\Gamma^q_{kj}(x^r_i - x^r_k)}{(x^q_i - x^q_k)^2} = 0,$$

as desired. Thus, $[\tilde{\Gamma}^r, \tilde{\Gamma}^s] = 0$.

($\Leftarrow$) Fix S in CS^d_n and Δ in S^d_n and let U, D, Γ, and $\tilde{\Gamma}$ be as in the discussion preceding Theorem 1. Assume

$$[D^r, \Gamma^s] = [D^s, \Gamma^r] \text{ and } \left[\tilde{\Gamma}^r, \tilde{\Gamma}^s\right] = 0, \tag{8}$$

for $1 \leqslant r, s \leqslant d$. Define a skew-Hermitian matrix Y as follows:

$$Y_{ij} := \begin{cases} \frac{\Gamma^q_{ij}}{x^q_j - x^q_i} & \text{if } x_i \neq x_j \\ 0 & \text{otherwise,} \end{cases}$$

where q is chosen so that $x^q_i - x^q_j \neq 0$. Observe that Y is independent of q because the ij^{th} entry of the first equation in (8) is

$$\Gamma^s_{ij}(x^r_i - x^r_j) = \Gamma^r_{ij}(x^s_i - x^s_j).$$

Now, define the curve $S(t)$ by

$$S^r(t) := Ue^{Yt}\left[D^r + t\tilde{\Gamma}^r\right]e^{-Yt}U^*,$$

for each $1 \leqslant r \leqslant d$. Then, $S(t)$ is continuously differentiable. Because Y is skew-Hermitian, e^{Yt} is unitary. Since D^r and $\tilde{\Gamma}^r$ are self-adjoint, $S(t)$ is in S_n^d. By a simple calculation using (8),

$$[S^r(t), S^s(t)] = 0,$$

for each pair $1 \leqslant r, s \leqslant d$. Thus, $S(t)$ is in CS_n^d. By definition, $S(0) = S$. For each r,

$$(S^r)'(t) = U\left(Ye^{Yt}\left[D^r + t\tilde{\Gamma}^r\right]e^{-Yt} + e^{Yt}\left[\tilde{\Gamma}^r\right]e^{-Yt} - e^{Yt}\left[D^r + t\tilde{\Gamma}^r\right]Ye^{-Yt}\right)U^*,$$

so that

$$(S^r)'(0) = U\left([Y, D^r] + \tilde{\Gamma}^r\right)U^* = \Delta^r.$$

Thus, $S'(0) = \Delta$, and $S(t)$ is the desired curve. □

Observe that by the construction in Theorem 1, if there is a C^1 curve $S(t)$ in CS_n^d with $S(0) = S$ and $S'(0) = \Delta$, there is actually a smooth curve $R(t)$ in CS_n^d with $R(0) = S$ and $R'(0) = \Delta$.

EXAMPLE 2. Let $I \in CS_n^d$ be the identity element. By Theorem 1, there is a smooth curve $S(t)$ in CS_n^d with

$$S(0) = I \text{ and } S'(0) = \Delta \text{ if and only if } \Delta \in CS_n^d.$$

Thus, the set of vectors tangent to CS_n^d at I is CS_n^d. For a Whitney stratification of CS_n^d and piece M_α containing I, the tangent space $T_I M_\alpha$ is linear. Since CS_n^d is not linear, $T_I M_\alpha$ is a strict subset of the set of tangent vectors at I.

The conditions of Theorem 1 actually imply that if $S \in CS_n^d$ has any repeated joint eigenvalues, the set of vectors tangent to CS_n^d at S is not a linear set. Then, for any Whitney stratification of CS_n^d and piece M_α containing S, the tangent space $T_S M_\alpha$ is a strict subset of the vectors tangent to CS_n^d at S. We will thus focus on differentiation along curves rather than differential maps.

To evaluate an induced matrix function along a curve in CS_n^d, we apply the original function to the curve's joint eigenvalues. We are therefore interested in the behavior of the joint eigenvalues of curves in CS_n^d.

If $S(t)$ is a continuous curve in S_n, a result by Rellich in [9] and [10] states that the eigenvalues of $S(t)$ can be represented by n continuous functions. A succinct proof is given by Kato in [8, pg 107-10]. With slight modification, the arguments show that the eigenvalues of a locally Lipschitz curve in S_n can be represented by locally Lipschitz functions. These results generalize as follows:

THEOREM 2. *Given a locally Lipschitz curve $S(t)$ in CS_n^d defined on an interval I, there exist locally Lipschitz functions $x_1(t), \ldots, x_n(t) : I \to \mathbb{R}^d$ with $\sigma(S(t)) = \{x_i(t) : 1 \leqslant i \leqslant n\}$.*

Proof. As the proof is a technical but straightforward modification of the one-variable case, it is left as an exercise. □

Theorem 2 provides a specific ordering of the joint eigenvalues of $S(t)$ at each t. This ordering may differ from the one in (3), where joint eigenvalues appear consecutively. However, Theorem 2 implies that the joint eigenvalues of a continuously differentiable, and hence locally Lipschitz, curve $S(t)$ are locally Lipschitz as an unordered n-tuple. Specifically, fix t^* and denote the eigenvalues of $S(t^*)$ by $\{x_i : 1 \leqslant i \leqslant n\}$. Then, for t near t^*, there is a constant c such that

$$\min\left(\max_{1\leqslant i\leqslant n} \|x_i - x_i(t)\|\right) \leqslant c|t^* - t|,$$

where the minimum is taken over all reorderings of the $\{x_i\}$. If we require that eigenvalues are ordered as in (3), we will use Theorem 2 to conclude that the eigenvalues are locally Lipschitz as an unordered n-tuple.

3. Differentiating Matrix Functions

Recall that every real-valued function defined on an open set $\Omega \subseteq \mathbb{R}^d$ induces a matrix function as in (1). We denote its domain, the space of d-tuples of pairwise-commuting $n\times n$ self-adjoint matrices with spectrum in Ω, by $CS_n^d(\Omega)$.

If the original function is continuous, the matrix function is as well. Specifically, Horn and Johnson proved in [6, pg 387-9] that a one-variable polynomial induces a continuous matrix polynomial. The arguments generalize easily to multivariate polynomials, and approximation arguments imply that the matrix function induced by a continuous function is continuous. We now consider differentiability and prove:

THEOREM 3. *Let $S(t)$ be a C^1 curve in CS_n^d defined on an interval I, and let Ω be an open set in $\mathbb{R}^d$ with $\sigma(S(t)) \subset \Omega$. If $f \in C^1(\Omega,\mathbb{R})$, then*

(i) $\frac{d}{dt}F(S(t))|_{t=t^*}$ *exists for all $t^* \in I$.*

(ii) *If $T(t)$ is another C^1 curve in CS_n^d with $T(0) = S(t^*)$ and $T'(0) = S'(t^*)$, then*

$$\tfrac{d}{dt}F(T(t))|_{t=0} = \tfrac{d}{dt}F(S(t))|_{t=t^*}.$$

Before proving Theorem 3, we assume f is real-analytic and prove Proposition 1. See [6] for the one-variable case. We first need some notation. We say an open set $\Omega \subseteq \mathbb{R}^d$ is a *rectangle* if $\Omega = I^1 \times \ldots \times I^d$ or more specifically,

$$\Omega = \{(x_1,\ldots,x_d) : x_r \in I^r \ \forall\ 1 \leqslant r \leqslant d\},$$

where each I^r is an open interval in $\mathbb{R}$, and an open set $\tilde{\Omega} \subseteq \mathbb{C}^d$ is a *complex rectangle* if $\tilde{\Omega} = (I^1 + iJ^1) \times \ldots \times (I^d + iJ^d)$ or specifically,

$$\tilde{\Omega} = \{(x_1 + iy_1,\ldots,x_d + iy_d) : x_r \in I_r, y_r \in J_r \ \forall\ 1 \leqslant r \leqslant d\},$$

where for each r, I^r and J^r are open intervals in $\mathbb{R}$.

PROPOSITION 1. *Let $S(t)$ be a C^1 curve in CS_n^d defined on an interval I. Let Ω be a rectangle in $\mathbb{R}^d$ with $\sigma(S(t)) \subset \Omega$. If f is a real-analytic function on Ω, then*

$$\tfrac{d}{dt}F(S(t))|_{t=t^*} \text{ exists and is continuous as a function of } t^* \text{ on } I.$$

The proof of Proposition 1 requires the following two lemmas.

LEMMA 1. *Let Ω be a rectangle in $\mathbb{R}^d$ and let S be in CS_n^d with $\sigma(S) \subset \Omega$. Each real-analytic function on Ω can be extended to an analytic function defined on a complex rectangle $\tilde{\Omega}$ such that $\sigma(S)$ is in $\tilde{\Omega}$.*

Proof. The result follows from basic properties of complex functions. It should be noted that $\tilde{\Omega}$ need not contain Ω. □

LEMMA 2. *Let $\tilde{\Omega}$ be a complex rectangle in $\mathbb{C}^d$ and let S be in CS_n^d with $\sigma(S) \subset \tilde{\Omega}$. If f is an analytic function on $\tilde{\Omega}$, then*

$$F(S) = \frac{1}{(2\pi i)^d} \int_{C^d} \cdots \int_{C^1} f(\zeta^1, ..., \zeta^d)(\zeta^1 I - S^1)^{-1} ...(\zeta^d I - S^d)^{-1}\, d\zeta^1 ... d\zeta^d,$$

where each C^r is a simple closed rectifiable curve strictly containing $\sigma(S^r)$, and $C^1 \times \ldots \times C^d \subset \tilde{\Omega}$.

Proof. Horn and Johnson prove the formula for a one-variable function in [6, pg 427]. Their derivation generalizes easily to multivariate functions. □

Proof. Proposition 1:
For ease of notation, assume $d = 2$ and for $r = 1, 2$, define

$$R^r(t) := (\zeta^r I - S^r(t))^{-1},$$

where ζ^r is in the resolvent set of $S^r(t)$. Fix $t_0 \in I$ and extend f to an analytic function on a complex rectangle $\tilde{\Omega}$ containing $\sigma(S(t_0))$. Choose simple closed rectifiable curves C^1 and C^2 such that $C^1 \times C^2 \subset \tilde{\Omega}$ and C^r strictly encloses the eigenvalues of $S^r(t_0)$. As the joint eigenvalues of $S(t)$ are continuous, we can use Lemma 2 to write

$$F(S(t)) = \frac{1}{(2\pi i)^2} \int_{C^2} \int_{C^1} f(\zeta^1, \zeta^2)\, R^1(t)\, R^2(t)\, d\zeta^1 d\zeta^2,$$

for t sufficiently close to t_0. For t_1, t_2 near t_0, we have

$$R^r(t_1) - R^r(t_2) = R^r(t_1)\,(S^r(t_1) - S^r(t_2))\,R^r(t_2),$$

which implies $R^r(t)$ is differentiable near t_0 and direct calculation gives

$$\tfrac{d}{dt}R^r(t)|_{t=t^*} = R^r(t^*)(S^r)'(t^*)R^r(t^*),$$

for $r=1,2$ and t^* near t_0. It can be easily shown that, for t^* sufficiently close to t_0, we can interchange integration and differentiation to yield

$$\begin{aligned}\tfrac{d}{dt}F(S(t))|_{t=t^*} &= \frac{1}{(2\pi i)^2}\int_{C_2}\int_{C_1} f(\zeta^1,\zeta^2)\tfrac{d}{dt}\left(R^1(t)R^2(t)\right)|_{t=t^*}\,d\zeta^1 d\zeta^2 \\ &= \frac{1}{(2\pi i)^2}\int_{C_2}\int_{C_1} f(\zeta^1,\zeta^2)\Big(R^1(t^*)(S^1)'(t^*)R^1(t^*)R^2(t^*) \\ &\quad + R^1(t^*)R^2(t^*)(S^2)'(t^*)R^2(t^*)\Big)\,d\zeta^1 d\zeta^2. \end{aligned} \tag{9}$$

As each $(S^r)'(t)$ is continuous in t and each $R^r(t)$ is continuous in t near t_0 (uniformly in ζ for ζ in $C^1\times C^2$), and $f(\zeta^1,\zeta^2)$ is uniformly bounded, $\frac{d}{dt}F(S(t))|_{t=t^*}$ is continuous at $t^*=t_0$. □

Proof. Theorem 3:
Observe that the theorem holds for polynomials: (i) follows from Proposition 1, and (ii) follows from the formula in (9). Fix $t^*\in I$. Let f be an arbitrary C^1 function, and let p be a polynomial that agrees with f to first order on $\sigma(S(t^*))$.

By Theorem 2, there are locally Lipschitz maps $x_i(t):=(x_i^1(t),\dots,x_i^d(t))$, for $1\leqslant i\leqslant n$, representing $\sigma(S(t))$ on I. From the multivariate mean value theorem, we have

$$\begin{aligned}\|(F-P)(S(t))\| &= \max_i |(f-p)(x_i(t))| \\ &= \max_i \left|(f-p)(x_i(t))-(f-p)(x_i(t^*))\right| \\ &= \max_i \left|\nabla(f-p)(x_i^*(t))\cdot(x_i(t)-x_i(t^*))\right| \\ &\leqslant \max_i \sum_{r=1}^{d}\left|\left(\tfrac{\partial f}{\partial x^r}-\tfrac{\partial p}{\partial x^r}\right)(x_i^*(t))\right|\left|x_i^r(t)-x_i^r(t^*)\right|, \end{aligned} \tag{10}$$

where $x_i^*(t)$ is on the line connecting $x_i(t)$ and $x_i(t^*)$ in $\mathbb{R}^d$. This makes sense because continuity implies that there is a convex set $U\subseteq\Omega$ such that $x_i(t^*),\ x_i(t)\in U$, for t sufficiently close to t^*. As f and p agree to first order on $\sigma(S(t^*))$ and the $x_i(t)$ are locally Lipschitz, (10) implies

$$\|(F-P)(S(t))\| = o(|t-t^*|).$$

Hence

$$\left\|\frac{F(S(t))-F(S(t^*))}{t-t^*}-\frac{P(S(t))-P(S(t^*))}{t-t^*}\right\|\to 0 \qquad \text{as } t\to t^*.$$

Therefore,

$$\tfrac{d}{dt}F(S(t))|_{t=t^*} \text{ exists and equals } \tfrac{d}{dt}P(S(t))|_{t=t^*}.$$

Applying the same argument to $F(T(t))$ at $t = 0$ gives

$$\tfrac{d}{dt}F(T(t))|_{t=0} \text{ exists and equals } \tfrac{d}{dt}P(T(t))|_{t=0}.$$

As (ii) holds for $P(t)$, we must have $\frac{d}{dt}F(T(t))|_{t=0} = \frac{d}{dt}F(S(t))|_{t=t^*}$. $\square$

In the following proposition, we calculate an explicit formula for the derivative.

PROPOSITION 2. *Let $S(t)$ be a C^1 curve in CS_n^d defined on an interval I, and let $t^* \in I$. Let Ω be an open set in $\mathbb{R}^d$ with $\sigma(S(t)) \subset \Omega$ and let $f \in C^1(\Omega,\mathbb{R})$. Then,*

$$\tfrac{d}{dt}F(S(t))|_{t=t*} = U\left(\sum_{r=1}^{d} \tilde{\Gamma}^r \tfrac{\partial F}{\partial x^r}(D) + [Y, F(D)]\right)U^*,$$

where U diagonalizes $S(t^*)$ as in (3), $\frac{\partial F}{\partial x^r}(D)$ is defined in (12), and the other matrices are as follows:

$$D^r := U^* S^r(t^*) U \qquad \Gamma^r := U^*(S^r)'(t^*)U$$

$$\tilde{\Gamma}^r_{ij} := \begin{cases} \Gamma^r_{ij} & \text{if } x_i = x_j \\ 0 & \text{otherwise} \end{cases} \qquad Y_{ij} := \begin{cases} \frac{\Gamma^q_{ij}}{x^q_j - x^q_i} & \text{if } x_i \neq x_j \\ 0 & \text{otherwise,} \end{cases}$$

where the joint eigenvalues of $S(t^*)$ are given by $\{x_i = (x_i^1, \ldots, x_i^d) : 1 \leqslant i \leqslant n\}$ and each q is chosen so $x_j^q - x_i^q \neq 0$.

Proof. Let $t^* \in I$ and define the C^1 curve $T(t)$ by

$$T^r(t) := U\, e^{Yt}\left[D^r + t\tilde{\Gamma}^r\right]e^{-Yt}\, U^*,$$

for $1 \leqslant r \leqslant d$. Then, $T(t)$ is the curve defined in the proof of Theorem 1 for $S := S(t^*)$ and $\Delta := S'(t^*)$. It is immediate that $T(t) \in CS_n^d$, $T(0) = S(t^*)$, and $T'(0) = S'(t^*)$. By Theorem 3, it now suffices to calculate $\frac{d}{dt}F(T(t))|_{t=0}$. First, we diagonalize each $D^r + t\tilde{\Gamma}^r$. Let p be the number of distinct joint eigenvalues of $S(t^*)$. By definition,

$$\tilde{\Gamma}^r = \begin{pmatrix} \Gamma_1^r & & \\ & \ddots & \\ & & \Gamma_p^r \end{pmatrix},$$

for $1 \leqslant r \leqslant d$, where each Γ_l^r is a $k_l \times k_l$ self-adjoint matrix corresponding to a distinct joint eigenvalue of S with multiplicity k_l. It follows from Theorem 1 that

$$\left[\tilde{\Gamma}^r, \tilde{\Gamma}^s\right] = 0, \text{ which implies: } \left[\Gamma_l^r, \Gamma_l^s\right] = 0,$$

for $1 \leqslant r, s \leqslant d$ and $1 \leqslant l \leqslant p$. Thus, for each l, there is a $k_l \times k_l$ unitary matrix V_l such that V_l diagonalizes each Γ_l^r. Let V be the $n \times n$ block diagonal matrix with blocks given by $V_1,\dots,V_p$. Then, V is a unitary matrix that diagonalizes each $\tilde{\Gamma}^r$. By the diagonalization in (3), the joint eigenvalues of D are positioned so that

$$D^r = \begin{pmatrix} c_1^r I_{k_1} & & \\ & \ddots & \\ & & c_p^r I_{k_p} \end{pmatrix}, \tag{11}$$

for $1 \leqslant r \leqslant d$, where I_{k_l} is the $k_l \times k_l$ identity matrix and each c_l^r is a constant. Equation (11) shows that V and V^* will commute with D^r. Define the diagonal matrix

$$\Lambda^r := V^* \tilde{\Gamma}^r V,$$

for $1 \leqslant r \leqslant d$ and rewrite $T(t)$ as follows:

$$T^r(t) = U e^{Yt} V \big(D^r + t\Lambda^r\big) V^* e^{-Yt} U^*,$$

for $1 \leqslant r \leqslant d$. Now we directly calculate $F(T(t))$ and $\frac{d}{dt}F(T(t))|_{t=0}$ as follows:

$$\begin{aligned} F(T(t)) &= U e^{Yt} V \; F\left(D^1 + t\Lambda^1, \dots, D^d + t\Lambda^d\right) \; V^* e^{-Yt} U^* \\ &= U e^{Yt} V \left(F(D) + t \sum_{r=1}^{d} \Lambda^r \tfrac{\partial F}{\partial x^r}(D) + o(|t|) \right) V^* e^{-Yt} U^*, \end{aligned}$$

where $\frac{\partial F}{\partial x^r}(D)$ is defined by

$$\frac{\partial F}{\partial x^r}(D) := \begin{pmatrix} \frac{\partial f}{\partial x^r}(x_1) & & \\ & \ddots & \\ & & \frac{\partial f}{\partial x^r}(x_n) \end{pmatrix}, \tag{12}$$

for $1 \leqslant r \leqslant d$ and the first-order approximation of F follows from the approximation of f. Differentiating $F(T(t))$ and setting $t = 0$ gives

$$\begin{aligned} \tfrac{d}{dt} F(T(t))|_{t=0} &= U \Big(\sum_{r=1}^{d} V \, \Lambda^r \tfrac{\partial F}{\partial x^r}(D) V^* + [Y, VF(D)V^*] \Big) U^* \\ &= U \Big(\sum_{r=1}^{d} \tilde{\Gamma}^r \tfrac{\partial F}{\partial x^r}(D) + [Y, F(D)] \Big) U^*, \end{aligned}$$

where V and V^* commute with $F(D)$ and each $\frac{\partial F}{\partial x^r}(D)$ because those matrices have decompositions akin to that of D^r in (11). □

We now prove that the derivative calculated in Proposition 2 is continuous in t^*.

THEOREM 4. *Let $S(t)$ be a C^1 curve in CS_n^d defined on an interval I. Let Ω be an open set in $\mathbb{R}^d$ with $\sigma(S(t)) \subset \Omega$. If $f \in C^1(\Omega, \mathbb{R})$, then*

$$\tfrac{d}{dt} F(S(t))|_{t=t^*} \text{ is continuous as a function of } t^* \text{ on } I.$$

For the proof, we will require the following lemma:

LEMMA 3. *Let $S(t)$ be a C^1 curve in CS_n^d defined on an interval I. Let Ω be an open, convex set in $\mathbb{R}^d$ with $\sigma(S(t)) \subset \Omega$. If $f \in C^1(\Omega, \mathbb{R})$ and $t_0 \in I$, then there is a neighborhood I_0 around t_0 such that*

$$\|\tfrac{d}{dt}F(S(t))|_{t=t^*}\| \leqslant C \max_{1\leqslant s\leqslant d; x\in \bar{E}} \left|\tfrac{\partial f}{\partial x^s}(x)\right|,$$

for all $t^* \in I_0$, where C is a constant and E a convex, bounded open set with $\bar{E} \subset \Omega$.

Proof. Let $t_0 \in I$ and fix a bounded interval I_0 around t_0 with $\bar{I}_0 \subset I$. By Theorem 2, the joint eigenvalues of $S(t)$ are continuous on I_0. Thus, there exists an open, bounded, convex set $E \subset \mathbb{R}^d$ such that $\bar{E} \subset \Omega$ and $\sigma(S(t^*)) \subset E$ for each $t^* \in I_0$. Fix $t^* \in I_0$. By Proposition 2,

$$\tfrac{d}{dt}F(S(t))|_{t=t*} = U\Big(\sum_{r=1}^{d} \tilde{\Gamma}^r \tfrac{\partial F}{\partial x^r}(D) + [Y, F(D)]\Big)U^*, \tag{13}$$

where U, D^r, $\tilde{\Gamma}^r$, and Y are functions of t^* defined in Proposition 2, and the joint eigenvalues of $S(t^*)$ are denoted by x_i, for $1 \leqslant i \leqslant n$. Observe that the matrix in (13) can be rewritten as

$$\Big[\sum_{r=1}^{d} \tilde{\Gamma}^r \tfrac{\partial F}{\partial x^r}(D) + [Y, F(D)]\Big]_{ij} = \begin{cases} \sum_{r=1}^{d} \Gamma_{ij}^r \frac{\partial f}{\partial x^r}(x_i) & \text{if } x_i = x_j \\ \Gamma_{ij}^q \frac{f(x_i)-f(x_j)}{x_i^q - x_j^q} & \text{if } x_i \neq x_j, \end{cases} \tag{14}$$

where q is such that $x_i^q \neq x_j^q$, and $\Gamma_{ij}^q/(x_i^q - x_j^q)$ is the same for any q with $x_i^q \neq x_j^q$. Recall that for a given $n \times n$ self-adjoint matrix A and an $n \times n$ unitary matrix U,

$$\max_{ij} |(UAU^*)_{ij}| \leqslant n\|UAU^*\| = n\|A\| \leqslant n^2 \max_{ij} |A_{ij}|. \tag{15}$$

It is immediate from (13), (14), and (15) that

$$\left|\left|\tfrac{d}{dt}F(S(t))\Big|_{t=t^*}\right|\right| \leqslant n \max\left|\sum_{r=1}^{d} \Gamma_{ij}^r \tfrac{\partial f}{\partial x^r}(x_i)\right| + n \max\left|\Gamma_{ij}^q \frac{f(x_i) - f(x_j)}{x_i^q - x_j^q}\right|, \tag{16}$$

where the first maximum is taken over (i, j) with $x_i = x_j$, the second maximum is taken over (i, j) with $x_i \neq x_j$, and q is such that $x_i^q \neq x_j^q$. Fix (i, j) with $x_i \neq x_j$. Since $f \in C^1(E)$, we can apply the multivariate mean value theorem as follows:

$$\begin{aligned} \left|f(x_i) - f(x_j)\right| &= \left|\nabla f(x^*) \cdot (x_i - x_j)\right| \\ &\leqslant \max_{s; x\in \bar{E}} \left|\tfrac{\partial f}{\partial x^s}(x)\right| \sum_{r=1}^{d} \left|x_i^r - x_j^r\right|, \end{aligned} \tag{17}$$

where x^* is on the line in E connecting x_i and x_j. If $x_i^q \neq x_j^q$, for each r with $x_i^r \neq x_j^r$,

$$\Gamma_{ij}^q \frac{x_i^r - x_j^r}{x_i^q - x_j^q} = \Gamma_{ij}^r.$$

It follows from (17) that, for each (i,j,q) with $x_i^q \neq x_j^q$,

$$\begin{aligned} \left|\Gamma_{ij}^q \frac{f(x_i)-f(x_j)}{x_i^q - x_j^q}\right| &\leqslant \left|\frac{\Gamma_{ij}^q}{x_i^q - x_j^q}\right| \max_{s;x\in\bar{E}} \left|\tfrac{\partial f}{\partial x^s}(x)\right| \sum_{r=1}^d \left|x_i^r - x_j^r\right| \\ &\leqslant \max_{s;x\in\bar{E}} \left|\tfrac{\partial f}{\partial x^s}(x)\right| \sum_{r=1}^d \left|\Gamma_{ij}^r\right| \\ &\leqslant dn^2 \max_{s;x\in\bar{E}} \left|\tfrac{\partial f}{\partial x^s}(x)\right| \max_{i,j,r} \left|(S^r)'(t^*)_{ij}\right|, \end{aligned} \tag{18}$$

where we used (15). Likewise,

$$\Big|\sum_{r=1}^d \Gamma_{ij}^r \tfrac{\partial f}{\partial x^r}(x_i)\Big| \leqslant dn^2 \max_{s;x\in\bar{E}} \left|\tfrac{\partial f}{\partial x^s}(x)\right| \max_{i,j,r} \left|(S^r)'(t^*)_{ij}\right|. \tag{19}$$

Let M be a constant bounding each $|(S^r)'(t^*)_{ij}|$ on $\bar{I}_0$ and let $C = 2dn^3M$. Substituting (18) and (19) into (16) gives

$$\left|\left|\tfrac{d}{dt}F(S(t))\big|_{t=t^*}\right|\right| \leqslant 2dn^3 \max_{s;x\in\bar{E}} \left|\tfrac{\partial f}{\partial x^s}(x)\right| \max_{i,j,r} \left|(S^r)'(t^*)_{ij}\right| \leqslant C \max_{s;x\in\bar{E}} \left|\tfrac{\partial f}{\partial x^s}(x)\right|,$$

for all t^* in I_0. □

Proof. Theorem 4:
First assume Ω is convex. Let $t_0 \in I$. Let I_0 be the interval around t_0 and E be the convex, bounded open set given in Lemma 3. Since f is a C^1 function and $\bar{E}$ is compact, a generalization of the Stone-Weierstrass theorem in [5, pg 55] guarantees a sequence $\{\phi_k\}$ of functions analytic on $\mathbb{R}^d$ such that

$$|\phi_k(x) - f(x)| < \tfrac{1}{k} \text{ and } \left|\tfrac{\partial \phi_k}{\partial x^r}(x) - \tfrac{\partial f}{\partial x^r}(x)\right| < \tfrac{1}{k},$$

for all $k \in \mathbb{N}$, $x \in \bar{E}$, and $1 \leqslant r \leqslant d$. Lemma 3 guarantees that, for each $t^* \in I_0$,

$$\begin{aligned} \left|\left|\tfrac{d}{dt}\Phi_k(S(t))\big|_{t=t^*} - \tfrac{d}{dt}F(S(t))\big|_{t=t^*}\right|\right| &= \left|\left|\tfrac{d}{dt}(F-\Phi_k)(S(t))\big|_{t=t^*}\right|\right| \\ &\leqslant C \max_{s;x\in\bar{E}} \left|\tfrac{\partial(f-\phi_k)}{\partial x^s}(x)\right| \\ &\leqslant \tfrac{C}{k}, \end{aligned}$$

where C is a fixed constant. This implies

$$\left\{\tfrac{d}{dt}\Phi_k(S(t))\big|_{t=t^*}\right\} \text{ converges uniformly to } \tfrac{d}{dt}F(S(t))\big|_{t=t^*} \text{ on } I_0.$$

By Proposition 1, each $\frac{d}{dt}\Phi_k(S(t))|_{t=t^*}$ is continuous on I. Since the uniform limit of continuous functions is continuous, $\frac{d}{dt}F(S(t))|_{t=t^*}$ is continuous on I_0.

Now, let $\Omega \subseteq \mathbb{R}^d$ be an arbitrary open set. Fix $t_0 \in I$ and let I_0 be a bounded open interval of t_0 with $\bar{I}_0 \subset I$. Let $E \subset \mathbb{R}^d$ be a bounded open set such that $\bar{E} \subset \Omega$ and $\sigma(S(t^*)) \subset E$ for all $t^* \in I_0$. Let O be an open set and K be a compact set such that $\bar{E} \subset O \subset K \subset \Omega$ and define a C^∞ bump function $b(x)$ such that

$$b(x) := \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \in O^c. \end{cases}$$

Now we can define a function g in $C^1(\mathbb{R}^d, \mathbb{R})$ by

$$g(x) := \begin{cases} b(x)f(x) & \text{if } x \in \Omega \\ 0 & \text{if } x \in \Omega^c. \end{cases}$$

As $\mathbb{R}^d$ is convex, it follows from the previous result that $\frac{d}{dt}G(S(t))|_{t=t^*}$ is continuous on I_0. Since $f(x) = g(x)$ in E, it follows from the formula in Proposition 2 that

$$\tfrac{d}{dt}F(S(t))|_{t=t^*} = \tfrac{d}{dt}G(S(t))|_{t=t^*}$$

for all $t^* \in I_0$, and thus, is continuous in I_0. □

Recall that CS_n^d possesses a Whitney stratification with pieces $\{M_\alpha\}$ that are smooth submanifolds of $\mathbb{R}^m$, where $m = dn^2$. Let Ω be an open set in $\mathbb{R}^d$ and let $f \in C^1(\Omega, \mathbb{R})$. Let V be an open set in CS_n^d such that for all $S \in V$, $\sigma(S) \subset \Omega$. Define $TV := \cup T(M_\alpha \cap V)$. Then, $F(S)$ exists for all $S \in V$, and we can use the derivative results to define a map $DF : TV \to TS_n$.

Specifically, fix an element in TV, which will consist of an $S \in V$ and $\Delta \in T_S M_\alpha$, where M_α is the piece containing S. Let $S(t)$ be a smooth curve in M_α such that $S(0) = S$ and $S'(0) = \Delta$. Define

$$DF(S,\Delta) := \left(F(S), \tfrac{d}{dt}F(S(t))|_{t=0} \right) = \left(F(S), U\Big(\sum_{r=1}^{d} \tilde{\Gamma}^r \tfrac{\partial F}{\partial x^r}(D) + [Y, F(D)] \Big) U^* \right),$$

where U, D, $\tilde{\Gamma}^r$, and Y are defined using S and Δ as in Proposition 2, and we can set

$$\|DF(S,\Delta)\| = \max\left(\|F(S)\|, \|\tfrac{d}{dt}F(S(t))|_{t=0}\| \right).$$

It is easy to see that the map is well-defined and that $DF(S,\cdot)$ is linear in Δ, for $\Delta \in T_S(V \cap M_\alpha)$. In the following theorem, let S be in a piece M_α and let R be in a piece M_β of a Whitney stratification of CS_n^d.

THEOREM 5. *Let Ω be an open set in $\mathbb{R}^d$ and V be an open set in CS_n^d with $\sigma(S) \subset \Omega$ for all $S \in V$. If $f \in C^1(\Omega, \mathbb{R})$, then*

$$DF : TV \to TS_n \text{ is continuous.}$$

Specifically, if $S \in V$ with $\Delta \in T_S M_\alpha$, then given $\varepsilon > 0$, there exist δ_1, $\delta_2 > 0$ such that if $R \in V$ with $\Lambda \in T_R M_\beta$, $\|S - R\| < \delta_1$, and $\|\Delta - \Lambda\| < \delta_2$, then

$$\|DF(S,\Delta) - DF(R,\Lambda)\| < \varepsilon.$$

Proof. The result for analytic functions follows from (9). For an arbitrary C^1 function f defined on a convex set, and for R and Λ sufficiently close to S and Δ, bound $\|DF(R,\Lambda)\|$ in a manner similar to Lemma 3. The remainder of the proof is almost identical to that of Theorem 4 and is left as an exercise. □

4. Higher Order Derivatives

We now consider higher-order differentiation and for ease of notation, discuss only two-variable functions. We first clarify some notation. In earlier sections, $(\zeta^1,...,\zeta^d)$ referred to a point in $\mathbb{C}^d$. In this section, (ζ_1,ζ_2) denotes a point in $\mathbb{C}^2$. Previously, $S(t)$ and $T(t)$ denoted two separate curves in CS_n^d. Now, $S(t)$ and $T(t)$ denote the two components of a single curve in CS_n^2.

Let $(S(t),T(t))$ be a C^m curve in CS_n^2 defined on an interval I. If $m \geqslant 1$, the curve is locally Lipschitz. By Theorem 2, for $1 \leqslant s \leqslant n$, there are locally Lipschitz curves

$$(x_s(t),y_s(t)) \tag{20}$$

defined on I representing the joint eigenvalues of $(S(t),T(t))$. Let $U(t)$ be a unitary matrix diagonalizing $(S(t),T(t))$ so that the joint eigenvalues are ordered as in (20). To simplify notation, we write $(S(t),T(t))$ as (S,T). For $l \in \mathbb{N}$ with $1 \leqslant l \leqslant m$, define

$$S^l := S^{(l)}(t) \text{ and } T^l := T^{(l)}(t) \tag{21}$$

and the set of pairs of index tuples

$$I_l := \left\{(i_1,...,i_k)\cup(i_{k+1},...,i_j) : i_1+...+i_j = l, i_q \in \mathbb{N}, i_q \neq 0, \text{ for } 1 \leqslant q \leqslant j\right\}.$$

For example, $I_2 = \{(2)\cup\emptyset,(1,1)\cup\emptyset,(1)\cup(1),\emptyset\cup(1,1),\emptyset\cup(2)\}$. For notational ease, for $1 \leqslant s \leqslant n$, define

$$U := U(t),$$
$$x_s := x_s(t),$$
$$y_s := y_s(t).$$

For some formulas, we will conjugate the derivatives in (21) by U^* and so define

$$\Gamma^l := U^*S^lU \text{ and } \Delta^l := U^*T^lU,$$

for $1 \leqslant l \leqslant m$. We will use the integral formula given in Lemma 2 and simplify it by defining

$$R_1 := (\zeta_1 I - S)^{-1} \text{ and } R_2 := (\zeta_2 I - T)^{-1},$$

where ζ_1 and ζ_2 are in the resolvent sets of S and T respectively. Now, let J_1 and J_2 be open intervals in $\mathbb{R}$ and let f be an element of $C^m(J_1\times J_2,\mathbb{R})$. Fix j and k in $\mathbb{N}$ such that $k \leqslant j \leqslant m$. Fix $k+1$ points $x_1,...,x_{k+1}$ in J_1 and $j-k+1$ points $y_1,...,y_{j-k+1}$ in J_2. Then

$$f^{[k,j-k]}(x_1,...,x_{k+1};y_1,...,y_{j-k+1})$$

denotes the divided difference of f taken in the first variable k times and the second variable $j-k$ times, evaluated at the given points. Finally, let $\odot$ denote the Schur (also called Hadamard) product of two matrices. We will prove the following differentiability result:

THEOREM 6. *Let J_1 and J_2 be open intervals in $\mathbb{R}$, and let $f \in C^m(J_1 \times J_2, \mathbb{R})$. Let (S,T) be a C^m curve in CS_n^2 defined on an interval I with joint eigenvalues in $J_1 \times J_2$. For $1 \leqslant l \leqslant m$ and $t^* \in I$, $\frac{d^l}{dt^l}F(S,T)|_{t=t^*}$ exists and*

$$\frac{d^l}{dt^l}F(S,T)\Big|_{t=t^*} = U\Bigg(\sum_{I_l}\sum_{s_2,..,s_j=1}^{n}\frac{l!}{i_1!\cdots i_j!}\left[f^{[k,j-k]}(x_{s_1},..,x_{s_{k+1}};y_{s_{k+1}},..,y_{s_{j+1}})\right]_{s_1,s_{j+1}=1}^{n} \odot \left[\Gamma^{i_1}_{s_1 s_2}\ldots\Gamma^{i_k}_{s_k s_{k+1}}\Delta^{i_{k+1}}_{s_{k+1}s_{k+2}}\ldots\Delta^{i_j}_{s_j s_{j+1}}\right]_{s_1,s_{j+1}=1}^{n}\Bigg)U^*,$$

where the U, U^*, Γ^i, Δ^j, x_q and y_r are evaluated at t^*.

Notice that the derivative formula in Theorem 6 requires f to be defined on pairs (x_q,y_r) for $1 \leqslant r,q \leqslant n$, rather than just at the joint eigenvalues (x_q,y_q) of (S,T). This condition was not needed in Theorem 3. Before proving Theorem 6, we consider the case where f is real-analytic and show:

PROPOSITION 3. *Let J_1 and J_2 be open intervals in $\mathbb{R}$, and let f be real-analytic on $J_1 \times J_2$. Fix $m \in \mathbb{N}$ and let (S,T) be a C^m curve in CS_n^2 defined on an interval I with joint eigenvalues in $J_1 \times J_2$. Then $\frac{d^m}{dt^m}F(S,T)$ exists, has the form in Theorem 6, and $\frac{d^m}{dt^m}F(S,T)|_{t=t^*}$ is continuous as a function of t^* on I.*

The proof of Proposition 3 requires the following two technical lemmas:

LEMMA 4. *Let (S,T) be a C^m curve in CS_n^2 defined on an interval I. Let $t^* \in I$, and let ζ_1 and ζ_2 be in the resolvent sets of $S(t^*)$ and $T(t^*)$ respectively. Then*

$$\frac{d^l}{dt^l}\left(R_1R_2\right)\Big|_{t=t^*} = \sum_{I_l}\frac{l!}{i_1!\cdots i_j!}R_1S^{i_1}R_1\ldots S^{i_k}R_1R_2T^{i_{k+1}}R_2\ldots T^{i_j}R_2,$$

for $1 \leqslant l \leqslant m$, where each R_1, R_2, S^r, and T^q is evaluated at t^*.

Proof. The proof is a technical calculation using induction on l and the formulas $\frac{d}{dt}R_1 = R_1S^1R_1$ and $\frac{d}{dt}R_2 = R_2T^1R_2$. □

LEMMA 5. *Let J_1 and J_2 be open intervals in $\mathbb{R}$, and let f be real-analytic on $J_1 \times J_2$. Let $j \geqslant k \in \mathbb{N}$. Choose $k+1$ points $x_1,\ldots,x_{k+1} \in J_1$ and $j-k+1$ points $y_1,\ldots,y_{j-k+1} \in J_2$. Extend f to be analytic on a complex rectangle $\tilde{\Omega} \subset \mathbb{C}^2$ such that each $(x_q,y_r) \in \tilde{\Omega}$. Then $f^{[k,j-k]}(x_1,\ldots,x_{k+1};y_1,\ldots,y_{j-k+1})$ exists and*

$$f^{[k,j-k]}(x_1,..,x_{k+1};y_1,..,y_{j-k+1}) = \frac{1}{(2\pi i)^2}\int_{C_2}\int_{C_1}\frac{f(\zeta_1,\zeta_2)}{\prod_{q=1}^{k+1}(\zeta_1-x_q)\prod_{r=1}^{j-k+1}(\zeta_2-y_r)}d\zeta,$$

where C_1 and C_2 are simple closed rectifiable curves strictly enclosing $x_1,\dots,x_{k+1}$ and $y_1,\dots,y_{j-k+1}$ respectively, such that $C_1\times C_2\subset\tilde{\Omega}$.

Proof. For a one-variable function, the formula is proven in [4, pg 2] and the two-variable analogue follows easily from the one variable case. □

Proof. Proposition 3:
Use the integral formula in Lemma 2 to establish an integral formula for $\frac{d^m}{dt^m}F(S,T)$ similar to the first line of (9). Simplify the formula using Lemma 4. This formula implies that the derivative is continuous. Then, let E_s denote the matrix that is 1 in the ssth entry and zero elsewhere. Rewrite each R_1 as

$$R_1 = U\left(\sum_{s=1}^{n}\frac{E_s}{\zeta_1 - x_s}\right)U^*$$

and R_2 similarly. Then, use Lemma 5 to convert the derivative into a formula involving the divided differences of f. The details are left as an exercise. □

Proof. Theorem 6:
The result follows via induction on l, and the base case is covered by Theorem 3. For the inductive step, fix $t^*\in I$. Let p be a polynomial such that p and its derivatives to l^{th} order agree with f at the points $(x_q(t^*),y_r(t^*))$ for $1\leqslant q,r\leqslant n$. Using the inductive hypothesis, find a constant C such that for t near t^*,

$$\|\tfrac{d^{l-1}}{dt^{l-1}}F(S,T)-\tfrac{d^{l-1}}{dt^{l-1}}P(S,T)\|\leqslant C\max\left|(f-p)^{[k,j-k]}(x_{s_1},..,x_{s_{k+1}};y_{s_{k+1}},..,y_{s_{j+1}})\right|,$$

where the joint eigenvalues of (S,T) are given by (x_q,y_q) and the maximum is over (k,j) with $k\leqslant j<l\in\mathbb{N}$ and sets $\left\{(s_1,..,s_{k+1})\cup(s_{k+1},..,s_{j+1}) : 1\leqslant s_1,..,s_{j+1}\leqslant n\right\}$. The proof now mirrors that of Theorem 3. Specifically, apply the multivariate mean value theorem to each $(f-p)^{[k,j-k]}$ and observe that, by our original assumptions, $(f-p)^{[k,j-k]}$ vanishes to first order at the points $(x_{s_1}(t^*),..,x_{s_{k+1}}(t^*);y_{s_{k+1}}(t^*),..,y_{s_{j+1}}(t^*))$. Then, use the locally Lipschitz property of the eigenvalues to conclude

$$\tfrac{d^l}{dt^l}F(S,T)|_{t=t^*}\text{ exists and equals }\tfrac{d^l}{dt^l}P(S,T)|_{t=t^*}.$$

The details are left as an exercise. □

We now show that the formula in Theorem 6 is continuous.

THEOREM 7. *Let J_1 and J_2 be open intervals in $\mathbb{R}$ and $f\in C^m(J_1\times J_2,\mathbb{R})$. Let (S,T) be a C^m curve in CS_n^2 defined on an interval I with joint eigenvalues in $J_1\times J_2$. Then for all $l\in\mathbb{N}$ with $1\leqslant l\leqslant m$,*

$$\tfrac{d^l}{dt^l}F(S,T)|_{t=t^*}\textit{ is continuous as a function of }t^*\textit{ on }I.$$

For the proof, we require the following lemma. The result is well-known for one-variable functions, and Brown and Vasudeva prove this two-variable analogue in [3]:

LEMMA 6. *Let J_1 and J_2 be open intervals in $\mathbb{R}$, and let $f \in C^m(J_1 \times J_2, \mathbb{R})$. Choose $j,k \in \mathbb{N}$ with $k \leqslant j \leqslant m$. Let $x_1,\dots,x_{k+1} \in J_1$ and $y_1,\dots,y_{j-k+1} \in J_2$, and choose closed subintervals $\tilde{J}_1$ and $\tilde{J}_2$ containing the x and y points respectively. Then, there exists $(x^*,y^*) \in \tilde{J}_1 \times \tilde{J}_2$ with*

$$f^{[k,j-k]}(x_1,\dots,x_{k+1};y_1,\dots,y_{j-k+1}) = \frac{f^{(k,j-k)}(x^*,y^*)}{k!(j-k)!}.$$

Proof. Theorem 7:
For $l < m$, the result follows from Theorem 6, which implies that $\frac{d^l}{dt^l}F(S,T)$ is differentiable and hence, continuous.

For $l = m$, fix $t_0 \in I$. Similarly to Lemma 3, find a constant C and closed, bounded intervals $\tilde{J}_1$ and $\tilde{J}_2$ such that if $\tilde{J} := \tilde{J}_1 \times \tilde{J}_2$, then $\tilde{J} \subset J_1 \times J_2$ and for all $g \in C^m(J_1 \times J_2, \mathbb{R})$ and t^* near t_0,

$$\left|\left|\tfrac{d^m}{dt^m}G(S,T)|_{t=t^*}\right|\right| \leqslant C \max_{\{j,k;(x,y)\in\tilde{J}\}} |g^{(k,j-k)}(x,y)|, \tag{22}$$

where $0 \leqslant k \leqslant j \leqslant m$. The estimates for this bound require Lemma 6. Now, approximate f to m^{th} order uniformly on $\tilde{J}$ by analytic functions $\{\phi_r\}$ and use (22) to show

$$\{\tfrac{d^m}{dt^m}\Phi_r(S,T)|_{t=t^*}\} \text{ converges uniformly to } \tfrac{d^m}{dt^m}F(S,T)|_{t=t^*}$$

for t^* in a neighborhood of t_0. The result then follows from Proposition 3. □

5. Applications

The formulas in Proposition 2 and Theorem 6 can be used to analyze monotonicity and convexity of matrix functions. A function $F : S_n \to S_n$ is *matrix monotone* if

$$F(A) \geqslant F(B) \text{ whenever } A \geqslant B, \ \ \forall\, A,B \in S_n.$$

For F continuously differentiable, an equivalent condition is

$$\tfrac{d}{dt}F(S(t))|_{t=t^*} \geqslant 0 \text{ whenever } S'(t^*) \geqslant 0, \ \ \forall\, C^1\ S(t) \subset S_n. \tag{23}$$

The local monotonicity condition in (23) extends to multivariate matrix functions: the only adjustment is that $S(t)$ is in CS_n^d. In [1], Agler, McCarthy, and Young characterized such locally matrix monotone functions on CS_n^d using a special case of Theorem 3 and Proposition 2. Specifically, they had to assume that $S(t)$ had distinct joint eigenvalues at each t. Our results in Section 3 extend the derivative formula to general C^1 curves in CS_n^d and show that the formula is continuous.

A matrix function $F : S_n \to S_n$ is *matrix convex* if

$$F(\lambda A + (1-\lambda)B) \leqslant \lambda F(A) + (1-\lambda)F(B) \ \ \forall\, A,B \in S_n \text{ and } \lambda \in [0,1]. \tag{24}$$

This condition extends to multivariate matrix functions with an additional restriction on the pairs A,B in CS_n^d; we also require $\lambda A + (1-\lambda)B \in CS_n^d$ for $\lambda \in (0,1)$. Given such

A,B, define the curve $S(t)$ on $[0,1]$ by

$$S^r(t) := tA^r + (1-t)B^r, \tag{25}$$

for $1 \leqslant r \leqslant d$. If F is twice continuously differentiable along C^2 curves, it can be shown that (24) is equivalent to

$$\tfrac{d^2}{dt^2}F(S(t))|_{t=t^*} \geqslant 0$$

for all $S(t)$ as in (25) and $t^* \in (0,1)$. For $d=2$, Theorem 6 tells us that, up to conjugation by a unitary matrix U diagonalizing $S(t^*)$,

$$\left[\tfrac{d^2}{dt^2}F(S(t))|_{t=t^*}\right]_{ij} = 2\sum_{k=1}^{n} f^{[2,0]}(x_i,x_k,x_j;y_j)\Gamma_{ik}\Gamma_{kj} + f^{[1,1]}(x_i,x_k;y_k,y_j)\Gamma_{ik}\Delta_{kj} + f^{[0,2]}(x_i;y_i,y_k,y_j)\Delta_{ik}\Delta_{kj}, \tag{26}$$

where $\{(x_i,y_i) : 1 \leqslant i \leqslant n\}$ are the joint eigenvalues of $t^*A + (1-t^*)B$ ordered as in the diagonalization given by U, and

$$\Gamma := U^*\left(A^1 - B^1\right)U \ \text{ and } \Delta := U^*\left(A^2 - B^2\right)U.$$

Although U might not diagonalize $S(t^*)$ so as to order the joint eigenvalues as in (3), the first relationship stated in Theorem 1 still applies to Γ and Δ. Specifically,

$$(x_i - x_j)\Delta_{ij} = (y_i - y_j)\Gamma_{ij}$$

for $1 \leqslant i,j \leqslant n$ and we can use this to simplify (26). Thus, this formula gives a characterization of convex matrix functions on CS_n^2.

REFERENCES

[1] J. AGLER AND J. MCCARTHY AND N. J. YOUNG, *Operator Monotone Functions and Loewner Functions of Several Variables*, Ann. of Math., to appear.
[2] R. BHATIA, *Matrix Analysis*, Springer, New York/Berlin, 1997.
[3] A.L. BROWN AND H.L. VASUDEVA, *The calculus of operator functions and operator convexity*, Dissertationes Mathematicae, Polska Akademia Nauk, Instytut Matematyczny **390**, 1 (2000).
[4] W.F. DONOGHUE, *Monotone Matrix Functions and Analytic Continuations*, Springer, Berlin, 1974.
[5] M.P. HEBLE, *Approximation Problems in Analysis and Probability*, Elsevier Science Publishers B.V., New York, 1989.
[6] R.A. HORN AND C.R. JOHNSON, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.
[7] V. KALOSHIN, *A geometric proof of the existence of Whitney stratifications* , Mosc. Math. J. **5**, 1 (2005), 125–133.
[8] T. KATO, *Perturbation Theory for Linear Operators*, Springer-Verlang, Berlin, 1966.
[9] F. RELLICH, *Störungstheorie der spektralzerlegung, I*, Ann. of Math. **113**, 1 (1937), 600–619.
[10] F. RELLICH, *Störungstheorie der spektralzerlegung, II*, Ann. of Math. **113**, 1 (1937), 677–685.
[11] H. WHITNEY, *Tangents to an analytic variety*, Ann. of Math. **81**, 2 (1965), 496–549.

(Received January 27, 2011)

Kelly Bickel
1 Brookings Drive Box 1146
St. Louis, MO 63110
USA
e-mail: kbickel@math.wustl.edu