PERTURBATION THEORY FOR HERMITIAN MATRIX-FUNCTIONS
BASED ON VECTOR-FIELDS

MARCUS CARLSSON

ABSTRACT. We consider “spectral” matrix-functions for Hermitian matrices, where the novelty is that the function applied to the spectrum is allowed to be a vector-field rather than a scalar function. We prove first order approximation formulas, generalizing the classical Daleskii-Krein theorem, as well as Lipschitz estimates. Proofs are new also in the scalar case, and lead to a shorter treatment than previous approaches.

1. Introduction
Let $H_n$ denote the set of $n \times n$ Hermitian self-adjoint matrices (possibly with complex off-diagonal entries), treated as a real vector space. By the spectral theorem, each matrix $A \in H_n$ can be decomposed as $A = U_A \Lambda_\alpha U_A^*$ where $\alpha$ are the eigenvalues, $\Lambda_\alpha$ the diagonal matrix with $\alpha$ on the diagonal, and $U_A$ is a unitary matrix whose columns are the eigenvectors. Given a function $f : \mathbb{R} \to \mathbb{R}$ the classical “functional calculus”, also known as “matrix function”, is defined as

$$f(A) = U_A \Lambda_{f(\alpha)} U_A^*$$

and is a natural extension of applying a polynomial $p(z) = \sum_{j=0}^k a_j z^j$ to $A$ via the rule

$$p(A) = \sum_{j=0}^k a_j A^j.$$ 

When discussing derivatives it is better to have separate notation for $f$ as a function (on $\mathbb{R}^n$) and as a function on $H_n$, so we will henceforth write $\mathcal{L}_f(A)$ for $f(A)$ as defined in (1.1). The first order perturbation of this calculus is completely understood by the so called Daleskii-Krein theorem, which shows that $\mathcal{L}_f$ is Fréchet differentiable and gives a concrete formula for its Fréchet derivative $\mathcal{L}_f'$, so that

$$\mathcal{L}_f(A + E) = \mathcal{L}_f(A) + \mathcal{L}_f'(E) + O(\|E\|^2).$$

(1.2)

We introduce a functional calculus for certain vector-fields $F$, and extend the result (1.2) to this case. Let $\mathbb{R}_{\geq}^n \subseteq \mathbb{R}^n$ be the set of non-increasing sequences, and let $F : \mathbb{R}_{\geq}^n \to \mathbb{R}^n$ be block-constant, i.e. such that $F_i(x) = F_j(x)$ whenever $x_i = x_j$. For block-constant vector-fields the functional calculus (1.1) is readily extended by defining

$$\mathcal{L}_F(A) = U_A \Lambda_{F(\alpha)} U_A^*.$$ 

(1.3)

Setting $F(x) = (f(x_1), \ldots, f(x_n))$ we get $\mathcal{L}_F = \mathcal{L}_f$. The extension (1.3) can be motivated in various ways, for instance it arises as derivatives of so called spectral functions [15]. More precisely, if $f : \mathbb{R}^n \to \mathbb{R}$ is a symmetric function (i.e. invariant under perturbations of...
its coordinates), then the functional \( A \mapsto f(\alpha) = f(eig(A)) \) is of interest e.g. in matrix optimization problems. We refer to [15] for more information, where it is also shown that

\[
\nabla f(eig(A)) = U_A \Lambda \nabla f(\alpha) U_A^* = \mathcal{L}_f(A),
\]

so the functional calculus (1.3) arise naturally when studying the gradient of \( f(eig(A)) \). In [14] a formula for the second derivative is given, which hence is a special case of the version of (1.2) that we will prove here, (in Theorem 2.7). The proof in [14] was further improved in [17], where also all higher order derivatives were computed. We remark that the proof given here does not reduce to the methods of [14, 17] for this particular case, neither is it reminiscent of any classical proof of the Daleckii-Krein theorem, which usually is based on polynomial approximation arguments [4, 7, 10, 16].

However, the main motivation behind this work is not to give a new proof of an established fact, but to provide an extension of (1.2) to a class of vector-fields much broader than just conservative vector-fields, (i.e. those arising as \( F = \nabla f \) for some scalar symmetric function \( f \)). This in turn is motivated by the fact that the new functional calculus we introduce is starting to appear more frequently as proximal operators in matrix optimization algorithms. Based on our extension of (1.2), we then move on to establish Lipschitz estimates. More precisely we show that

\[
\| \mathcal{L}_F(A) - \mathcal{L}_F(B) \|_2 \leq \| F \|_{Lip} \| A - B \|_2,
\]

which is the best estimate one could hope for (in terms of the Frobenius norm), shown in Section 3.

Interesting examples of vector-fields \( F \) arise as proximal operators in low-rank approximation theory [1, 9, 13]. It is our hope that the theory provided here can be used for faster evaluation of such operators in iterative algorithms, but this has to be investigated elsewhere. In this case, the Lipschitz-estimate is however of lesser use, since proximal operators are known to be firmly-nonexpansive (see e.g. Theorem 21.2 and Corollary 23.8 in [3]), and the estimate (1.4) usually only gives nonexpansiveness.

2. Perturbation theory for \( \mathcal{L}_F \)

Let \( A = U_A \Lambda_\alpha U_A^* \) be the spectral decomposition of a Hermitian matrix \( A \in \mathcal{H}_n \), where \( \Lambda_\alpha \) thus is diagonal with diagonal elements \( \alpha \) ordered non-increasingly. We will also use the notation \( \text{diag}(\alpha) \) for \( \Lambda_\alpha \). Let \( \mathbb{R}^n_\geq \) be the subset of \( \mathbb{R}^n \) of non-increasing sequences. Given any function on \( F : \mathbb{R}^n_\geq \to \mathbb{R}^n \) we define

\[
\mathcal{L}_F(A) = U_A \text{diag}(F(\alpha)) U_A^*.
\]

For this to be well defined it is necessary that \( F(\alpha) \) is constant for indices where \( \alpha \) is constant, and such functions will be referred to as block-constant at \( \alpha \) (following [17]). The logic behind this terminology is that \( \alpha \) determines certain blocks of subindices where \( \alpha_i = \alpha_j \). A function which is block constant for all \( \alpha \) in its domain of definition will simply be called block-constant. Given \( f : \mathbb{R} \to \mathbb{R} \) the traditional functional calculus is retrieved by setting \( F(x) = (f(x_1), \ldots, f(x_n)) \), which obviously is block-constant.

We now analyze how perturbations affect the functional calculus \( \mathcal{L}_F \), where \( F \) is any given function on \( \mathbb{R}^n_\geq \), i.e. we are interested in \( \mathcal{L}_F(A + E) \) for small \( E \). More precisely we shall analyze the Fréchet derivative of this map. We of course assume that \( F \) is block constant at \( \alpha \), but not necessarily in the whole of \( \mathbb{R}^n_\geq \). For example, \( F \) could be \( F(x) = (1, 0, \ldots, 0) \),
and in this case $\mathcal{L}_F(A)$ equals the orthogonal projection onto the subspace spanned by the first eigenvector, which is well defined as long as $\alpha_1$ has multiplicity 1. Note that

$$\mathcal{L}_F(A + E) - \mathcal{L}_F(A) = U_A \left( \mathcal{L}_F(\Lambda_\alpha + U_A^*EU) - \mathcal{L}_F(\Lambda_\alpha) \right) U_A^*,$$

by which it follows that it suffices to compute the derivative in the case when $A$ is diagonal. The matrix $U_A^*EU_A$ will henceforth be denoted $\tilde{E}$ (in analogy with [6, 5]).

2.1. **Point-symmetric functions and vector-fields.** We first introduce the function class for which $\mathcal{L}_F$ is Fréchet differentiable. Given a vector $x \in \mathbb{R}^n$, we let $\text{per}(x)$ be the set of permutation matrices $\Pi$ such that $\Pi x = x$. Any vector-field $F$ that is differentiable at a point $x$ gives rise to a (matrix) derivative $F'|_x$, and the property that

$$\Pi^* F'|_x \Pi = F'|_x, \quad \Pi \in \text{per}(x),$$

will turn out to be crucial. To simplify verification of this fact, we introduce “point-symmetric functions”, where the terminology is adopted from [17].

Let $\text{sort}(x)$ be the set of permutation matrices $\Sigma$ such that $\Sigma x \in \mathbb{R}^n_{\geq}$. Given a fixed $\Sigma \in \text{sort}(x)$, note that

$$\text{sort}(x) = \Sigma \cdot \text{per}(x).$$

If $F : \mathbb{R}^n_{\geq} \to \mathbb{R}^n$ is block-constant, we can therefore uniquely extend it to a function on $\mathbb{R}^n$ by setting

$$F^{\text{ext}}(x) = \Sigma^* F(\Sigma x).$$

To see this, we use (2.3) and note that $\Sigma^*$ is the inverse of $\Sigma$. As a silly example, consider the vector-field $F : \mathbb{R}^2_{\geq} \to \mathbb{R}^2$ defined by $F(x_1, x_2) = (x_1 - x_2, 0)$. The the extension to $x_1 < x_2$ via (2.4) then becomes $F^{\text{ext}}(x_1, x_2) = (0, x_2 - x_1)$.

Following [17], we denote by $T^{k,n}$ the set of $k$-tensors on $\mathbb{R}^n$. The set $T^{0,n}$ is defined as $\mathbb{R}$, $T^{1,n}$ is readily identified with $\mathbb{R}^n$ and $T^{2,n}$ with the set of $n \times n$-matrices over $\mathbb{R}$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is thus identified with a $T^{0,n}$-valued map, a vector-field with a $T^{1,n}$-valued map and so on.

A $k$-tensor valued map $f : \mathbb{R}^n \to T^{k,n}$ is called point-symmetric if

$$f(x)[h_1, \ldots, h_k] = f(\Pi x)[\Pi h_1, \ldots, \Pi h_k]$$

for all permutation-matrices $\Pi$. Note in particular that a point symmetric map satisfies $f(x)[h_1, \ldots, h_k] = f(x)[\Pi h_1, \ldots, \Pi h_k]$ whenever $\Pi \in \text{per}(x)$. This means that if $k = 0$, point symmetric functions coincides with “symmetric functions” as defined e.g. in [15], and it is easily seen that a vector-field $F : \mathbb{R}^n_{\geq} \to \mathbb{R}^n$ is block-constant if and only if $F^{\text{ext}}$ is point symmetric.

A key feature of point symmetric maps is that the property is invariant under differentiation. More precisely, a Fréchet derivative of a $T^{k,n}$-valued map naturally identifies with a new $T^{k+1,n}$-valued map, and if the former is point-symmetric then so is the latter. This is easy to show, we refer to [17] for the details. In particular, if we identify a tensor $\mathcal{T} \in T^{2,n}$ with an $n \times n$-matrix $M$ as usual (i.e. by the formula $\mathcal{T}[h_1, h_2] = h_2^T M h_1$) then $\mathcal{T}(\Pi h_1, \Pi h_2)$ translates to the matrix $\Pi^T M \Pi$. Summing up, we have proved the following:

**Proposition 2.1.** If a vector-field $F$ is block-constant in a neighborhood of some point $x$, then $F^{\text{ext}}$ is point-symmetric there. In particular, if $F^{\text{ext}}$ is $C^1$ at $x$, then $F'|_x$ satisfies (2.2).
2.2. \textbf{The partial derivatives of $\mathcal{L}_F$.} We can now start to compute Gateaux derivatives of $\mathcal{L}_F$ at $\Lambda_\alpha$ for certain directions $E$, i.e. the derivative of $\mathbb{R} \ni h \mapsto \mathcal{L}_F(\Lambda_\alpha + hE)$, which we denote by $\frac{\partial \mathcal{L}_F}{\partial E}(\Lambda_\alpha)$. Let $E_{i,j}$ be the matrix which is 1 at indices $i, j$ and has zeroes elsewhere, and let $(e_i)_{i=1}^n$ denote the canonical basis in $\mathbb{C}^n$.

\textbf{Lemma 2.3.} If $F$ is $C^1$ point-symmetric at $\alpha$, then
\[
\frac{\partial \mathcal{L}_F}{\partial E_{i,i}}(\Lambda_\alpha) = \text{diag}(F'|_{\alpha}e_i).
\]

\textit{Proof.} Let $e \in \mathbb{R}^n$ be given and let $\Pi \in \text{per}(\alpha)$ be such that $\alpha + h\Pi e$ is non-increasing, for all small enough $h > 0$ (we assume for the remainder that $h$ is arbitrary but positive). Note that this means that $\Pi$ only permutes within each subblock $S_i$, in which it orders the elements of $h$ non-increasingly. If $E = \text{diag}(e)$ then the spectral decomposition of $\Lambda_\alpha + hE$ equals $\Pi^* \text{diag}(\alpha + h\Pi e) \Pi$, so
\[
\frac{1}{h} \mathcal{L}_F(\Lambda_\alpha + hE) - \mathcal{L}_F(\Lambda_\alpha) = \Pi^* \text{diag}(F(\alpha + h\Pi e) - F(\alpha)) \Pi.
\]
Upon taking a limit as $h \to 0^+$ we obtain $\Pi^* \text{diag}(F'|_{\alpha} \Pi e) = \text{diag}(\Pi^* F'|_{\alpha} \Pi e)$. In order for $\mathcal{L}_F$ to be Fréchet differentiable this expression needs to be linear in $E$, which happens if and only if \textbf{(2.2)} holds for all permutations $\Pi \in \text{Per}(\alpha)$, which indeed is the case by Proposition \textbf{2.1}. Thus we can remove the restriction that $h > 0$ and we conclude that
\[
\lim_{h \to 0} \frac{1}{h} \mathcal{L}_F(\Lambda_\alpha + hE) - \mathcal{L}_F(\Lambda_\alpha) = F'|_{\alpha} e.
\]
In particular, if $E = E_{i,i}$ we get the statement in the lemma. \hfill \Box

We now consider the off diagonal elements. We will consider $E_{i,j} + E_{j,i}$ and $iE_{i,j} - jE_{j,i}$, which combined with $E_{i,i}$ provides a natural basis for $\mathcal{H}_\alpha$ as a real vector space. A perturbation in the direction of $E_{i,j}$ changes the eigenvalues $\alpha_i, \alpha_j$ in a nontrivial way (see e.g. [6]), but we basically consider a $2 \times 2$ matrix problem since the others are unaffected by the perturbation. Given a $C^1$ point-symmetric vector-field $F$ we first introduce the notation
\[
[F, \alpha](i, j) = \begin{cases} \frac{F_i(\alpha) - F_j(\alpha)}{\alpha_i - \alpha_j} & \alpha_i \neq \alpha_j \\ \frac{\partial_i F_i(\alpha) - \partial_j F_j(\alpha)}{\partial_i F_i(\alpha)} & \alpha_i = \alpha_j, \ i \neq j \\ i = j \end{cases}
\]
Note that there seems to be a lack of symmetry in \(i\) and \(j\) on the second line, but this is not so due to the assumption that \(F\) is \(C^1\) point-symmetric. In terms of the numbers introduced before Definition 2.2 we have, for \(i \in S_i\) and \(j \in S_j\) that \([F,\alpha](i,j) = \frac{s_i-s_j}{\alpha_i-\alpha_j}\) when \(i \neq j\), \([F,\alpha](i,i) = r_i\) when \(i \neq j\) belong to the same block, whereas \([F,\alpha](i,j) = r_i + t_{i,i}\) on the diagonal \(i = j\).

**Lemma 2.4.** Let \(\tau \in \mathbb{C}\) be unimodular. If \(F\) is \(C^1\) point-symmetric at \(\alpha\) and \(i \neq j\), then

\[
\frac{\partial L_F}{\partial (\tau E_{i,j} + \bar{\tau} E_{j,i})}(\Lambda_\alpha) = [F,\alpha] \circ (\tau E_{i,j} + \bar{\tau} E_{j,i}).
\]

**Proof.** We may assume that \(i = 1\), \(j = 2\) and that \(\Lambda_\alpha\) is a \(2 \times 2\) matrix, as noted earlier. Then \(\Lambda_\alpha + h(\tau E_{1,2} + \bar{\tau} E_{2,1}) = \left(\begin{array}{cc} \alpha_1 & \tau h \\ \tau h & \alpha_2 \end{array}\right)\). If \(\alpha_1 \neq \alpha_2\), some simple calculations give

\[
\left(\begin{array}{cc} \alpha_1 & \tau h \\ \tau h & \alpha_2 \end{array}\right) = \left(\begin{array}{cc} \frac{\tau h}{\alpha_1 - \alpha_2} - \frac{\tau h}{1} \end{array}\right) + O(h^2) \left(\begin{array}{cc} \alpha_1 + O(h^2) & \left(-\frac{1}{\alpha_1 - \alpha_2} \frac{\tau h}{1}\right) + O(h^2) \right)
\]

which yields that

\[
\frac{1}{h} \left(L_F(\left(\begin{array}{cc} \alpha_1 & \tau h \\ \tau h & \alpha_2 \end{array}\right)) - L_F(\left(\begin{array}{cc} \alpha_1 & 0 \\ 0 & \alpha_2 \end{array}\right))\right) = \frac{1}{h} \left(\begin{array}{cc} \frac{\tau h}{\alpha_1 - \alpha_2} - \frac{\tau h}{1} \end{array}\right) \left(F_1(\alpha) \begin{array}{cc} 0 \\ 0 \end{array} F_2(\alpha) \right) \left(-\frac{1}{\alpha_1 - \alpha_2} \frac{\tau h}{1}\right) + O(h^2) = \left(\begin{array}{cc} 0 & \left[-\frac{1}{\alpha_1 - \alpha_2} \frac{\tau h}{1}\right] \end{array}\right) + O(h),
\]

as desired. If \(\alpha_1 = \alpha_2\) we instead get

\[
\left(\begin{array}{cc} \alpha_1 & \tau h \\ \tau h & \alpha_1 \end{array}\right) = \frac{1}{2} \left(\begin{array}{cc} 1 & \tau \\ \tau & -1 \end{array}\right) \left(\begin{array}{cc} \alpha_1 + h & 0 \\ 0 & \alpha_1 - h \end{array}\right) \left(\begin{array}{cc} 1 & \tau \\ \tau & -1 \end{array}\right)
\]

(without ordo terms). Since \(F\) is assumed to be \(C^1\) point-symmetric at \(\alpha\), there are values \(s, r, t\) such that \(F_1(\alpha_1 + h, \alpha_1 - h) = s + (r + t, (h, -h)) + o(h) = s + rh + o(h)\), and similarly \(F_2(\alpha_1 + h, \alpha_1 - h) = s - rh + o(h)\). Thus

\[
L_F(\left(\begin{array}{cc} \alpha_1 & \tau h \\ \tau h & \alpha_1 \end{array}\right)) = \frac{1}{2} \left(\begin{array}{cc} 1 & \tau \\ \tau & -1 \end{array}\right) \left(F_1(\alpha_1 + h, \alpha_1 - h) \begin{array}{cc} 0 \\ 0 \end{array} F_2(\alpha_1 + h, \alpha_1 - h) \right) \left(\begin{array}{cc} 1 & \tau \\ \tau & -1 \end{array}\right) = \frac{1}{2} \left(\begin{array}{cc} (s + rh) + (s - rh) \tau \\ (s + rh) - (s - rh) \tau \end{array}\right) + o(h) = \left(\begin{array}{cc} s & rh \tau \\ rh \tau & s \end{array}\right) + o(h).
\]

It follows that

\[
L_F(\left(\begin{array}{cc} \alpha_1 & \tau h \\ \tau h & \alpha_1 \end{array}\right)) - L_F(\left(\begin{array}{cc} \alpha_1 & 0 \\ 0 & \alpha_1 \end{array}\right)) = h \left(\begin{array}{cc} 0 & r \tau \\ r \tau & 0 \end{array}\right) + o(h)
\]

from which the desired conclusion easily follows. 

\[\square\]
eigenvalues in \( \Lambda_\alpha \), see e.g. Ch. VII of [4] or [8] or [6]. With this in mind, it is a bit surprising that \( \mathcal{L}_F \) (which inherently relies on the eigenvectors) is differentiable, as we shall show next.

2.3. Fréchet differentiability of \( \mathcal{L}_F \). We recall that a function \( F \) on \( \mathcal{H}_n \) is Fréchet differentiable at \( A \in \mathcal{H}_n \) if there exists a real-linear operator \( \mathcal{F}' : \mathcal{H}_n \to \mathcal{H}_n \) such that

\[
\mathcal{F}(A + E) = \mathcal{F}(A) + \mathcal{F}'(E) + o(\|E\|).
\]

Since all norms on finite dimensional spaces are equivalent, it is not important to specify which norm we use, but often we work with the Frobenius-norm \( \|E\|_2 \) for simplicity. The expression \( \|E\| \) with no subindex will be reserved for the operator norm.

**Lemma 2.5.** Let \( A \in \mathcal{H}_n \) be given and let \( F, G : \mathbb{R}^n \to \mathbb{C}^n \) be block-constant at \( \alpha \) such that \( F(x) = G(x) + o(\|x - \alpha\|_2) \). Then \( \mathcal{L}_F \) is Fréchet differentiable at \( A \) if and only if \( \mathcal{L}_G \) is, and the derivatives coincide.

**Proof.** Let \( \Lambda_\alpha + E = U \Lambda \xi U^* \) be the spectral decomposition of \( \Lambda_\alpha + E \). Since \( \mathcal{L}_F(A) = \mathcal{L}_G(A) \) we get

\[
\mathcal{L}_F(A + E) - \mathcal{L}_F(A) = U(\text{diag}(F(\xi) - G(\xi)))U^* + \mathcal{L}_G(A + E) - \mathcal{L}_G(A) = O(F(\xi) - G(\xi)) + \mathcal{L}_G(A + E) - \mathcal{L}_G(A) = o(\|\xi - \alpha\|_2) + \mathcal{L}_G(A + E) - \mathcal{L}_G(A) = o(\|E\|_2) + \mathcal{L}_G(A + E) - \mathcal{L}_G(A)
\]

where the last identity follows from the Hoffman-Wielandt inequality. From this it is clear that Fréchet differentiability of \( G \) implies that of \( F \) and vice versa.

We remind the reader that \( \tilde{\alpha}_1, \ldots, \tilde{\alpha}_k \) is a (non-increasing) enumeration of the distinct eigenvalues of \( A \), and that \( S_j = \{i : \alpha_i = \tilde{\alpha}_j\} \). It is a basic result that \( A \) can be written

\[
A = \sum_{j=1}^k \tilde{\alpha}_j P_j
\]

where \( P_j \) is the orthogonal projection onto the eigenspace corresponding to \( \tilde{\alpha}_j \). Also, if \( \Gamma_j \) is a suitably chosen circle around \( \tilde{\alpha}_j \), then

\[
P_j(A) = \int_{\Gamma_j} (zI - A)^{-1} \frac{dz}{2\pi i}.
\]

**Proposition 2.6.** If \( F \) is a \( C^1 \) point-symmetric vector-field, then \( \mathcal{L}_F \) is Fréchet differentiable.

**Proof.** By the lemma it suffices to prove the above statement for

\[
G(x) = F(\alpha) + F'_\alpha(x - \alpha).
\]

Let \( B \in \mathcal{H}_n \) denote a matrix in the vicinity of \( A \) and denote its eigenvalues by \( \beta \). We then have

\[
\mathcal{L}_G(B) = U_B(\text{diag}(F(\alpha)) + \text{diag}(F'_\alpha(\beta - \alpha)))U_B^* = \sum_{i=1}^k U_B(\text{diag}(F(\alpha)) + \text{diag}(F'_\alpha(\beta - \alpha)))U_B^* P_i(B)
\]
where diagonal values outside of the block $S_i$ are irrelevant for the $i$:th term. Therefore the computation can be continued as follows

$$
\sum_{i=1}^{k} U_B \left( s_i I + r_i \text{diag}(\beta - \alpha) + \left( \sum_{j=1}^{k} t_{i,j} \sum_{l \in S_j} (\beta_l - \tilde{\alpha}_j) \right) I \right) U_B^* P_i(B) = \sum_{i=1}^{k} r_i BP_i(B) + \sum_{i=1}^{k} \left( s_i - r_i \tilde{\alpha}_i + \left( \sum_{j=1}^{k} t_{i,j} \sum_{l \in S_j} (\beta_l - \tilde{\alpha}_j) \right) \right) P_i(B).
$$

Now note that we can write $BP_i(B) = \int_{\Gamma_j} \frac{z - \tilde{\alpha}_j}{(zI - B) \det(zI - B)} \frac{dz}{2\pi i}$ and that

$$
\sum_{l \in S_j} (\beta_l - \tilde{\alpha}_j) = \int_{\Gamma_j} \frac{(z - \tilde{\alpha}_j) \frac{d}{dz} \det(zI - B)}{\det(zI - B)} \frac{dz}{2\pi i},
$$

for $B$ sufficiently close to $A$. Summing up we have shown that

$$\mathcal{L}_G(B) = \sum_{i=1}^{k} \int_{\Gamma_j} \frac{r_i z}{zI - B} \frac{dz}{2\pi i} + \int_{\Gamma_j} \frac{(z - \tilde{\alpha}_j) \frac{d}{dz} \det(zI - B)}{\det(zI - B)} \frac{dz}{2\pi i}.
$$

This calculation was made under the assumption that $B \in \mathcal{H}_n$ but the final expression is actually a valid expression for all matrices $B$, and as such it is holomorphic in each variable $B_{i,j}$ separately. If we identify the space of $n \times n$-matrices with $\mathbb{C}^{n^2}$, the function is then holomorphic and it is well known that this implies that the function is $C^\infty$, see e.g. Theorem 2.2.1 and Corollary 2.2.2 in [12]. In particular it is differentiable, and hence the expression in (2.8) is also Fréchet differentiable on the space of $n \times n$-matrices. Since the restriction to the subspace $\mathcal{H}_n$ equals $\mathcal{L}_G(B)$ for $B$ in a neighborhood of $A$, it follows that $\mathcal{L}_G$ is Fréchet differentiable at $A$. 

We now come to the first main theorem, a generalization of the so called Daleskii-Krein theorem [4, 7, 10, 16] to the matrix-functions based on vector-fields introduced in this paper. Given any matrix $M$ we use the notation $M^\circ = M - M \circ I$, i.e. $M^\circ$ coincides with $M$ off the diagonal, whereas it is 0 on the diagonal.

**Theorem 2.7.** Let $A = U \Lambda_\alpha U^*$ be given and let $F$ be $C^1$ point-symmetric vector-field at $\alpha$. Given $E$ in $\mathcal{H}_n$, set $\tilde{E} = U^* EU$ and let $\hat{e}$ be the vector with the diagonal elements of $\tilde{E}$. Then

$$\mathcal{L}'_F(E) = U \left( [F, \alpha] \circ \hat{E} + \text{diag}(F^\circ(\alpha)^\circ \hat{e}) \right) U^*.
$$

If $f$ is real valued and $F(x) = (f(x_1), \ldots, f(x_n))$, then $\mathcal{L}_F$ reduces to the traditional matrix functions (functional calculus) and the below theorem implies the Daleskii-Krein theorem, which we elaborate more on in Section 2.4. If $F = \nabla f$ for some symmetric function $f : \mathbb{R}^n \to \mathbb{R}$, the theorem provides a formula for the second order term in a Taylor type expansion of $f(eig(A + E))$, and the above theorem reduces to Theorem 3.2 of [14]. The proof given here is more general and also shorter, it seems. Formulas for all possible orders was subsequently found in [17], but we will not pursue a similar quest here.
Proof. \( \mathcal{L}_F \) is Fréchet differentiable by the above proposition. Write
\[
\hat{E} = \sum_i \frac{\partial_i E_{i,i}}{i} + \sum_{j>i} b_{i,j}(E_{i,j} + E_{j,i}) + \sum_{j>i} c_{i,j}i(E_{i,j} - E_{j,i})
\]
and note that the above matrices provide an orthogonal basis for \( \mathcal{H}_n \) as a real vector space. Using Lemma 2.3 for the first sum and 2.4 for the latter two (with \( \tau \) equal to 1 and \( i \) respectively), the formula \( \mathcal{L}'_F(E) = U\left([F,\alpha]^* \circ \hat{E} + \text{diag}_{F'|_{\alpha^*}}\right)U^* \) follows from basic multivariable calculus, and clearly
\[
[F,\alpha]^* \circ \hat{E} + \text{diag}_{F'|_{\alpha^*}} = [F,\alpha] \circ \hat{E} + \text{diag}_{(F'|_{\alpha})^*\hat{e}}.
\]
\( \square \)

2.4. The scalar case. We now specialize to the scalar valued functional calculus, i.e. when \( F(x) = (f(x_1), \ldots, f(x_n)) \) where \( f : \mathbb{R} \to \mathbb{R} \) (or some subset including the actual spectrum). It is clear that any differentiable \( f \) yields a \( C^1 \) point-symmetric \( F \). Given \( \alpha \in \mathbb{R}_n^\ast \), the matrix \( [F,\alpha] \) will then be written \( [f,\alpha] \) and simplifies to
\[
[f,\alpha](i,j) = \begin{cases} \frac{f(\alpha_i) - f(\alpha_j)}{\alpha_i - \alpha_j} & \alpha_i \neq \alpha_j \\ \partial_{i} f(\alpha_i) & \alpha_i = \alpha_j \end{cases}
\]
With this notation Theorem 2.7 implies the Daleskii-Krein theorem;

Corollary 2.8. Let \( f \) be a \( C^1 \)-function and \( \Lambda = U_A \Lambda_n U_A^\ast \in \mathcal{H}_n \). Given \( E \in \mathcal{H}_n \) set \( \check{E} = U_A E U_A \). Then
\[
\mathcal{L}'_f(E) = U\left([f,\lambda] \circ \hat{E}\right)U^*.
\]
If \( \Lambda \) has a non-trivial kernel, this is not enough to analyze the perturbation of e.g. \( \sqrt{\Lambda + \check{E}} \), which turns out to be a rather delicate issue, we refer to the adjacent paper [5] for more on this particular case.

3. Lipschitz continuity of \( \mathcal{L}_F \)

We can now prove that \( \mathcal{L}_F \) is Lipschitz continuous (with respect to the Frobenius norm) with the same constant as \( F \). We recall that for \( F : \mathbb{R}^n \to \mathbb{R}^n \) we have that
\[
\|F\|_{\text{Lip}} = \sup_{x \neq y} \frac{\|F(x) - F(y)\|_2}{\|x - y\|_2},
\]
which for \( C^1 \)-functions equals \( \sup_x \|F'|_x\| \) where the last norm refers to the operator norm
\[
\|F'|_x\| = \sup_{y : \|y\| = 1} \frac{\|F'\|_x y\|_2}{\|y\|_2}.
\]

Lemma 3.1. Let \( F \) be \( C^1 \) point-symmetric at a particular point \( \alpha \). Then \( |[F,\alpha](i,j)| \leq \|F\|_{\text{Lip}} \)

Proof. The statement is obvious if \( i = j \). Suppose that \( i > j \) and that they belong to the same block. Then \( [F,\alpha](i,j) = \partial_i F_i - \partial_j F_j, \partial_i F_i = \partial_j F_j \) and \( \partial_i F_j = \partial_j F_i \). The inequality then follows by noting that
\[
|2[F,\alpha](i,j)| = |\langle e_i - e_j, F'(e_i - e_j) \rangle| \leq \|F'\|_2 \|e_i - e_j\|_2^2 = 2\|F\|_{\text{Lip}}.
\]
Finally, if \( i \in S_i \) and \( j \in S_j \) with \( i < j \), then we have to prove that
\[
|s_i - s_j| \leq \|F\|_{Lip}(\tilde{\alpha}_i - \tilde{\alpha}_j)
\] (3.1)

If we show that this holds for any two adjacent numbers, we can write \( s_i - s_j = \sum_{k=i+1}^{j-1} s_k - s_{k+1} \) and obtain (3.1) by using the triangle inequality. We thus assume that \( j = i + 1 \). Since \( F \) is block-constant it is no restriction to assume that \( j = i + 1 \), and hence we have to show that
\[
|F_i(\alpha) - F_{i+1}(\alpha)| \leq \|F\|_{Lip}(\alpha_i - \alpha_{i+1}).
\]

If we let \( \gamma \) be obtained from \( \alpha \) by replacing the values on positions \( i \) and \( i + 1 \) with \( \frac{\alpha_i + \alpha_{i+1}}{2} \), we have \( F_i(\gamma) = F_{i+1}(\gamma) \), again using that \( F \) is block-constant. Thus
\[
F_i(\alpha) - F_{i+1}(\alpha) = (F_i(\alpha) - F_i(\gamma)) + (F_{i+1}(\gamma) - F_{i+1}(\alpha)) = \langle e_i - e_{i+1}, F(\alpha) - F(\gamma) \rangle.
\]
Since \( \|\alpha - \gamma\| = \frac{\alpha_i - \alpha_{i+1}}{\sqrt{2}} \) and the modulus of the right hand side can be estimated by \( \|e_i - e_{i+1}\| \|F\|_{Lip} \|\alpha - \gamma\| \), the desired inequality follows.

\(\square\)

**Lemma 3.2.** Let \( F \) be a continuous block-constant vector-field in \( \mathbb{R}^n \) and define \( G : \mathbb{R}^n \to \mathbb{R}^n \) as the convolution \( G(x) = F^{ext} * \Psi(x) \), where \( \Psi(y) = \psi(y_1)\psi(y_2)\ldots\psi(y_n) \) and \( \psi : \mathbb{R} \to \mathbb{R} \) is any \( C^\infty \)-function with compact support. Then \( G \) is block-constant and \( C^\infty \).

**Proof.** That \( G \) becomes \( C^\infty \) is a standard fact whose proof we omit. Note that \( \Psi \) is permutation invariant, i.e. \( \Psi(\Pi x) = \Psi(x) \) for all permutations \( \Pi \). We fix \( x \) and consider \( y \) as a variable. Given any point \( x - y \) the value of \( F^{ext}(x - y) \) is given by \( F^{ext}(x - y) = \Sigma_y F(\Sigma_y(x - y)) \) with \( \Sigma_y \in sort(x - y) \), and since \( F \) is block-constant this value is independent of the particular choice of \( \Sigma_y \) in case of ambiguity (i.e. when \( x - y \) is on the boundary of \( \mathbb{R}^n \)). Let \( \Pi \) be an arbitrary perturbation. Then
\[
\Pi^* G(\Pi x) = \int_{\mathbb{R}^n} \Pi^* F^{ext}(\Pi x - y) \Psi(y) dy = \int_{\mathbb{R}^n} \Pi^* F^{ext}(\Pi(x - \Pi^* y)) \Psi(\Pi^* y) dy = \int_{\mathbb{R}^n} \Pi^* F^{ext}(\Pi(x - y)) \Psi(y) dy = \int_{\mathbb{R}^n} \Pi^* \Upsilon_{\Pi y} F(\Upsilon_{\Pi y} \Pi(x - y)) \Psi(y) dy
\]

where each \( \Upsilon_{\Pi y} \) is such that \( \Upsilon_{\Pi y} \Pi(x - y) \in \mathbb{R}^n_+ \). By the comments before the computation, we can consider \( \Upsilon_{\Pi y} \Pi \) as \( \Sigma_y \) and it follows that \( \Pi \) has no effect on the outcome, so the above computation equals \( \int_{\mathbb{R}^n} \Sigma_y F(\Sigma_y(x - y)) \Psi(y) dy = G(x) \), as desired. \(\square\)

We now come to the final theorem of the paper. This type of results can also be proved using the convex theory of complex sub-stochastic matrices, see [2].

**Theorem 3.3.** Assume that \( F \) is block-constant on \( \mathbb{R}^n_+ \) and Lipschitz continuous. Then
\[
\|\mathcal{L}_F(A) - \mathcal{L}_F(B)\|_2 \leq \|F\|_{Lip} \|A - B\|_2.
\] (3.2)

**Proof.** It is easy to see that \( F^{ext} \) is continuous. By Lemma 3.2 and a standard approximation argument, we may assume that \( F \) is \( C^\infty \) on \( \mathbb{R}^n \) with the same Lipschitz constant. Moreover a simple matrix approximation argument shows that we may assume that \( A \) has only simple eigenvalues. Set \( E = A - B \) and let \( U_i \Delta_\xi U_i^* \) be the spectral decomposition of \( B + tE \). By the discussion in Chapter II, Sec 1.1 [11], we know that \( \xi_i \) has distinct elements for all \( t \) except finitely many. Since \( C^1 \) vector-fields are automatically \( C^1 \) point-symmetric at all vectors
with distinct elements, \( F \) is \( C^1 \) point-symmetric for all \( \xi \) except finitely many values of \( t \). By simply ignoring these points and appealing to Theorem 2.7 we get that

\[
\left\| \mathcal{L}_F(A) - \mathcal{L}_F(B) \right\|_2 = \left\| \int_0^1 \frac{d}{dt} \mathcal{L}_F(B + tE) dt \right\|_2 \leq \\
\int_0^1 \left\| U_t \left( \left[ F, \xi_t \right] \circ \hat{E} + \text{diag}(F'_{|\xi_t})\hat{e} \right) U_t^* \right\|_2 dt \leq \sup_{0 \leq t \leq 1} \left\| U_t \left( \left[ F, \xi_t \right] \circ \hat{E} + \text{diag}(F'_{|\xi_t})\hat{e} \right) U_t^* \right\|_2.
\]

Note that \( \hat{E} = U_t E U_t^* \) implicitly depends on \( t \). As noted in the proof of Theorem 2.7 the supindex \( o \) can be moved from \( (F'_{|\xi_t})^o \) to \( [F, \xi_t] \). Thus we get

\[
\left\| U_t \left( \left[ F, \xi_t \right] \circ \hat{E} + \text{diag}(F'_{|\xi_t})\hat{e} \right) U_t^* \right\|_2^2 \leq \sum_{i \neq j} \left( \left| F_{i\xi_t} \hat{E}_{ij} \right|^2 + \left| F'_{i\xi_t} \hat{e} \right|^2 \right) \leq \sum_{i \neq j} \left( \left| F_{i\xi_t} \right|^2 \left| \hat{E}_{ij} \right|^2 + \left| F'_{i\xi_t} \hat{e} \right|^2 \right) = \left\| F \right\|_{Lip}^2 \left\| \hat{E} \right\|_2^2
\]

where we used Lemma 3.1. Since \( \left\| \hat{E} \right\|_2 = \left\| E \right\|_2 = \left\| A - B \right\|_2 \), the theorem follows by inserting this estimate in the above supremum. \( \square \)

The assumption that \( F \) is block-constant in all of \( \mathbb{R}^n \) is crucial for the above result to be true. To see this consider the case when \( \mathcal{L}_F \) is the orthogonal projection onto the first eigenspace and \( n = 2 \) say. The matrices

\[
\begin{pmatrix}
1 + \pm \varepsilon & 0 \\
0 & 1 + \mp \varepsilon
\end{pmatrix}
\]

then show that \( \mathcal{L}_F \) is not continuous, despite \( F(x) = (1, 0) \) being constant (and hence having Lipschitz constant 0).

REFERENCES

[1] Fredrik Andersson, Marcus Carlsson, and Carl Olsson. Convex envelopes for fixed rank approximation. Optimization Letters, pages 1–13, 2017.
[2] Fredrik Andersson, Marcus Carlsson, and Carl-Mikael Perfekt. Operator-lipschitz estimates for the singular value functional calculus. Proc. Amer. Math. Soc., 144:1867–1875, 2016.
[3] Heinz H Bauschke, Patrick L Combettes, et al. Convex analysis and monotone operator theory in Hilbert spaces, volume 2011. Springer, 2017.
[4] Rajendra Bhatia. Matrix analysis, volume 169. Springer Science & Business Media, 2013.
[5] Marcus Carlsson. Perturbation theory for the matrix square root and matrix modulus. https://arxiv.org/pdf/1810.01464.pdf.
[6] Marcus Carlsson. Perturbation theory for the spectral decomposition of hermitian matrices. arXiv preprint arXiv:1809.09480, 2018.
[7] Ju L Daletskii and SG Krein. Integration and differentiation of functions of hermitian operators and applications to the theory of perturbations. AMS Translations (2), 47(1-30), 1965.
[8] Chandler Davis and William Morton Kahan. The rotation of eigenvectors by a perturbation. iii. SIAM Journal on Numerical Analysis, 7(1):1–46, 1970.
[9] Christian Grussler and Pontus Giselsson. Low-rank inducing norms with optimality interpretations. SIAM Journal on Optimization, 28(4):3057–3078, 2018.
[10] Nicholas J Higham. Functions of matrices: theory and computation, volume 104. Siam, 2008.
[11] Tosio Kato. Perturbation theory for linear operators, volume 132. Springer Science & Business Media, 2013.
[12] Steven George Krantz. Function theory of several complex variables, volume 340. American Mathematical Soc., 2001.
[13] Viktor Larsson and Carl Olsson. Convex low rank approximation. International Journal of Computer Vision, 120(2):194–214, 2016.
[14] Adrian S Lewis and Hristo S Sendov. Twice differentiable spectral functions. SIAM Journal on Matrix Analysis and Applications, 23(2):368–386, 2001.
[15] Adrian Stephen Lewis. Derivatives of spectral functions. *Mathematics of Operations Research*, 21(3):576–588, 1996.

[16] Horn Roger and R Johnson Charles. *Topics in matrix analysis*. Cambridge University Press, 1991.

[17] Hristo S Sendov. The higher-order derivatives of spectral functions. *Linear algebra and its applications*, 424(1):240–281, 2007.

**Centre for Mathematical Sciences, Lund University, Box 118, SE-22100, Lund, Sweden, E-mail address: marcus.carlsson@math.lu.se**