Differentiating Matrix Functions *†

Kelly Bickel ‡
Washington University
St. Louis, Missouri 63130
kbickel@math.wustl.edu

Abstract

Multivariate, real-valued functions on $\mathbb{R}^d$ induce matrix-valued functions 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 the best notion of differentiation of these matrix functions is differentiation along curves. We prove that $C^1$ real-valued functions induces $C^1$ matrix functions and give a formula for the derivative. We also show that real-valued $C^m$ functions defined on open rectangles in $\mathbb{R}^2$ induce matrix functions that can be $m$-times continuously differentiable along $C^m$ 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 $\Omega$. Let $S = (S^1, \ldots, S^d)$ be such a $d$-tuple diagonalized by a unitary matrix $U$ as follows:

$$S^r = U \begin{pmatrix} x_1^r & \cdots & \vdots \\ \vdots & \ddots & \vdots \\ x_n^r \end{pmatrix} U^* \quad \forall 1 \leq r \leq d.$$  

Denote the joint spectrum of $S$ by $\sigma(S) := \{ x_i = (x_i^1, \ldots, x_i^d) : 1 \leq i \leq n \}$ and define

$$F(S) := U \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{pmatrix} U^* \quad (1.1)$$

*2010 MSC: Primary 26B05, Secondary 15A60.
†Key Words: Matrix functions, Differentiability, Derivative formulas.
‡This research was partially supported by National Science Foundation grant DMS-0966845.
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 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 [8], eigenvector behavior at points where distinct eigenvalues coalesce can be unpredictable. Specifically, let

\[
S(t) = e^{-\frac{1}{t^2}} \begin{pmatrix}
\cos(\frac{1}{t}) & \sin(\frac{1}{t}) \\
\sin(\frac{1}{t}) & -\cos(\frac{1}{t})
\end{pmatrix}
\]

for \( t \neq 0 \), 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

\[
\pm \begin{pmatrix}
\cos(\frac{1}{t}) \\
\sin(\frac{1}{t})
\end{pmatrix}
\quad \text{and} \quad
\pm \begin{pmatrix}
\sin(\frac{1}{t}) \\
-\cos(\frac{1}{t})
\end{pmatrix}.
\]

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 \( m \)-times continuously differentiable real functions induce \( m \)-times continuously Fréchet differentiable matrix functions.

If a matrix function is defined using a real-valued function on \( \mathbb{R}^d \) as in (1.1), 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 \( S_n \).

In Section 2, we analyze the geometry of \( CS_n^d \) and conclude that the best notion of differentiability for functions on this space is differentiation along curves. If we fix \( S \) in \( CS_n^d \), Theorem 2.3 characterizes the directions \( \Delta \) 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 2.5, we show that the joint eigenvalues of Lipschitz curves in \( CS_n^d \) can be represented by Lipschitz functions.

In Section 3, we examine the differentiability properties of induced matrix functions. Specifically, in Theorem 3.1, 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 3.6, prove that it is continuous.
In Section 4, we consider higher-order differentiation. With additional domain restrictions, in Theorem 4.1, we show that an induced matrix function is $m$-times continuously differentiable along $C^m$ curves. We also calculate a formula for the derivatives and in Theorem 4.5, show they are continuous. In Section 5, we discuss several applications of the differentiability results.

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

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 $CS^d_n$

Let $S = (S^1, \ldots, S^d)$ be in $CS^d_n$ (or $S^d_n$) and let $x_i = (x^1_i, \ldots, x^d_i)$ be in $\sigma(S)$. Define

\[ \|S\| := \max_{1 \leq r \leq d} \|S^r\| \quad \text{and} \quad \|x_i\| := \max_{1 \leq r \leq d} |x^r_i|, \]

where $\|S^r\|$ is the usual operator norm. Observe that $CS^d_n$ 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 $CS^d_n$ because both require the function to be defined on linear sets around each point.

Recall that $CS^d_n$ is the zero set of the polynomials associated with $d(d-1)/2$ commutator operations and so is an algebraic variety. A result by Whitney [10] says every algebraic variety can be decomposed into submanifolds 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 $N$. To make Whitney’s conditions more precise, we need the following definition:

**Definition 2.1** A stratification of $X$ is a locally finite partition $Z$ of $X$ such that

(i) Each piece $M_\alpha \in Z$ is a smooth submanifold of $X$.

(ii) The frontier of each piece $\overline{M_\alpha} \setminus M_\alpha$ is either trivial or a union of other pieces.

Then $X$ is called a stratified space with stratification $Z$.

**Example 2.2** Consider $CS^2_2$, the space of pairs of self-adjoint, commuting $2 \times 2$ matrices. In the following definitions, $a, b, c, d \in \mathbb{R}$. Define

\[ M_1 := \left\{ \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} U^*, \begin{pmatrix} c & 0 \\ 0 & d \end{pmatrix} U^* : U \in S_2 \text{ is unitary, } a \neq b, c \neq d \right\} \]
\[
    M_2 := \left\{ \left( \begin{array}{cc} a & 0 \\ 0 & a \end{array} \right), \left( \begin{array}{cc} c & 0 \\ 0 & d \end{array} \right) \right\} U^* : U \in S_2 \text{ is unitary, } c \neq d \}
\]
\[
    M_3 := \left\{ \left( \begin{array}{cc} a & 0 \\ 0 & b \end{array} \right) U^*, \left( \begin{array}{cc} c & 0 \\ 0 & c \end{array} \right) \right\} : U \in S_2 \text{ is unitary, } a \neq b \}
\]
\[
    M_4 := \left\{ \left( \begin{array}{cc} a & 0 \\ 0 & a \end{array} \right), \left( \begin{array}{cc} c & 0 \\ 0 & c \end{array} \right) \right\}
\]

It is clear that \( CS_2 \times S_2 = \bigcup M_i \). Moreover, each \( M_i \) is a manifold and \( \overline{M_i} \backslash M_i \) is either trivial or a union of other \( M_j \). Thus, the partition \( \{ M_i \} \) is a stratification of \( CS_2 \times S_2 \).

In general, a decomposition of \( CS_n \) into pieces will be related to the number and multiplicity of the repeated joint eigenvalues of the elements of \( CS_n \).

Whitney’s result says \( CS_n \) has a stratification \( Z \) with further regularity. Specifically, let \( \{ M_\alpha \} \) denote the pieces of \( Z \) and define \( TCS_n := \bigcup TM_\alpha \). Then, \( TCS_n \) is also a stratified space, and we call \( Z \) a Whitney stratification of \( CS_n \). Given a function \( F : CS_n \to S_n \), one type of derivative is a map
\[
    DF : TCS_n \to TS_n \quad \text{such that } \quad DF|_{TM_\alpha} : TM_\alpha \to TS_n
\]
is the usual differential map for each \( M_\alpha \). In Theorem 3.8, we analyze such maps. However, these differential maps cannot be easily generalized to analyze higher-order differentiation. Furthermore, the space \( TCS_n \) will only contain a subset of the vectors tangent to \( CS_n \). Example 2.4 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 \( S_n \) are vectors tangent to \( CS_n \) at a given point \( S \). This is equivalent to the following question:

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

For an element \( S \in CS_n \) with distinct joint eigenvalues, Agler, McCarthy, and Young in [1] gave necessary and sufficient conditions on \( S \) and \( \Delta \) for such a \( C^1 \) curve to exist. We extend their result to an arbitrary element \( S \). Fix \( S \in CS_n \) and \( \Delta \in S_n \). Let \( U \) be a unitary matrix diagonalizing each component of \( S \) such that the repeated joint eigenvalues appear consecutively. Renumbering the \( x_i \)'s if necessary, define
\[
    D^r := U^* S^r U = \begin{pmatrix} x_1^r \\ \vdots \\ x_n^r \end{pmatrix} \quad \forall \ 1 \leq r \leq d.
\]
(2.2)

For each \( r \), define the two matrices
\[
    \Gamma^r := U^* \Delta^r U
\]
\[
    \tilde{\Gamma}^r_{ij} := \begin{cases} 
        \Gamma^r_{ij} & \text{if } x_i = x_j \\
        0 & \text{otherwise.}
    \end{cases}
\]
(2.3)
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 2.3** Let $S \in CS_d^n$ and $\Delta \in S_d^n$. There exists a $C^1$ curve $S(t)$ in $CS_d^n$ with $S(0) = S$ and $S'(0) = \Delta$ iff

$$[D^r, \Gamma^s] = [D^s, \Gamma^r] \text{ and } [\tilde{\Gamma}^r, \tilde{\Gamma}^s] = 0 \quad \forall \ 1 \leq r, s \leq d.$$ 

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

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

where $U$ diagonalizes $S$ as in (2.2). Then $R(t)$ is a $C^1$ curve in $CS_d^n$ with $S(0) = D$ and $S'(0) = \Gamma$. We will first prove that

$$[D^r, \Gamma^s] = [D^s, \Gamma^r] \text{ and } [\Gamma^r, \Gamma^s]_{ij} = 0 \quad \forall \ 1 \leq r, s \leq d \text{ and } (ij) \text{ such that } x_i = x_j.$$

We will use those commutativity results to conclude

$$[\tilde{\Gamma}^r, \tilde{\Gamma}^s] = 0 \quad \forall \ 1 \leq r, s \leq d.$$

Since $R(t)$ is $C^1$ in a neighborhood of $t = 0$, we can write

$$R^r(t) = D^r + \Gamma^r t + h^r(t) \quad \forall \ 1 \leq r \leq d,$$

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

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

$$= \left[ D^r + \Gamma^r t + h^r(t), D^s + \Gamma^s t + h^s(t) \right]$$

$$= \left( [D^r, h^s(t)] + [h^r(t), D^s] + [h^r(t), h^s(t)] \right)$$

$$+ \left( [D^r, \Gamma^s] + [\Gamma^r, D^s] + [\Gamma^r, h^s(t)] + [h^r(t), \Gamma^s] \right) t$$

$$+ [\Gamma^r, \Gamma^s] t^2, \quad (2.4)$$

where the term $[D^r, D^s]$ was omitted because it vanishes. Fix $t \neq 0$ and divide each term in (2.4) by $t$. Letting $t$ tend towards zero yields

$$0 = [D^r, \Gamma^s] - [D^s, \Gamma^r]. \quad (2.5)$$

Choose $i$ and $j$ such that $x_i = x_j$. Then, the $ij^{th}$ entry of (2.4) reduces to

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

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

$$0 = [\Gamma^r, \Gamma^s]_{ij}. \quad (2.6)$$

5
Define a skew-Hermitian matrix $Y$

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

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

$$[\tilde{\Gamma}^r, \tilde{\Gamma}^s]_{ij} = \sum_{k=1}^{n} \tilde{\Gamma}^r_{ik} \tilde{\Gamma}^s_{kj} - \tilde{\Gamma}^s_{ik} \tilde{\Gamma}^r_{kj}$$

$$= \sum_{\{k: x_k = x_i\}} \Gamma^r_{ik} \Gamma^s_{kj} - \Gamma^s_{ik} \Gamma^r_{kj}$$

$$= \sum_{\{k: x_k \neq x_i\}} \Gamma^r_{ik} \Gamma^s_{kj} - \Gamma^s_{ik} \Gamma^r_{kj},$$

where the last equality uses (2.6). 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 $x_k^q \neq x_i^q$. Apply (2.5) to pairs $r, q$ and $s, q$ to get

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

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

$$\Gamma^r_{ik}(x_i^q - x_k^q) = \Gamma^q_{ik}(x_i^r - x_k^r) \quad \Gamma^r_{kj}(x_k^q - x_j^q) = \Gamma^q_{kj}(x_k^r - x_j^r)$$

$$\Gamma^s_{ik}(x_i^q - x_k^q) = \Gamma^q_{ik}(x_i^s - x_k^s) \quad \Gamma^s_{kj}(x_k^q - x_j^q) = \Gamma^q_{kj}(x_k^s - x_j^s).$$

Since $x_i = x_j$ and $x_k^q \neq x_i^q$, we can replace all the $x_j$'s with $x_i$'s in (2.7) and solve for the $\Gamma^r$ and $\Gamma^s$ entries. Using these relations gives

$$\Gamma^r_{ik} \Gamma^s_{kj} - \Gamma^s_{ik} \Gamma^r_{kj} = \frac{\Gamma^q_{ik}(x_i^r - x_k^r)}{(x_i^q - x_k^q)^2} \left( \frac{\Gamma^q_{ij}(x_i^q - x_j^q)}{(x_i^q - x_j^q)^2} - \frac{\Gamma^q_{kj}(x_k^q - x_j^q)}{(x_i^q - x_j^q)^2} \right) = 0,$$

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

$(\Leftarrow)$ Fix $S \in CS^d_n$ and $\Delta \in S^d_n$ and let $U, D$, and $\Gamma$ be as in the discussion preceding Theorem 2.3. Assume

$$[D^r, \Gamma^s] = [D^s, \Gamma^r] \quad \text{and} \quad [\tilde{\Gamma}^r, \tilde{\Gamma}^s] = 0 \quad \forall \ 1 \leq r, s \leq d.$$ 

Define a skew-Hermitian matrix $Y$ as follows:

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

6
where the $q$ is chosen so that $x_i^q - x_j^q \neq 0$. Observe that $Y$ is independent of $q$ because the $ij^{th}$ entry of the first equation in (2.8) is
\[ \Gamma_{ij}^r(x_i^r - x_j^r) = \Gamma_{ij}^r(x_i^s - x_j^s). \]
Now, define the curve $S(t)$ by
\[ S^r(t) := U e^{Y^t [D^r + t\tilde{\Gamma}^r]} e^{-Y^t} U^* \quad \forall \ 1 \leq r \leq d. \]
Then, $S(t)$ is continuously differentiable. Because $Y$ is skew-Hermitian, $e^{Y^t}$ is unitary. Since $D^r$ and $\tilde{\Gamma}^r$ are self-adjoint, $S(t) \in S_n^d$. By a simple calculation using (2.8),
\[ [S^r(t), S^s(t)] = 0 \quad \forall \ 1 \leq r, s \leq d. \]
Thus, $S(t) \in CS_n^d$. By definition, $S(0) = S$. For each $r$,
\[ (S^r)'(t) = U \left( Y e^{Y^t [D^r + t\tilde{\Gamma}^r]} e^{-Y^t} + e^{Y^t [\tilde{\Gamma}^r]} e^{-Y^t} - e^{Y^t [D^r + t\tilde{\Gamma}^r]} Y e^{-Y^t} \right) U^*, \]
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.

\[ \square \]

**Example 2.4** Let $I \in CS_n^d$ be the identity element. By Theorem 2.3, there is a continuously differentiable curve $S(t)$ in $CS_n^d$ with
\[ S(0) = I \text{ and } S'(0) = \Delta \quad \text{if and only if} \quad \Delta \in CS_n^d. \]
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 $M_\alpha$ 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 2.3 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 $M_\alpha$ 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 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 [8] and [9] states that the eigenvalues of $S(t)$ can be represented by $n$ continuous functions. A succinct proof is given by Kato in [7, pg 107-10]. With slight modification, the arguments show that the eigenvalues of a Lipschitz curve in $S_n$ can be represented by Lipschitz functions. These results generalize as follows:
Theorem 2.5  Given a Lipschitz curve $S(t)$ in $CS^n_d$ defined on an interval $I$, there exist Lipschitz functions $x_1(t), \ldots, x_n(t) : I \rightarrow \mathbb{R}^d$ with $\sigma(S(t)) = \{x_i(t) : 1 \leq i \leq n\}$.

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

Theorem 2.5 provides a specific ordering of the eigenvalues of $S(t)$ at each $t$. This ordering may differ from the one in (2.2), where joint eigenvalues appear consecutively. However, Theorem 2.5 implies that the eigenvalues of a Lipschitz curve $S(t)$ are Lipschitz as an unordered $n$-tuple. Specifically, fix $t^*$ and denote the eigenvalues of $S(t^*)$ by $\{x_i : 1 \leq i \leq n\}$. Then, for $t$ near $t^*$, there is a constant $c$ such that

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

where the minimum is taking over all reorderings of the $\{x_i\}$. If we require that eigenvalues are ordered as in (2.2), we will use Theorem 2.5 to conclude that the eigenvalues are 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 (1.1). We denote its domain, the space of $d$-tuples of pairwise-commuting $n \times n$ self-adjoint matrices with spectrum in $\Omega$, 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 of a continuous function is continuous. We now consider differentiability and prove:

Theorem 3.1  Let $S(t)$ be a $C^1$ curve in $CS^n_d$ defined on an interval $I$ and let $\Omega$ 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

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

We say an open set $\Omega \subset \mathbb{R}^d$ is a rectangle if $\Omega = I^1 \times \cdots \times I^d$, and an open set $\Omega \subset \mathbb{C}^d$ is a complex rectangle if $\Omega = (I^1 + iJ^1) \times \cdots \times (I^d + iJ^d)$, where each $I^r$ and $J^r$ is an open interval in $\mathbb{R}$. Before proving Theorem 3.1, we assume $f$ is real-analytic and prove Proposition 3.2. See [6] for the one-variable case.
Proposition 3.2 Let \( S(t) \) be a \( C^1 \) curve in \( CS_n^d \) defined on an interval \( I \). Let \( \Omega \) be an open rectangle in \( \mathbb{R}^d \) with \( \sigma(S(t)) \subset \Omega \). If \( f \) is a real-analytic function on \( \Omega \), then

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

The proof of Proposition 3.2 requires the following two lemmas.

Lemma 3.3 Let \( \Omega \) be an open rectangle in \( \mathbb{R}^d \) and let \( S \in CS_n^d \) with \( \sigma(S) \subset \Omega \). Each real-analytic function on \( \Omega \) 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 \( \Omega \) need not contain \( \tilde{\Omega} \). \( \square \)

Lemma 3.4 Let \( \tilde{\Omega} \) be an open rectangle in \( \mathbb{C}^d \) and let \( S \in CS_n^d \) with \( \sigma(S) \subset \tilde{\Omega} \). If \( f \) is an analytic function on \( \tilde{\Omega} \), then

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

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

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

Proof of Proposition 3.2:
For ease of notation, assume \( d = 2 \) and define

\[
R^r(t) := (\zeta^r I - S^r(t))^{-1} \quad \forall 1 \leq r \leq 2,
\]

where \( \zeta^r \) is in the resolvent of \( S^r(t) \). Fix \( t_0 \in I \) and extend \( f \) to an analytic function on a complex open rectangle \( \tilde{\Omega} \) containing \( \sigma(S(t_0)) \). Choose rectifiable curves \( C^1 \) and \( C^2 \) such that \( C^1 \times C^2 \subset \tilde{\Omega} \) and each \( C^r \) strictly encloses the eigenvalues of \( S^r(t_0) \). By Theorem 2.5, the joint eigenvalues of \( S(t) \) are continuous and by Lemma 3.4,

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

for \( t \) sufficiently close to \( t_0 \). Direct calculation gives

\[
\frac{d}{dt} F^r(t)|_{t=t^*} = R^r(t^*) (S^r(t^*))' R^r(t^*) \quad \text{for } 1 \leq r \leq 2 \text{ and } t^* \text{ near } t_0.
\]

It can be easily shown that, for \( t^* \) sufficiently close to \( t_0 \), we can interchange integration and differentiation to yield

\[
\frac{d}{dt} F(S(t))|_{t=t^*} = \frac{1}{(2\pi i)^2} \int_{C^2} \int_{C^1} f(\zeta^1, \zeta^2) \frac{d}{dt} \left( R^1(t) R^2(t) \right) |_{t=t^*} d\zeta^1 d\zeta^2
\]

\[
= \frac{1}{(2\pi i)^2} \int_{C^2} \int_{C^1} f(\zeta^1, \zeta^2) \left( R^1(t^*) (S^1)'(t^*) R^1(t^*) R^2(t^*)
\right.
\]

\[
+ R^1(t^*) R^2(t^*) (S^2)'(t^*) R^2(t^*) \bigg) \, d\zeta^1 d\zeta^2.
\]

(3.9)
As each \((S')'(t)\) is continuous and all other terms in (3.9) are uniformly bounded near \(t_0\), we get \(\frac{d}{dt}F(S(t))|_{t=t^*}\) is continuous at \(t^* = t_0\). \(\square\)

Proof of Theorem 3.1:
Observe that the theorem holds for polynomials: \((i)\) follows from Proposition 3.2 and \((ii)\) follows from the formula in (3.9). 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 2.5, there are Lipschitz maps \(x_i(t) := (x^1_i(t), \ldots, x^d_i(t))\), for \(1 \leq i \leq n\), representing \(\sigma(S(t))\) on \(I\). From the multivariate Mean Value Theorem, we have

\[
\| (F - P)(S(t)) \| = \max_i |(f - p)(x_i(t))| \\
= \max_i |(f - p)(x_i(t)) - (f - p)(x_i(t^*))| \\
= \max_i |\nabla (f - p)(x_i^*(t)) \cdot (x_i(t) - x_i(t^*))| \\
\leq \max_i \sum_{r=1}^d \left| \frac{\partial f}{\partial x^r} - \frac{\partial p}{\partial x^r} \right| (x_i^*(t)) \| x_i^*(t) - x_i^*(t^*) \|, \tag{3.10}
\]

where \(x_i^*(t)\) is on the line connecting \(x_i(t)\) and \(x_i(t^*)\) in \(\mathbb{R}^d\). For \(t\) near \(t^*\), continuity implies \(x_i^*(t) \in \Omega\). As \(f\) and \(p\) agree to first order on \(\sigma(S(t^*))\), from (3.10), we have

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

Hence

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

Therefore,

\[
\frac{d}{dt}F(S(t))|_{t=t^*}\quad \text{exists and equals} \quad \frac{d}{dt}P(S(t))|_{t=t^*}.
\]

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

\[
\frac{d}{dt}F(T(t))|_{t=0}\quad \text{exists and equals} \quad \frac{d}{dt}P(T(t))|_{t=0}.
\]

As \((ii)\) holds for \(P(t)\), we must have

\[
\frac{d}{dt}F(T(t))|_{t=0} = \frac{d}{dt}F(S(t))|_{t=t^*}. \quad \square
\]

In the following proposition, we calculate an explicit formula for the derivative.
Proposition 3.5 Let $S(t)$ be a $C^1$ curve in $CS_n^d$ defined on an interval $I$ and let $t^* \in I$. Let $\Omega$ be an open set in $\mathbb{R}^d$ with $\sigma(S(t)) \subset \Omega$ and let $f \in C^1(\Omega, \mathbb{R})$. Then,

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

where $U$ diagonalizes $S(t^*)$ as in (2.2) and the other matrices are as follows:

$$D^r := U^* \left[ S^r(t^*) \right] U \quad \Gamma^r := U^* \left[ (S^r)'(t^*) \right] U$$

$$\tilde{\Gamma}^r_{ij} = \begin{cases} \Gamma^r_{ij} & \text{if } x_i = x_j \\ 0 & \text{otherwise} \end{cases} \quad Y_{ij} := \begin{cases} \frac{r^q_{ij}}{x_j^q - x_i^q} & \text{if } x_i \neq x_j \\ 0 & \text{otherwise}, \end{cases}$$

where the joint eigenvalues of $S(t^*)$ are given by $\left\{ x_i = (x_1^1, \ldots, x_i^d) : 1 \leq i \leq n \right\}$ and $q$ 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

$$T^r(t) := U e^{Y(t)[D^r + t^{\tilde{\Gamma}^r}] e^{-Yt}} U^* \quad \forall 1 \leq r \leq d.$$ 

Then, $T(t)$ is the curve defined in the proof of Theorem 2.3 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 3.1, 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,

$$\tilde{\Gamma}^r = \begin{pmatrix} \Gamma^1_{ij} & \cdots \\ \cdots & \cdots \\ \Gamma^p_{ij} \end{pmatrix} \quad \forall 1 \leq r \leq d,$$

where each $\Gamma^r_{ij}$ 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 2.3 that

$$[\tilde{\Gamma}^r, \tilde{\Gamma}^s] = 0, \text{ which implies: } [\Gamma^r_{ij}, \Gamma^s_{ij}] = 0 \quad \forall 1 \leq r, s \leq d \text{ and } 1 \leq l \leq p.$$ 

Thus, for each $l$, there is a $k_l \times k_l$ unitary matrix $V_l$ such that $V_l$ diagonalizes each $\Gamma^r_{ij}$. Let $V$ be the $n \times n$ block diagonal matrix with blocks given by $V_1, \ldots, V_p$. Then, $V$ is a unitary matrix that diagonalizes each $\tilde{\Gamma}^r$. By the diagonalization in (2.2), the joint eigenvalues of $D$ are positioned so that

$$D^r = \begin{pmatrix} c_{i}^r I_{k_1} & \cdots \\ \cdots & \cdots \\ c_{p}^r I_{k_p} \end{pmatrix} \quad \forall 1 \leq r \leq d,$$

where $I_{k_l}$ is the $k_l \times k_l$ identity matrix and $c_{i}^r$ is a constant. Equation (3.11) shows that conjugation by $V$ will not affect $D^r$. Define the diagonal matrix

$$\Lambda^r := V^* \tilde{\Gamma}^r V \quad \forall 1 \leq r \leq d,$$
and rewrite $T(t)$ as follows

$$T^r(t) = U e^{Y^t V [D^r + t\Lambda^r]} V^* e^{-Y^t} U^* \quad \forall 1 \leq r \leq d.$$ 

Now we directly calculate $F(T(t))$ and $\frac{d}{dt} F(T(t))|_{t=0}$

$$F(T(t)) = U e^{Y^t V F\left(D^1 + t\Lambda^1, \ldots, D^d + t\Lambda^d\right)} V^* e^{-Y^t} U^*$$

$$= U e^{Y^t V \left(F(D) + t \sum_{r=1}^d \Lambda^r \frac{\partial f}{\partial x^r}(D) + o(|t|)\right)} V^* e^{-Y^t} U^*,$$

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

$$\frac{\partial f}{\partial x^r}(D) := \begin{pmatrix} \frac{\partial f}{\partial x^1}(x_1) \\ \vdots \\ \frac{\partial f}{\partial x^n}(x_n) \end{pmatrix} \quad \forall 1 \leq r \leq d,$$

and the first-order approximation of $F$ follows from the approximation of $f$ on each diagonal entry of the $d$-tuple of diagonal matrices. Differentiating $F(T(t))$ and setting $t = 0$ gives

$$\frac{d}{dt} F(T(t))|_{t=0} = U \left( \sum_{r=1}^d V \Lambda^r \frac{\partial f}{\partial x^r}(D)V^* + [Y, V F(D)V^*] \right) U^*$$

$$= U \left( \sum_{r=1}^d \tilde{T}^r \frac{\partial f}{\partial x^r}(D) + [Y, F(D)] \right) U^*,$$

where conjugation by $V$ leaves $F(D)$ and each $\frac{\partial f}{\partial x^r}(D)$ unchanged because those matrices have decompositions akin to that of $D^r$ in (3.11).

We now prove that the derivative calculated in Proposition 3.5 is continuous in $t^*$. 

**Theorem 3.6** Let $S(t)$ be a $C^1$ curve in $CS^d_n$ defined on an interval $I$. Let $\Omega$ be an open set in $\mathbb{R}^d$ with $\sigma(S(t)) \subset \Omega$. If $f \in C^1(\Omega, \mathbb{R})$, then

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

is continuous as a function of $t^*$ on $I$.

For the proof, we will require the following lemma:

**Lemma 3.7** Let $S(t)$ be a $C^1$ curve in $CS^d_n$ defined on an interval $I$. Let $\Omega$ 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

$$\|\frac{d}{dt} F(S(t))|_{t=t^*}\| \leq C \max_{1 \leq s \leq d; x \in E} |\frac{\partial f}{\partial x^s}(x)| \quad \text{for all } t^* \in I_0,$$

where $C$ is a constant and $E$ is a convex, precompact open set with $\bar{E} \subset \Omega$. 

12
PROOF: Let \( t_0 \in I \) and fix a bounded interval \( I_0 \) around \( t_0 \) with \( I_0 \subset I \). By Theorem 2.5, the joint eigenvalues of \( S(t^*) \) are continuous on \( I_0 \). Thus, there exists an open, precompact, 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 3.5,

\[
\frac{d}{dt} F(S(t))|_{t=t^*} = U \left( \sum_{r=1}^{d} \tilde{\Gamma}^r \frac{\partial f}{\partial x^r}(D) + [Y, f(D)] \right) U^*, \tag{3.12}
\]

where \( U, D^r, \tilde{\Gamma}^r \), and \( Y \) are functions of \( t^* \) defined in Proposition 3.5, and the joint eigenvalues of \( S(t^*) \) are denoted by \( x_i \), for \( 1 \leq i \leq n \). Observe that the matrix in (3.12) can be rewritten as

\[
\begin{bmatrix}
\sum_{r=1}^{d} \Gamma^r \frac{\partial f}{\partial x^r}(D) + [Y, F(D)]
\end{bmatrix}_{ij} = \begin{cases}
\sum_{r=1}^{d} \Gamma^r_{ij} \frac{\partial f}{\partial x^r}(x_i) & \text{if } x_i = x_j \\
\Gamma^q_{ij} \frac{f(x_i) - f(x_j)}{x_i - x_j} & \text{if } x_i \neq x_j,
\end{cases} \tag{3.13}
\]

where \( q \) is such that \( x_i^q \neq x_j^q \). As shown in the proof of Theorem 2.3, the value \( \frac{\Gamma^q_{ij}}{x_i^q - x_j^q} \) is independent of \( q \) whenever \( x_i^q \neq x_j^q \).

Recall that for a given \( n \times n \) self-adjoint matrix \( A \) and an \( n \times n \) unitary matrix \( U \),

\[
\max_{ij} |(U A U^*)_{ij}| \leq n \|U A U^*\| = n \|A\| \leq n^2 \max_{ij} |A_{ij}|. \tag{3.14}
\]

It is immediate from (3.12), (3.13) and (3.14) that

\[
\left| \frac{d}{dt} F(S(t)) \right|_{t=t^*} \leq n \max \left| \sum_{r=1}^{d} \Gamma^r \frac{\partial f}{\partial x^r}(x_i) \right| + n \max \left| \Gamma^q_{ij} \frac{f(x_i) - f(x_j)}{x_i^q - x_j^q} \right|, \tag{3.15}
\]

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:

\[
|f(x_i) - f(x_j)| = |\nabla f(x^*) \cdot (x_i - x_j)| \leq \max_{x \in E} \left| \frac{\partial f}{\partial x^r}(x) \right| \sum_{r=1}^{d} |x_i^r - x_j^r|, \tag{3.16}
\]

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 \( x_i^r \neq x_j^r \),

\[
\Gamma^q_{ij} \frac{x_i^r - x_j^r}{x_i^q - x_j^q} = \Gamma^r_{ij}.
\]
It follows from (3.16) that, for each \((i, j, q)\) with \(x_i^q \neq x_j^q\),
\[
\left| \Gamma_{ij}^q \frac{f(x_i) - f(x_j)}{x_i^q - x_j^q} \right| \leq \left| \Gamma_{ij}^q \frac{1}{x_i^q - x_j^q} \right| \max_{s \in \mathcal{E}} \left| \frac{\partial f}{\partial x^s}(x) \right| \sum_{r=1}^{d} |x_i^r - x_j^r| \\
\leq \max_{s \in \mathcal{E}} \left| \frac{\partial f}{\partial x^s}(x) \right| \sum_{r=1}^{d} |\Gamma_{ij}^r| \\
\leq d n^2 \max_{s \in \mathcal{E}} \left| \frac{\partial f}{\partial x^s}(x) \right| \max_{i,j,r} \left| (S^r)'(t^*)_{ij} \right|. \tag{3.17}
\]
Likewise,
\[
\left| \sum_{r=1}^{d} \Gamma_{ij}^r \frac{\partial f}{\partial x^r}(x_i) \right| \leq d n^2 \max_{s \in \mathcal{E}} \left| \frac{\partial f}{\partial x^s}(x) \right| \max_{i,j,r} \left| (S^r)'(t^*)_{ij} \right|. \tag{3.18}
\]
Let \(M\) be a constant bounding each \(|(S^r)'(t^*)_{ij}|\) on \(I_0\) and let \(C = 2d n^3 M\). Substituting (3.17) and (3.18) into (3.15) gives
\[
\left| \frac{d}{dt} F(S(t)) \right|_{t=t^*} \leq 2d n^3 \max_{s \in \mathcal{E}} \left| \frac{\partial f}{\partial x^s}(x) \right| \max_{i,j,r} \left| (S^r)'(t^*)_{ij} \right| \\
\leq C \max_{s \in \mathcal{E}} \left| \frac{\partial f}{\partial x^s}(x) \right| \forall t^* \in I_0. \quad \Box
\]

**Proof of Theorem 3.6:**
First assume \(\Omega\) is convex. Let \(t_0 \in I\). Let \(I_0\) be the interval around \(t_0\) and \(E\) be the convex, precompact open set given in Lemma 3.7. Since \(f\) is a \(C^1\) function and \(E\) 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
\[
|\phi_k(x) - f(x)| < \frac{1}{k} \quad \text{and} \quad \left| \frac{\partial \phi_k}{\partial x^r}(x) - \frac{\partial f}{\partial x^r}(x) \right| < \frac{1}{k} \quad \forall x \in \bar{E} \quad \text{and} \quad 1 \leq r \leq d.
\]
Lemma 3.7 guarantees that, for each \(t^* \in I_0\),
\[
\left| \frac{d}{dt} \Phi_k(S(t)) \right|_{t=t^*} - \frac{d}{dt} F(S(t)) \right|_{t=t^*} \right| = \left| \frac{d}{dt} \Phi_k(S(t)) - (S^r)'(t^*)_{ij} \right| \\
\leq C \max_{s \in \mathcal{E}} \left| \frac{\partial f}{\partial x^s}(x) \right| \\
\leq \frac{C}{k},
\]
where \(C\) is a fixed constant. This implies
\[
\frac{d}{dt} \Phi_k(S(t)) \right|_{t=t^*} \text{ converges uniformly to } \frac{d}{dt} F(S(t)) \right|_{t=t^*} \text{ on } I_0.
\]
By Proposition 3.2, each \(\frac{d}{dt} \Phi_k(S(t)) \right|_{t=t^*}\) is continuous on \(I\). Since the uniform limit of continuous functions is continuous, \(\frac{d}{dt} F(S(t)) \right|_{t=t^*}\) is continuous on \(I_0\).
Now, let $\Omega$ be an arbitrary domain. Fix $t_0 \in I$ and let $I_0$ be a bounded open interval of $t_0$ with $I_0 \subset I$. Let $E \subset \mathbb{R}^d$ be an open precompact set such that $E \subset \Omega$ 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 $E \subset O \subset K \subset \Omega$ and define a $C^\infty$ bump function $b(x)$ such that

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

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 3.5 that

$$\frac{d}{dt}F(S(t))|_{t=t^*} = \frac{d}{dt}G(S(t))|_{t=t^*} \quad \forall t^* \in I_0,$$

and thus, is continuous in $I_0$. \hfill \square

Recall that $CS_n^d$ possesses a Whitney stratification with pieces $\{M_\alpha\}$. Let $\Omega$ 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 $TV := \bigcup T(M_\alpha \cap V)$. 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 $\Delta \in TS\alpha$, where $M_\alpha$ is the piece containing $S$. Let $S(t)$ be a $C^1$ curve in $CS_n^d$ such that $S(0) = S$ and $S'(0) = \Delta$. Define

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

where $U$, $D$, $\tilde{\Gamma}$, and $Y$ are defined using $S$ and $\Delta$ as in Proposition 3.5. It is easy to see that the map is well-defined and $DF(S, \cdot)$ is linear in $\Delta$. In the following theorem, let $S$ be in a piece $M_\alpha$ and let $R$ be in a piece $M_\beta$ of a Whitney stratification of $CS_n^d$.

**Theorem 3.8** Let $\Omega$ be an open set in $\mathbb{R}^d$ and $V$ be an open set in $CS_n^d$ with $\sigma(S)$ in $\Omega$ for all $S \in V$. If $f \in C^1(\Omega, \mathbb{R})$, then

$$DF : TV \to TS_n$$

is continuous.

Specifically, if $S \in V$ with $\Delta \in TS\alpha$, then given $\epsilon > 0$, there exist $\delta_1$, $\delta_2 > 0$ such that if $R \in V$ with $\Lambda \in TR\beta$, $\|S - R\| < \delta_1$, and $\|\Delta - \Lambda\| < \delta_2$, then

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

**Proof:** The result for analytic functions follows from Equation (3.9). For an arbitrary function $f$, and for $R$ and $\Lambda$ sufficiently close to $S$ and $\Delta$, bound $\|DF(R, \Lambda)\|$ in a manner similar to Lemma 3.7. The remainder of the proof is almost identical to that of Theorem 3.6 and is left as an exercise. \hfill \square
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, \((\zeta^1, \ldots, \zeta^d)\) referred to a point in \(C^d\). In this section, \((\zeta_1, \zeta_2)\) denotes a point in \(C^2\). 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 \(m \geq 1\), the curve is Lipschitz. By Theorem 2.5, there are Lipschitz curves

\[
(x_s(t), y_s(t)) \quad \text{for } 1 \leq s \leq n,
\]

defined on \(I\) representing the joint eigenvalues of \((S(t), T(t))\). Let \((S(t), T(t))\) as \((S, T)\). For \(l \in \mathbb{N}\) with \(1 \leq l \leq m\), define

\[
S^l := S^{(l)}(t) \quad \text{and} \quad T^l := T^{(l)}(t)
\]

and the set of pairs of index tuples

\[
I_l := \{(i_1, \ldots, i_k) \cup (i_{k+1}, \ldots, i_j) : i_1 + \cdots + i_j = l, i_q \in \mathbb{N} \text{ for } 1 \leq q \leq j\}.
\]

For example, \(I_2 = \{(2) \cup \emptyset, (1, 1) \cup \emptyset, (1) \cup (1), \emptyset \cup (1, 1), \emptyset \cup (2)\}\). For notational ease, define

\[
U := U(t), \quad x_s := x_s(t), \quad y_s := y_s(t) \quad \text{for } 1 \leq s \leq n.
\]

For some formulas, we will conjugate the derivatives in (4.20) by \(U^*\) and so define

\[
\Gamma^l := U^* S^{(l)} U \quad \text{and} \quad \Delta^l := U^* T^{(l)} U, \quad \text{for } 1 \leq l \leq m.
\]

We will use the integral formula given in Lemma 3.4 and simplify it by defining

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

where \(\zeta_1\) and \(\zeta_2\) are in the resolvents 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 \(k \leq j \leq m\). Fix \(k + 1\) points \(x_1, \ldots, x_{k+1}\) in \(J_1\) and \(j - k + 1\) points \(y_1, \ldots, y_{j-k+1}\) in \(J_2\). Then

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

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 product of two matrices. We will prove the following differentiability result:
Lemma 4.3 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 $1 \leq l \leq m$ and $t^* \in I$, $\frac{d^l}{dt^l} F(S, T)|_{t=t^*}$ exists and

$$\frac{d^l}{dt^l} F(S, T)|_{t=t^*} = U \left( \sum_{I_1} \sum_{I_2} \frac{l!}{i_1! \ldots i_j!} \left[ f^{[k,j-k]}(x_{s_1}, \ldots, x_{s_{k+1}}; y_{s_{k+1}}, \ldots, y_{s_{j+1}}) \right]_{s_1, s_{j+1}=1}^n \Delta_{s_{k+1}}^{i_{k+1}} \Delta_{s_{k+2}}^{i_{k+2}} \ldots \Delta_{s_{j+1}}^{i_{j+1}} \right)_{s_1, s_{j+1}=1}^n U^*,$$

where the $U$, $U^*$, $\Gamma^i$, $\Delta^j$, $x_q$ and $y_r$ are evaluated at $t^*$.

Notice that the derivative formula in Theorem 4.1 requires $f$ to be defined on pairs $(x_q, y_r)$ for $1 \leq r, q \leq n$, rather than just at the joint eigenvalues $(x_q, y_q)$ of $(S, T)$. This condition was not needed in Theorem 3.1. Before proving Theorem 4.1, we consider the case where $f$ is real-analytic and show:

Proposition 4.2 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 4.1, and $\frac{d^m}{dt^m} F(S, T)|_{t=t^*}$ is continuous as a function of $t^*$ on $I$.

The proof of Proposition 4.2 requires the following two technical lemmas:

Lemma 4.3 Let $(S, T)$ be a $C^m$ curve in $CS_n^2$ defined on an interval $I$. Let $t^* \in I$ and let $\zeta_1$ and $\zeta_2$ be elements in the resolvents of $S(t^*)$ and $T(t^*)$ respectively. Then

$$\frac{d^l}{dt^l} (R_1 R_2)|_{t=t^*} = \sum_{I_1} \frac{l!}{i_1! \ldots i_j!} R_1^{S^i_1} R_1 \ldots S^{i_2}_2 R_1 R_2 T^{i_{k+1}} R_2 \ldots T^{i_j} R_2 \quad \forall \ 1 \leq l \leq m,$n

where each $R_1$, $R_2$, $S^i$, and $T^j$ is evaluated at $t^*$.

Proof: The proof is a technical calculation using induction on $l$ and the formulas

$$\frac{d}{dt} R_1 = R_1 S^1 R_1 \quad \text{and} \quad \frac{d}{dt} R_2 = R_2 T^1 R_2. \quad \square$$

Lemma 4.4 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 $j \geq k \in \mathbb{N}$. 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 $\Omega \subset \mathbb{C}^2$ such that each $(x_q, y_r) \in \Omega$. 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, \ldots, x_{k+1}; y_1, \ldots, y_{j-k+1}) = \frac{1}{(2\pi i)^2} \int_{C_1} \int_{C_2} \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_1 d\zeta_2,$n

where $C_1$ and $C_2$ are rectifiable curves strictly enclosing $x_1, \ldots, x_{k+1}$ and $y_1, \ldots, y_{j-k+1}$ respectively, such that $C_1 \times C_2 \subset \Omega$. 17
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 of Proposition 4.2:
Use the integral formula in Lemma 3.4 to establish an integral formula for $\frac{d^m}{dt^m} F(S, T)$ similar to the first line of (3.9). Simplify the formula using Lemma 4.3. This formula implies that the derivative is continuous. Then, use Lemma 4.4 to convert the derivative into a formula involving the divided differences of $f$. The details are left as an exercise. □

Proof of Theorem 4.1:
The result follows via induction on $l$, and the base case is covered by Theorem 3.1. 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 $1 \leq q, r \leq n$. Find a constant $C$ such that for $t$ near $t^*$,
\[
\| \frac{d^{l+1}}{dt^{l+1}} F(S, T) - \frac{d^{l+1}}{dt^{l+1}} P(S, T) \| \leq C \max \{ |f - p|^{[k,j,k]}(x_{s_1}, \ldots, x_{s_{k+1}}; y_{s_{k+1}}, \ldots, y_{s_{j+1}}) \},
\]
where the joint eigenvalues of $(S, T)$ are given by $(x_q, y_q)$ and the maximum is over $(k, j)$ with $k \leq j < l \in \mathbb{N}$ and sets $\{(s_1, \ldots, s_{k+1}) \cup (s_{k+1}, \ldots, s_{j+1}) : 1 \leq s_1 \ldots s_{j+1} \leq n \}$. As in Theorem 3.1, apply the multivariate Mean Value theorem to each $(f - p)^{[k,j,k]}$ and use the Lipschitz property of the eigenvalues to conclude
\\
\[ \frac{d^l}{dt^l} F(S, T)|_{t=t^*} \text{ exists and equals } \frac{d^l}{dt^l} P(S, T)|_{t=t^*}. \]

The details are left as an exercise. □

We now show that the formula in Theorem 4.1 is continuous.

**Theorem 4.5** 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 $1 \leq l \leq m$,
\[ \frac{d^l}{dt^l} F(S, T)|_{t=t^*} \text{ is continuous as a function of } t^* \text{ on } I. \]

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 4.6** 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 $k \leq j \leq m$. Let $x_1, \ldots, x_{k+1} \in J_1$ and $y_1, \ldots, y_{j-k+1} \in J_2$ 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
\[ f^{[k,j-k]}(x_1, \ldots, x_{k+1}; y_1, \ldots, y_{j-k+1}) = \frac{f^{(k,j-k)}(x^*, y^*)}{k!(j-k)!}. \]
Proof of Theorem 4.5:
For $l < m$, the result follows from Theorem 4.1, which implies that $\frac{d^l}{dt^l} F(S, T)$ is differentiable and, hence, continuous.

For $l = m$, fix $t_0 \in I$. As in Lemma 3.7, find a constant $C$ and an open precompact convex set $J$ with $\bar{J} \subset J_1 \times J_2$ such that, for all $g \in C^m(J_1 \times J_2, \mathbb{R})$ and $t^* \near t_0$,

$$\left| \frac{d^m}{dt^m} G(S, T)_{|t = t^*} \right| \leq C \max_{\{i, j ; (x, y) \in J\}} |f^{(k, j-k)}(x, y)|,$$

where $0 \leq k \leq j \leq m$. The estimates for this bound require Lemma 4.6. Then, approximate $f$ to $m$th order uniformly on $\bar{J}$ by analytic functions $\{\phi_r\}$ and show

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

in a neighborhood of $t_0$. The result then follows from Proposition 4.2. \qed

5 Applications

The formulas in Proposition 3.5 and Theorem 4.1 can be used to analyze monotonicity and convexity of matrix functions. A function $F : S_n \to S_n$ is matrix monotone if

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

For $F$ continuously differentiable, an equivalent condition is

$$\frac{d}{dt} F(S(t))_{|t = t^*} \geq 0 \text{ whenever } S'(t^*) \geq 0, \quad \forall C^1 S(t) \in S_n. \quad (5.20)$$

The local monotonicity condition in (5.20) extends to multivariate matrix functions: the only adjustment is that $S(t)$ is in $CS^d_n$. In [1], Agler, McCarthy, and Young characterized such locally monotone matrix functions on $CS^d_n$ using a special case of Theorem 3.1 and Proposition 3.5. 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 $CS^d_n$ and show that the formula is continuous.

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

$$F(\lambda A + (1 - \lambda)B) \leq \lambda F(A) + (1 - \lambda)F(B) \quad \forall A, B \in S_n, \quad \lambda \in [0, 1]. \quad (5.21)$$

This condition extends to multivariate matrix functions with an additional restriction on the pairs $A, B$ in $CS^d_n$: we also require $\lambda A + (1 - \lambda)B \in CS^d_n$ for $\lambda \in [0, 1]$. Given such $A, B$, define $S(t)$ on $[0, 1]$ by

$$S^r(t) := tA^r + (1 - t)B^r \quad \forall 1 \leq r \leq d.$$
If $F$ is twice continuously differentiable, it can be shown that (5.21) is equivalent to
\[ \frac{d^2}{dt^2} F(S(t))|_{t=t^*} \geq 0 \quad \forall \text{ such } S(t), \ t^* \in [0, 1]. \quad (5.22) \]
Assume $F$ was defined using a real function $f$ as in (1.1). For $d = 2$, Theorem 4.1 tells us that, up to conjugation by a unitary matrix $U$,
\[
\left[ \frac{d^2}{dt^2} F(S(t))|_{t=t^*} \right]_{ij} = 2 \sum_{k=1}^{n} f^{[2,0]}(x_i, x_k, x_j; y_i) \Gamma_{ik} \Gamma_{kj} + f^{[1,1]}(x_i, x_k; y_k, y_j) \Gamma_{ik} \Delta_{kj} + f^{[0,2]}(x_i; y_i, y_k, y_j) \Delta_{ik} \Delta_{kj},
\]
where \( \{(x_i, y_i) : 1 \leq i \leq n\} \) are the joint eigenvalues of \( t^* A + (1-t^*) B \) and
\[ \Gamma := U^* (A_1 - B_1) U \ \text{ and } \ \Delta := U^* (A_2 - B_2) U. \]
This formula can be simplified using the relationship between $\Gamma$ and $\Delta$ discussed in Theorem 2.3. Specifically, we know
\[ (x_i - x_j) \Delta_{ij} = (y_i - y_j) \Gamma_{ij} \quad \forall \ 1 \leq i, j \leq n. \]
Thus, this formula gives a characterization of convex matrix functions on $CS_n^2$.

References

[1] J. Agler, J. McCarthy, and N. J. Young. Operator Monotone Functions and Loewner Functions of Several Variables. 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, 2000.
[4] W.F. Donoghue. Monotone matrix functions and analytic continuation. 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] T. Kato. Perturbation Theory for Linear Operators. Springer-Verlang, Berlin, 1966.
[8] F. Rellich. Störungstheorie der Spektralzerlegung, I. Ann. of Math., 113:600–619, 1937.
[9] F. Rellich. Störungstheorie der Spektralzerlegung, II. Ann. of Math., 113:677–685, 1937.
[10] H. Whitney. Tangents to an analytic variety. Ann. of Math., 81:496–549, 1965.