Generalized Hadamard Product and the Derivatives of Spectral Functions

Hristo S. Sendov*

April 3, 2006

Abstract

Real valued function, $F(X)$, on a symmetric matrix argument are called spectral if $F(U^T XU) = F(X)$ for every orthogonal matrix U and $X \in \operatorname{dom} F$. We are interested in a description of the higher order derivatives (when they exists) of F with respect to X. Formulae for the gradient and the Hessian of F are given in [7] and [11]. In this work we present common features of these two formulae, that are preserved in the higher order derivatives.

Keywords: spectral function, twice differentiable, higher-order derivatives, eigenvalues, symmetric function, perturbation theory, multi-linear algebra.
Mathematics Subject Classification (2000): primary: 49R50; 47A75, secondary: 15A18; 15A69.

1 Introduction

Spectral functions, are real valued functions on a symmetric matrix argument invariant under conjugation by orthogonal matrices. More precisely, $F : S^n \to \mathbb{R}$ is spectral if

$$F(U^T XU) = F(X),$$

*Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, Canada N1G 2W1. Email: `hssendov@uoguelph.ca`. Research supported by NSERC.

for all $X \in \operatorname{dom} F$ and $U \in O^n$ — the orthogonal group on $\mathbb{R}^n$. By restricting F to the subspace of diagonal matrices, it is not difficult to see that spectral functions can be represented as the composition

$$F = f \circ \lambda,$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a symmetric function ($f(Px) = f(x)$ for any permutation matrix P and vector x), and $\lambda : S^n \to \mathbb{R}^n$ is the eigenvalue map: $\lambda(X) = (\lambda_1(X), ..., \lambda_n(X))$ — the vector of eigenvalues of X. We assume throughout that,

$$\lambda_1(X) \geq \cdots \geq \lambda_n(X).$$

The study of spectral functions generalizes the study of the individual eigenvalues of a symmetric matrix since if we let

$$\begin{aligned} &\phi_k(x) : \mathbb{R}^n \to \mathbb{R}, \\ &\phi_k(x) := \text{ the } k^{\text{th}} \text{ largest element of } \{x_1, ..., x_n\} \end{aligned}$$

then, $\phi_k(x)$ is symmetric and

$$\lambda_k(X) = (\phi_k \circ \lambda)(X).$$

Various differential properties of eigenvalues have been studied for a long time. They find a lot of applications in areas ranging from matrix perturbation theory [17], and eigenvalue optimization [10], [9], to quantum mechanics [4]. The Taylor directional expansion (when it exists) of the eigenvalues of symmetric matrices depending on one scalar parameter is described in the monograph by Kato [3]. This naturally raises the questions about the differentiability properties of the more general spectral functions. Many such questions have already been investigated in the literature and the answer to most of them follows the same pattern: $f \circ \lambda$ has a property at the matrix X if, and only if, f has the same property at the vector $\lambda(X)$. In this way, properties of the function $f \circ \lambda$ on S^n are reduced to properties of the simpler function f on $\mathbb{R}^n$.

Some of the properties of $f \circ \lambda$ at (or around) a matrix X that hold if, and only if, f has the same property at (or around) the vector $\lambda(X)$ are:

(i) F is lower semicontinuous at X if, and only if, f is at $\lambda(X)$, [6].

(ii) F is lower semicontinuous and convex if, and only if, f is, [2], [6].

(iii) The symmetric function corresponding to the Fenchel conjugate of F is the Fenchel conjugate of f, [14], [6]. (A similar statement holds for the recession function of F, [14].)

(iv) F is pointed, has good asymptotic behaviour or is a barrier function on the set $\lambda^{-1}(C)$ if, and only if, f is such on C, [14].

(v) F is Lipschitz around X if, and only if, f is such around $\lambda(X)$, [7]

(vi) F is (continuously) differentiable at X if, and only if, f is at $\lambda(X)$, [7].

(vii) F is strictly differentiable at X if, and only if, f is at $\lambda(X)$, [7], [8].

(viii) $\nabla(f \circ \lambda)$ is semismooth at X if, and only if, ∇f is at $\lambda(X)$, [13].

(ix) If f is l.s.c. and convex then, F is twice epi-differentiable at X relatively to Ω if, and only if, f is twice epi-differentiable at $\lambda(X)$ relative to $\lambda(\Omega)$, [18], where Ω is an arbitrary epi-gradient.

(x) F has a quadratic expansion at X if, and only if, f has a quadratic expansion at $\lambda(X)$, [12].

(xi) F is twice (continuously) differentiable at X if, and only if, f is twice (continuously) differentiable at $\lambda(X)$, [11].

(xii) $F \in \mathcal{C}^\infty$ at $X \Leftrightarrow f \in \mathcal{C}^\infty$ at $\lambda(X)$, [1].

(xiii) F is analytic at X if, and only if, f is at $\lambda(X)$, [19].

(xiv) F is a polynomial of the entries of X if, and only if, f is a polynomial. This is a consequence of the Chevalley Restriction Theorem, [20, p. 143].

There are of course exceptions to that pattern. For example, f being directionally differentiable at $\lambda(X)$ does not imply that $f \circ \lambda$ is such at X, see [7].

Formulae for the gradient and the Hessian of the spectral function F given in terms of the derivatives of the symmetric function f were derived in [7] and [11]. In order to reproduce them here we need a bit more notation. For any vector x in $\mathbb{R}^n$, denote by $\mathrm{Diag}\, x$ the diagonal matrix with vector x on the main diagonal, and denote by $\mathrm{diag}\colon M^n \to \mathbb{R}^n$ its dual operator defined by $\mathrm{diag}\,(X) = (x_{11}, ..., x_{nn})$. Recall that the Hadamard product of two matrices $A = [A^{ij}]$ and $B = [B^{ij}]$ of the same dimensions is the matrix $A \circ B = [A^{ij}B^{ij}]$. Thus we have

$$\nabla(f \circ \lambda)(X) = V\big(\mathrm{Diag}\,\nabla f(\lambda(X))\big)V^T, \text{ and} \tag{1}$$

$$\begin{aligned} \nabla^2(f \circ \lambda)(X)[H_1, H_2] = \nabla^2 f(\lambda(X))[\mathrm{diag}\,\tilde{H}_1, \mathrm{diag}\,\tilde{H}_2]+ \\ + \langle \mathcal{A}(\lambda(X)), \tilde{H}_1 \circ \tilde{H}_2 \rangle, \end{aligned} \tag{2}$$

where V is any orthogonal matrix such that $X = V\big(\mathrm{Diag}\,\lambda(X)\big)V^T$ is the *ordered* spectral decomposition of X; $\tilde{H}_i = V^T H_i V$ for $i = 1, 2$, and $x \in \mathbb{R}^n \to \mathcal{A}(x)$ is a matrix valued map that is continuous if $\nabla^2 f(x)$ is.

In [11] a conjecture was made that F is k-times (continuously) differentiable at X if, and only if, f is such at $\lambda(X)$. When that happens, a natural issue is to find a practical description of the k^{th} derivative of F and an efficient way to compute it. In addition, explicit formulae for the first k^{th} derivatives of F generalize the terms in the k^{th} order Taylor directional expansion (when it exists) of the individual eigenvalues, given in [3].

This work aims to generalize some common features in Formulae (1) and (2), that are preserved in the higher order derivatives of $f \circ \lambda$. The language we present simplifies the description of the higher order derivatives of spectral functions and offers a systematic way for evaluating them, when those derivatives are viewed as multi-linear functions on the space of symmetric matrices. In Section 2 we introduce a multi-linear map on the space of square matrices, that generalizes the Hadamard product between two matrices. In Section 4 we present its multi-linear dual operator that generalizes the Diag operator. The connections with the derivatives of spectral functions are pointed out throughout.

The current paper is the first of three. In [15] we formulate calculus-type rules for the interaction between that generalization of the Hadamard product and the eigenvalues of symmetric matrices. Then in [16] we describe how to compute the higher order derivatives of spectral functions in two general cases. For example, we show that Conjecture 4.1 holds for the derivatives of any function (non necessarily symmetric) of the eigenvalues of symmetric matrices, at a matrix X with distinct eigenvalues. And second, we show that it holds for the derivatives of separable spectral functions at an arbitrary symmetric matrix X. (Separable spectral functions are those arising from symmetric functions $f(x) = g(x_1) + \cdots + g(x_n)$ for some function g on a scalar argument.) The computation of the maps $\mathcal{A}_\sigma(x)$ (see Equation (16) below) in these two cases is particularly simple.

2 Generalizations of the Hadamard product

By $\{H_{pq} : 1 \leq p, q \leq n\}$ we denote the standard basis of the space M^n of all $n \times n$ real matrices. That is, the matrices H_{pq} are such that $(H_{pq})^{ij}$ is 1 if $(i, j) = (p, q)$, and 0 otherwise.

The Hadamard product, $H_1 \circ H_2$, between two matrices H_1 and H_2 from M^n is a matrix valued function on two matrix arguments, linear in each argument separately. Thus, it is uniquely determined by its values on the pairs of basic matrices $(H_{p_1q_1}, H_{p_2q_2})$. On such basic pairs the Hadamard product is defined as:

$$(H_{p_1q_1} \circ H_{p_2q_2})^{ij} = \begin{cases} 1, & \text{if } i = p_1 = p_2 \text{ and } j = q_1 = q_2, \\ 0, & \text{otherwise.} \end{cases}$$

Analogous object is obtained if a *cross Hadamard product* is defined as follows

$$(H_{p_1q_1} \circ_{(12)} H_{p_2q_2})^{ij} := \begin{cases} 1, & \text{if } i = p_1 = q_2 \text{ and } j = p_2 = q_1, \\ 0, & \text{otherwise,} \end{cases}$$

and then, extended to a bilinear function on $M^n \times M^n$. The Hadamard product and the cross Hadamard product are essentially the same:

$$H_{p_1q_1} \circ_{(12)} H_{p_2q_2} = H_{p_1q_1} \circ H^T_{p_2q_2} = H_{p_1q_1} \circ H_{q_2p_2}.$$

These observations can be generalized in the following way. Denote by $\mathbb{N}$ the set of all natural numbers and by $\mathbb{N}_k$ the set $\{1, 2, ..., k\}$. A *k-tensor on* $\mathbb{R}^n$ is a real-valued map on $\mathbb{R}^n \times \cdots \times \mathbb{R}^n$ (k-times) linear in each argument separately. When a basis in $\mathbb{R}^n$ is fixed, a k-tensor can be viewed as an $n \times \cdots \times n$ (k-times) "block" of numbers. We index the elements of a tensor in a similar way we index the entries of a matrix, thus by $T^{i_1 \ldots i_k}$ we denote the $(i_1, ..., i_k)$-th entry of T. The space of all k-tensors on $\mathbb{R}^n$ will be denoted by $T^{k,n}$. The set of all permutations on $\mathbb{N}_k$ as well as the set of all $n \times n$ permutation matrices will be denoted by P^k. (It will be clear from the context which one we mean.)

Definition 2.1 For a fixed permutation σ on $\mathbb{N}_k$, define the *σ-Hadamard* product between k basic matrices, $H_{p_1q_1}$, $H_{p_2q_2}$,...,$H_{p_kq_k}$, to be a k-tensor on $\mathbb{R}^n$ as follows:

$$(H_{p_1q_1} \circ_\sigma H_{p_2q_2} \circ_\sigma \cdots \circ_\sigma H_{p_kq_k})^{i_1i_2\ldots i_k} = \begin{cases} 1, & \text{if } i_s = p_s = q_{\sigma(s)}, \forall s = 1, ..., k, \\ 0, & \text{otherwise.} \end{cases}$$

Now, extend this product to a k-tensor valued map on k matrix arguments, linear in each of them separately.

Another way to write the above definition is using the Kronecker delta symbol. Recall that δ_{ij} is equal to 1 if $i = j$, and 0 otherwise. Thus,

$$
\begin{aligned}
(H_{p_1q_1} \circ_\sigma H_{p_2q_2} \circ_\sigma \cdots \circ_\sigma H_{p_kq_k})^{i_1i_2\ldots i_k} &= \delta_{i_1p_1}\delta_{i_1q_{\sigma(1)}} \cdots \delta_{i_kp_k}\delta_{i_kq_{\sigma(k)}} \\
&= \delta_{i_1p_1}\delta_{p_1q_{\sigma(1)}} \cdots \delta_{i_kp_k}\delta_{p_kq_{\sigma(k)}}.
\end{aligned} \tag{3}
$$

The next lemma gives the formula for the general entry of the σ-Hadamard product between arbitrary matrices.

Lemma 2.2 *The σ-Hadamard product of arbitrary matrices is given by*

$$
\begin{aligned}
(H_1 \circ_\sigma H_2 \circ_\sigma \cdots \circ_\sigma H_k)^{i_1i_2\ldots i_k} &= H_1^{i_1i_{\sigma^{-1}(1)}} \cdots H_k^{i_ki_{\sigma^{-1}(k)}} \\
&= H_{\sigma(1)}^{i_{\sigma(1)}i_1} \cdots H_{\sigma(k)}^{i_{\sigma(k)}i_k}.
\end{aligned}
$$

Proof. Let σ be a permutation on $\mathbb{N}_k$ and let H_1,...,H_k be arbitrary matrices. Since the product is linear in each argument separately, we compute

$$
\begin{aligned}
&(H_1 \circ_\sigma H_2 \circ_\sigma \cdots \circ_\sigma H_k)^{i_1i_2\ldots i_k} \\
&\quad = \sum_{p_1,q_1=1}^{n,n} \cdots \sum_{p_k,q_k=1}^{n,n} H_1^{p_1q_1} \cdots H_k^{p_kq_k} (H_{p_1q_1} \circ_\sigma H_{p_2q_2} \circ_\sigma \cdots \circ_\sigma H_{p_kq_k})^{i_1i_2\ldots i_k} \\
&\quad = \sum_{p_1,q_1=1}^{n,n} \cdots \sum_{p_k,q_k=1}^{n,n} H_1^{p_1q_1} \cdots H_k^{p_kq_k} \delta_{i_1p_1}\delta_{i_1q_{\sigma(1)}} \cdots \delta_{i_kp_k}\delta_{i_kq_{\sigma(k)}} \\
&\quad = \sum_{p_1,q_1=1}^{n,n} \cdots \sum_{p_k,q_k=1}^{n,n} H_1^{p_1q_1} \cdots H_k^{p_kq_k} \delta_{i_1p_1}\delta_{i_{\sigma^{-1}(1)}q_1} \cdots \delta_{i_kp_k}\delta_{i_{\sigma^{-1}(k)}q_k} \\
&\quad = H_1^{i_1i_{\sigma^{-1}(1)}} \cdots H_k^{i_ki_{\sigma^{-1}(k)}} \\
&\quad = H_{\sigma(1)}^{i_{\sigma(1)}i_1} \cdots H_{\sigma(k)}^{i_{\sigma(k)}i_k}.
\end{aligned}
$$

■

Corollary 2.3 *When the first $k-1$ of the matrices involved in the product are basic we get*

$$(H_{p_1q_1} \circ_\sigma \cdots \circ_\sigma H_{p_{k-1}q_{k-1}} \circ_\sigma H)^{i_1i_2\ldots i_k}$$

$$
= \begin{cases} H^{i_k i_k}\Big(\prod_{s=1}^{k-1} \delta_{i_s p_s}\delta_{i_s q_{\sigma(s)}}\Big), & \textit{if } k = \sigma^{-1}(k) \\ H^{i_{\sigma(l)} i_l}\big(\delta_{i_l p_l}\delta_{i_k q_{\sigma(k)}}\big)\Big(\prod_{\substack{s=1 \\ s\neq l}}^{k-1} \delta_{i_s p_s}\delta_{i_s q_{\sigma(s)}}\Big), & \textit{if } l := \sigma^{-1}(k) \neq k. \end{cases}
$$

Proof. Suppose first that $l := \sigma^{-1}(k) \neq k$. Using the result of the previous lemma we calculate

$$
\begin{aligned}
&(H_{p_1 q_1} \circ_\sigma \cdots \circ_\sigma H_{p_{k-1} q_{k-1}} \circ_\sigma H)^{i_1 i_2 \ldots i_k} = H_{p_1 q_1}^{i_1 i_{\sigma^{-1}(1)}} \cdots H_{p_{k-1} q_{k-1}}^{i_{k-1} i_{\sigma^{-1}(k-1)}} H^{i_k i_{\sigma^{-1}(k)}} \\
&= \delta_{i_1 p_1}\delta_{i_{\sigma^{-1}(1)} q_1} \cdots \delta_{i_{k-1} p_{k-1}}\delta_{i_{\sigma^{-1}(k-1)} q_{k-1}} H^{i_k i_{\sigma^{-1}(k)}} \\
&= \delta_{i_1 p_1}\delta_{i_1 q_{\sigma(1)}} \cdots \delta_{i_{l-1} p_{l-1}}\delta_{i_{l-1} q_{\sigma(l-1)}} H^{i_{\sigma(l)} i_l} \delta_{i_{l+1} p_{l+1}}\delta_{i_{l+1} q_{\sigma(l+1)}} \cdots \\
&\qquad\qquad \cdots \delta_{i_{k-1} p_{k-1}}\delta_{i_{k-1} q_{\sigma(k-1)}} \big(\delta_{i_l p_l}\delta_{i_k q_{\sigma(k)}}\big).
\end{aligned}
$$

The case $l = k$ follows as well. ■

The above corollary can be easily modified when the matrix H is in arbitrary position in the product.

We often represent a permutation by its cycle decomposition. For example, (123)(45) is the permutation on $\mathbb{N}_5$ that maps 1 to 2, 2 to 3, 3 to 1, in addition to 4 to 5 and 5 to 4.

Example 2.4 We already saw that, when $k = 2$ and $\sigma = (12)$ the σ-Hadamard product is essentially the ordinary Hadamard product:

$$
H_1 \circ_{(12)} H_2 = H_1 \circ H_2^T.
$$

If we restrict our attention to the space of symmetric matrices then, the two products coincide. In the case when $\sigma = (1)(2)$ we get

$$
H_1 \circ_{(1)(2)} H_2 = (\operatorname{diag} H_1)(\operatorname{diag} H_2)^T.
$$

Example 2.5 In the case $k = 1$, there is one permutation, $\sigma = (1)$, on the elements of the set $\mathbb{N}_1$ and the σ-Hadamard product corresponding to it is a vector valued linear map:

$$
(\circ_\sigma H_{p_1 q_1})^{i_1} = \begin{cases} 1, & \text{if } i_1 = p_1 = q_1 \\ 0, & \text{otherwise} \end{cases}
$$

$$= (\operatorname{diag} H_{q_1p_1})^{i_1}.$$

Extending by linearity we get

$$\circ_\sigma H = \operatorname{diag} H.$$

The standard scalar product between any two k-tensors, T_1, and T_2 is given by:

$$\langle T_1, T_2 \rangle = \sum_{i_1,\ldots,i_k=1}^{n,\ldots,n} T_1^{i_1\ldots i_k} T_2^{i_1\ldots i_k}.$$

Lemma 2.6 *Let T be a k-tensor on $\mathbb{R}^n$, and H be a matrix in M^n. Let $H_{p_1q_1},\ldots,H_{p_{k-1}q_{k-1}}$ be basic matrices in M^n, and let σ be a permutation on $\mathbb{N}_k$. Then, the following identities hold.*

(i) If $\sigma^{-1}(k) = k$ then,

$$\langle T, H_{p_1q_1} \circ_\sigma \cdots \circ_\sigma H_{p_{k-1}q_{k-1}} \circ_\sigma H \rangle = \Big(\prod_{t=1}^{k-1} \delta_{p_t q_{\sigma(t)}} \Big) \sum_{t=1}^{n} T^{p_1\ldots p_{k-1} t} H^{tt}.$$

(ii) If $\sigma^{-1}(k) = l$, where $l \neq k$ then,

$$\langle T, H_{p_1q_1} \circ_\sigma \cdots \circ_\sigma H_{p_{k-1}q_{k-1}} \circ_\sigma H \rangle = \Big(\prod_{\substack{t=1 \\ t\neq l}}^{k-1} \delta_{p_t q_{\sigma(t)}} \Big) T^{p_1\ldots p_{k-1} q_{\sigma(k)}} H^{q_{\sigma(k)} p_{\sigma^{-1}(k)}}.$$

Proof. Using the definitions and observation (3), we calculate.

$$\langle T, H_{p_1q_1} \circ_\sigma H_{p_2q_2} \circ_\sigma \cdots \circ_\sigma H_{p_{k-1}q_{k-1}} \circ_\sigma H \rangle$$

$$= \sum_{p_k,q_k=1}^{n,n} H^{p_kq_k} \langle T, H_{p_1q_1} \circ_\sigma H_{p_2q_2} \circ_\sigma \cdots \circ_\sigma H_{p_{k-1}q_{k-1}} \circ_\sigma H_{p_kq_k} \rangle$$

$$= \sum_{p_k,q_k=1}^{n,n} H^{p_kq_k} \sum_{i_1,\ldots,i_k=1}^{n,\ldots,n} T^{i_1\ldots i_k} (H_{p_1q_1} \circ_\sigma H_{p_2q_2} \circ_\sigma \cdots \circ_\sigma H_{p_{k-1}q_{k-1}} \circ_\sigma H_{p_kq_k})^{i_1\ldots i_k}$$

$$= \sum_{p_k,q_k=1}^{n,n} H^{p_kq_k} \sum_{i_1,\ldots,i_k=1}^{n,\ldots,n} T^{i_1\ldots i_k} \delta_{i_1p_1} \delta_{p_1q_{\sigma(1)}} \cdots \delta_{i_kp_k} \delta_{p_kq_{\sigma(k)}}$$

$$= \sum_{p_k,q_k=1}^{n,n} H^{p_kq_k} T^{p_1\ldots p_k} \delta_{p_1q_{\sigma(1)}} \cdots \delta_{p_kq_{\sigma(k)}}.$$

The result follows easily by considering the two cases separately. ■

3 A partial order on P^k and a property of the σ-Hadamard product

Given two permutations σ, μ on $\mathbb{N}_k$, we say that σ *refines* μ if for every $s \in \mathbb{N}_k$ there is an $r \in \mathbb{N}_k$ such that

$$\{\sigma^l(s) \,:\, l = 1, 2, ...\} \subseteq \{\mu^l(r) \,:\, l = 1, 2, ...\},$$

where $\sigma^l(s) = \sigma(\sigma(\cdots(\sigma(s))\cdots)$ - l times. Informally, σ refines μ if the elements of every cycle of σ are contained in a cycle of μ, thus, the cycles of σ partition the cycles of μ. If σ refines μ we denote it by

$$\mu \preceq \sigma.$$

The refinement relationship is a pre-order on P^k (it is reflexive, transitive, but not antisymmetric). With respect to this pre-order, the identity permutation is the biggest element (that is, bigger than any one else) and every permutation with only one cycle is a smallest element (that is, it is smaller than any other element).

There is a natural map between the set P^k and the *diagonal subspaces* of $\mathbb{R}^k$, given as follows:

$$\mathcal{D}(\sigma) = \{x \in \mathbb{R}^k \,:\, x_s = x_{\sigma(s)} \ \forall s \in \mathbb{N}_k\}.$$

This map is onto but is not one-to-one since, for example, when $k = 3$ $\mathcal{D}((123)) = \mathcal{D}((132)) = \{x \in \mathbb{R}^3 \,:\, x_1 = x_2 = x_3\}$. The image of the identity permutation is $\mathbb{R}^k$. The following relationship helps to visualize the partial order on P^k

$$\mu \preceq \sigma \ \Leftrightarrow\ \mathcal{D}(\mu) \subseteq \mathcal{D}(\sigma).$$

Given a permutation $\mu \in P^k$ and a tensor $T \in T^{k,n}$, we denote by $P_\mu(T)$ the tensor in $T^{k,n}$ defined by

$$(P_\mu(T))^{i_1 \ldots i_k} = \begin{cases} T^{i_1 \ldots i_k}, & \text{if } i_s = i_{\mu(s)},\ \forall s \in \mathbb{N}_k \\ 0, & \text{otherwise.} \end{cases}$$

Informally, the tensor $P_\mu(T)$ preserves the entries of T lying on the "subspace" $\mathcal{D}(\mu)$ of T and replaces the rest of the entries with zeros.

Next is the main result of this section. It describes exactly when one can "transfer" diagonal "subspaces" of T between different σ-Hadamard products.

Theorem 3.1 *Let σ_1, σ_2, and μ be three permutations on $\mathbb{N}_k$. Then, the identity*

$$\langle P_\mu(T), H_1 \circ_{\sigma_1} \cdots \circ_{\sigma_1} H_k \rangle = \langle P_\mu(T), H_1 \circ_{\sigma_2} \cdots \circ_{\sigma_2} H_k \rangle \tag{4}$$

holds for any matrices H_1,...,H_k, and any tensor T in $T^{k,n}$ if, and only if, $\mu \preceq \sigma_2^{-1} \circ \sigma_1$.

Proof. Since both sides are linear in each of the matrices H_1,..,H_k separately, it is enough to prove the theorem when these matrices are basic. In other words, we show that

$$\langle P_\mu(T), H_{p_1q_1} \circ_{\sigma_1} \cdots \circ_{\sigma_1} H_{p_kq_k} \rangle = \langle P_\mu(T), H_{p_1q_1} \circ_{\sigma_2} \cdots \circ_{\sigma_2} H_{p_kq_k} \rangle,$$

for any indexes p_1,...,p_k, q_1,...,q_k, and any $T \in T^{k,n}$ if, and only if, $\mu \preceq \sigma_2^{-1} \circ \sigma_1$. Direct calculation gives:

$$\begin{aligned}
&\langle P_\mu(T), H_{p_1q_1} \circ_{\sigma_1} \cdots \circ_{\sigma_1} H_{p_kq_k} \rangle \\
&\qquad = \sum_{i_1,\dots,i_k=1}^{n,\dots,n} (P_\mu(T))^{i_1\dots i_k} (H_{p_1q_1} \circ_{\sigma_1} \cdots \circ_{\sigma_1} H_{p_kq_k})^{i_1\dots i_k} \\
&\qquad = \sum_{i_1,\dots,i_k=1}^{n,\dots,n} (P_\mu(T))^{i_1\dots i_k} H_{p_1q_1}^{i_1 i_{\sigma_1^{-1}(1)}} \cdots H_{p_kq_k}^{i_k i_{\sigma_1^{-1}(k)}} \\
&\qquad = \sum_{i_1,\dots,i_k=1}^{n,\dots,n} (P_\mu(T))^{i_1\dots i_k} \delta_{i_1p_1} \delta_{i_1 q_{\sigma_1(1)}} \cdots \delta_{i_kp_k} \delta_{i_k q_{\sigma_1(k)}} \\
&\qquad = (P_\mu(T))^{p_1\dots p_k} \delta_{p_1 q_{\sigma_1(1)}} \cdots \delta_{p_k q_{\sigma_1(k)}}.
\end{aligned}$$

The last expression is equal to $T^{p_1\dots p_k}$ when $p_s = p_{\mu(s)} = q_{\sigma_1(s)}$ for all $s \in \mathbb{N}_k$, and is equal to 0 otherwise.

Analogously, the right-hand side of (4) is

$$\langle P_\mu(T), H_{p_1q_1} \circ_{\sigma_2} \cdots \circ_{\sigma_2} H_{p_kq_k} \rangle = (P_\mu(T))^{p_1\dots p_k} \delta_{p_1 q_{\sigma_2(1)}} \cdots \delta_{p_k q_{\sigma_2(k)}},$$

which is equal to $T^{p_1\dots p_k}$ when $p_s = p_{\mu(s)} = q_{\sigma_2(s)}$ for all $s \in \mathbb{N}_k$, and is equal to 0 otherwise.

Suppose that $\mu \preceq \sigma_2^{-1} \circ \sigma_1$. We consider three cases.

If there is an s_0 such that $p_{s_0} \neq p_{\mu(s_0)}$ then, both sides of (4) are zero and the equality is trivial.

If $p_s = p_{\mu(s)}$ for all $s \in \mathbb{N}_k$ but for some s_0 we have that $p_{s_0} \neq q_{\sigma_1(s_0)}$ then, it is not possible to have $p_s = q_{\sigma_2(s)}$ for all $s \in \mathbb{N}_k$. Indeed, suppose on the contrary that $p_s = q_{\sigma_2(s)}$ for all $s \in \mathbb{N}_k$. Letting $r = \sigma_2(s)$ we get $p_{\sigma_2^{-1}(r)} = q_r$ for every $r \in \mathbb{N}_k$. Therefore $p_{\sigma_2^{-1}(\sigma_1(s))} = q_{\sigma_1(s)}$ for every $s \in \mathbb{N}_k$ and in particular $p_{\sigma_2^{-1}(\sigma_1(s_0))} = q_{\sigma_1(s_0)} \neq p_{s_0}$. But $\mu \preceq \sigma_2^{-1} \circ \sigma_1$ implies that $\sigma_2^{-1}(\sigma_1(s_0))$ and s_0 belong to the same cycle of μ, that is $\mu^l(s_0) = \sigma_2^{-1}(\sigma_1(s_0))$ for some $l \in \mathbb{N}$. By the assumption in this case we have that $p_{s_0} = p_{\mu^l(s_0)}$ for every l, a contradiction. Thus, for some $s_1 \in \mathbb{N}_k$ we have $p_{s_1} \neq q_{\sigma_2(s_1)}$ and again both sides of (4) are equal to zero.

Suppose finally that $p_s = p_{\mu(s)} = q_{\sigma_1(s)}$ for all $s \in \mathbb{N}_k$. Then, the left-hand side of (4) is equal to $T^{p_1 \cdots p_k}$. We are done if we show that $p_s = q_{\sigma_2(s)}$ for every $s \in \mathbb{N}_k$. Suppose this is not true, that is, for some s_0, $p_{s_0} \neq q_{\sigma_2(s_0)}$. Then, for $r_0 = \sigma_2(s_0)$ we have $p_{\sigma_2^{-1}(r_0)} \neq q_{r_0}$, and for $s_1 = \sigma_1^{-1}(r_0)$ we have $p_{\sigma_2^{-1}(\sigma_1(s_1))} \neq q_{\sigma_1(s_1)}$. The condition $\mu \preceq \sigma_2^{-1} \circ \sigma_1$ implies that $\sigma_2^{-1}(\sigma_1(s_1))$ and s_1 belong to the same cycle of μ and we reach a contradiction as in the previous case.

To prove the opposite direction of the theorem, suppose that

$$(P_\mu(T))^{p_1 \cdots p_k} \delta_{p_1 q_{\sigma_1(1)}} \cdots \delta_{p_k q_{\sigma_1(k)}} = (P_\mu(T))^{p_1 \cdots p_k} \delta_{p_1 q_{\sigma_2(1)}} \cdots \delta_{p_k q_{\sigma_2(k)}}, \tag{5}$$

for every choice of the indexes p_1,...,p_k and q_1,...,q_k and every T. Take T to be such that $T^{i_1 \dots i_k} \neq 0$ for every choice of the indexes $i_1, ..., i_k$ satisfying $i_s = i_{\mu(s)}$ for every $s \in \mathbb{N}_k$. Suppose that $\mu \not\preceq \sigma_2^{-1} \circ \sigma_1$. This means that there is a number $s_0 \in \mathbb{N}_k$ such that $\sigma_2^{-1}(\sigma_1(s_0))$ and s_0 are not in the same cycle of μ. Choose the indexes p_1,...,p_k and q_1,...,q_k so that $p_s = p_{\mu(s)}$ and $p_s = q_{\sigma_1(s)}$, for every $s \in \mathbb{N}_k$. Moreover, choose the indexes p_1,...,p_k so that if $s, r \in \mathbb{N}_k$ are not in the same cycle of μ then, $p_s \neq p_r$. This in particular means that

$$p_{\sigma_2^{-1}(\sigma_1(s_0))} \neq p_{s_0}. \tag{6}$$

With those choices, the left-hand side of (5) is equal to $T^{p_1 \cdots p_k} \neq 0$. We reach a contradiction by showing that for some r_0, $p_{r_0} \neq q_{\sigma_2(r_0)}$ implying that the right-hand side of (5) is zero. Suppose on the contrary that $p_r = q_{\sigma_2(r)}$ for every $r \in \mathbb{N}_k$. Then, $p_{\sigma_2^{-1}(\sigma_1(s))} = q_{\sigma_1(s)} = p_s$, for every $s \in \mathbb{N}_k$, contradicting (6). We are done. ■

Notice that if $\mu \preceq \nu$ then, for arbitrary permutation σ in P^k we have

$$\mu \preceq \nu^{-1} = (\sigma \circ \nu)^{-1} \circ \sigma.$$

This observation leads to the next corollary.

Corollary 3.2 *Suppose μ and ν are permutations in P^k such that $\mu \preceq \nu$. Then, for an arbitrary permutation $\sigma \in P^k$, any matrices H_1,...,H_k, and a tensor T in $T^{k,n}$ we have the identity:*

$$\langle P_\mu(T), H_1 \circ_\sigma \cdots \circ_\sigma H_k \rangle = \langle P_\mu(T), H_1 \circ_{\sigma\circ\nu} \cdots \circ_{\sigma\circ\nu} H_k \rangle.$$

In particular, the result holds when $\nu = \mu$ or $\nu = \mu^{-1}$.

It is useful to see explicitly the conclusions of the above theorem when $k \leq 3$. We summarize them in the next corollary.

Corollary 3.3 *For any $T \in T^{2,n}$ and any two matrices H_1 and H_2 we have*

$$\langle P_{_{(12)}}(T), H_1 \circ_{_{(1)(2)}} H_2 \rangle = \langle P_{_{(12)}}(T), H_1 \circ_{_{(12)}} H_2 \rangle.$$

For any $T \in T^{3,n}$ and any three matrices H_1, H_2, and H_3 we have

$$\begin{aligned}
\langle P_{_{(13)}}(T), H_1 \circ_{_{(132)}} H_2 \circ_{_{(132)}} H_3 \rangle &= \langle P_{_{(13)}}(T), H_1 \circ_{_{(12)(3)}} H_2 \circ_{_{(12)(3)}} H_3 \rangle, \\
\langle P_{_{(23)}}(T), H_1 \circ_{_{(123)}} H_2 \circ_{_{(123)}} H_3 \rangle &= \langle P_{_{(23)}}(T), H_1 \circ_{_{(12)(3)}} H_2 \circ_{_{(12)(3)}} H_3 \rangle,
\end{aligned}$$

and

$$\begin{aligned}
\langle P_{_{(13)}}(T), H_1 \circ_{_{(13)(2)}} H_2 \circ_{_{(13)(2)}} H_3 \rangle &= \langle P_{_{(13)}}(T), H_1 \circ_{_{(1)(2)(3)}} H_2 \circ_{_{(1)(2)(3)}} H_3 \rangle, \\
\langle P_{_{(23)}}(T), H_1 \circ_{_{(1)(23)}} H_2 \circ_{_{(1)(23)}} H_3 \rangle &= \langle P_{_{(23)}}(T), H_1 \circ_{_{(1)(2)(3)}} H_2 \circ_{_{(1)(2)(3)}} H_3 \rangle.
\end{aligned}$$

Finally, for any two permutations σ_1, σ_2 on $\mathbb{N}_3$ we have

$$\langle P_{_{(123)}}(T), H_1 \circ_{\sigma_1} H_2 \circ_{\sigma_1} H_3 \rangle = \langle P_{_{(123)}}(T), H_1 \circ_{\sigma_2} H_2 \circ_{\sigma_2} H_3 \rangle.$$

Example 3.4 Let us have another look at Formula (1) for the first derivative of a spectral function at X. Let $X = V(\operatorname{Diag}\lambda(X))V^T$ and $\tilde{E} = V^T E V$, where E is a symmetric matrix. Using the definitions and notation in the previous subsection we have:

$$\begin{aligned}
\nabla(f \circ \lambda)(X)[E] &= \langle V\big(\operatorname{Diag}\nabla f(\mu)\big)V^T, E \rangle \\
&= \langle \nabla f(\mu), \operatorname{diag} \tilde{E} \rangle \\
&= \langle \nabla f(\mu), \circ_{_{(1)}} \tilde{E} \rangle.
\end{aligned}$$

Example 3.5 Let X be a symmetric matrix with ordered spectral decomposition $X = V(\mathrm{Diag}\,\lambda(X))V^T$. Take two symmetric matrices E_1 and E_2 and let $\tilde{E}_i = V^T E_i V$ for $i = 1, 2$. As we saw in the examples in Section 2 we have:

$$E_1 \circ_{(1)(2)} E_2 = (\mathrm{diag}\, E_1)(\mathrm{diag}\, E_2)^T \quad \text{and} \quad E_1 \circ_{(12)} E_2 = E_1 \circ E_2.,$$

Then, Formula (2) for the Hessian of the spectral function $f \circ \lambda$ becomes:

$$\begin{aligned} \nabla^2(f \circ \lambda)(X)[E_1, E_2] &= \nabla^2 f(\lambda(X))[\mathrm{diag}\, \tilde{E}_1, \mathrm{diag}\, \tilde{E}_2] + \langle \mathcal{A}(\lambda(X)), \tilde{E}_1 \circ \tilde{E}_2 \rangle \\ &= \langle \nabla^2 f(\lambda(X)), \tilde{E}_1 \circ_{(1)(2)} \tilde{E}_2 \rangle + \langle \mathcal{A}(\lambda(X)), \tilde{E}_1 \circ_{(12)} \tilde{E}_2 \rangle. \end{aligned}$$

These examples support the following conjecture, describing the structure of the higher-order derivatives of spectral functions. (More instances of when the conjecture holds are given after its equivalent reformulation in Conjecture 4.1.)

Conjecture 3.1 The spectral function $f \circ \lambda$ is k times (continuously) differentiable at X if, and only if, $f(x)$ is k times (continuously) differentiable at the vector $\lambda(X)$. Moreover, there are k-tensor valued maps $\mathcal{A}_\sigma : \mathbb{R}^n \to T^{k,n}$, $\sigma \in P^k$, depending only on the symmetric function f, such that for any symmetric matrices E_1,...,E_k we have

$$\nabla^k(f \circ \lambda)(X)[E_1, ..., E_k] = \sum_{\sigma \in P^k} \langle \mathcal{A}_\sigma(\lambda(X)), \tilde{E}_1 \circ_\sigma \cdots \circ_\sigma \tilde{E}_k \rangle, \tag{7}$$

where $X = V(\mathrm{Diag}\,\lambda(X))V^T$ and $\tilde{E}_i = V^T E_i V$, for $i = 1, .., k$.

The left-hand side of Formula (7) is the k-th derivative of the spectral function evaluated at the matrices $E_1, ..., E_k$ while on the right side these matrices are conjugated by V and "jumbled" into the σ-Hadamard products $\tilde{E}_1 \circ_\sigma \cdots \circ_\sigma \tilde{E}_k$. Our goal in the next section is to identify more clearly the multi-linear operator on the right-hand side of (7) acting on the matrices $E_1, ..., E_k$.

4 The Diag^σ operator

Recall that the adjoint of the linear operator $\mathrm{Diag} : \mathbb{R}^n \to M^n$ is the operator $\mathrm{diag} : M^n \to \mathbb{R}^n$. That is, we have the identity

$$\langle \mathrm{Diag}\, x, H \rangle = \langle x, \mathrm{diag}\, H \rangle, \tag{8}$$

for any vector x and any matrix H. It is also easy to verify that for any vector x, matrix H, and orthogonal matrix U we have

$$(9) \qquad \langle U(\operatorname{Diag} x)U^T, H\rangle = \langle x, \operatorname{diag}(U^T HU)\rangle = \langle x, \circ_{(1)}(U^T HU)\rangle,$$

where the last equality is trivial from Example 2.5.

In this section we generalize Equations (8) and (9) for an arbitrary k-tensor in place of x and an arbitrary σ-Hadamard product in place of $\circ_{(1)}$.

Let T be an arbitrary k-tensor on $\mathbb{R}^n$ and let σ be a permutation on $\mathbb{N}_k$. We define $\operatorname{Diag}^\sigma T$ to be a $2k$-tensor on $\mathbb{R}^n$ in the following way

$$(\operatorname{Diag}^\sigma T)^{i_1\ldots i_k}_{j_1\cdots j_k} = \begin{cases} T^{i_1\ldots i_k}, & \text{if } i_s = j_{\sigma(s)}, \forall s \in \mathbb{N}_k \\ 0, & \text{otherwise.} \end{cases}$$

Informally speaking, viewing tensors as cubes placed at the origin of the positive orthant of a Euclidean space and its indices as coordinates, then the operator $\operatorname{Diag}^\sigma T$ lifts T onto the k-dimensional diagonal plane defined by

$$\{(x, y) \in \mathbb{R}^k \times \mathbb{R}^k \,|\, x_s = y_{\sigma(s)}, \text{ for all } s \in \mathbb{N}_k\}.$$

Notice that this map between the permutations on $\mathbb{N}_k$ and k-dimensional diagonal subspaces of $\mathbb{R}^k \times \mathbb{R}^k$ is one-to-one.

When $k = 1$ and $\sigma = (1)$, the definition of $\operatorname{Diag}^\sigma T$ coincides with the definition of the Diag operator in Equation (8). An equivalent way to define $\operatorname{Diag}^\sigma T$ useful for calculations is:

$$(\operatorname{Diag}^\sigma T)^{i_1\ldots i_k}_{j_1\cdots j_k} = T^{i_1\ldots i_k}\delta_{i_1 j_{\sigma(1)}}\cdots\delta_{i_k j_{\sigma(k)}}.$$

We now consider an action, call it *conjugation*, of the group, O^n, of all $n \times n$ orthogonal matrices on the space of all k-tensors on $\mathbb{R}^n$. For any k-tensor T, and $U \in O^n$ this action will be denoted by UTU^T, and defined by:

$$(10) \qquad (UTU^T)^{i_1\ldots i_k} = \sum_{p_1=1}^{n}\cdots\sum_{p_k=1}^{n}\Big(T^{p_1\ldots p_k}U^{i_1 p_1}\cdots U^{i_k p_k}\Big).$$

When $k = 1$, this is exactly the action of O^n on $\mathbb{R}^n$, and when $k = 2$ the definition coincides with the conjugate action of O^n on the space of $n \times n$

square matrices. In general, it is not difficult to see that it is possible to order the entries of $T \in T^{k,n}$ into a vector $\mathrm{vec}\,(T) \in \mathbb{R}^{n^k}$ such that

$$UTU^T = \mathrm{vec}^{\,-1}\big((\otimes^k U)\mathrm{vec}\,(T)\big), \tag{11}$$

where $\otimes^k U$ is the k-th tensor power of U and $\mathrm{vec}^{\,-1}$ is the inverse of the linear operation vec. The fact that $\otimes^k U$ is an orthogonal matrix whenever U is, the well known identity $(\otimes^k V)(\otimes^k U) = \otimes^k (VU)$, and (11) show the following lemma.

Lemma 4.1 *The conjugate action is associative and norm preserving. That is, for any k-tensor, T, on $\mathbb{R}^n$ and any two orthogonal matrices U, V in O^n*

$$V(UTU^T)V^T = (VU)T(VU)^T,$$

and

$$\|UTU^T\| = \|T\|.$$

Any $2k$-tensor, T, on R^n can naturally be viewed as a k-tensor on the space M^n in the following way. Let H_1,...,H_k be any $n \times n$ matrices then,

$$T[H_1, ..., H_k] := \sum_{p_1,q_1=1}^{n,n} \cdots \sum_{p_k,q_k=1}^{n,n} T^{p_1 \ldots p_k}_{q_1 \ldots q_k} H_1^{p_1 q_1} \cdots H_k^{p_k q_k}.$$

Let P be an $n \times n$ permutation matrix and σ its corresponding permutation on $\mathbb{N}_n$, that is, $P^T e^i = e^{\sigma(i)}$ for all $i = 1, ..., n$, where $\{e^i \,|\, i = 1, ..., n\}$ is the standard basis in $\mathbb{R}^n$. The action of P on the tensors is what one expects it to be:

$$(PTP^T)^{i_1 \ldots i_k} = \sum_{p_1=1}^{n} \cdots \sum_{p_k=1}^{n} \Big(T^{p_1 \ldots p_k} \prod_{\nu=1}^{k} P^{i_\nu p_\nu} \Big) = T^{\sigma(i_1) \ldots \sigma(i_k)}.$$

The conjugation by an orthogonal matrix is defined on tensors on $\mathbb{R}^n$ of any dimension. The next lemma shows that the conjugation by a permutation matrix commutes with the lifting operation $\mathrm{Diag}^{\,\mu}$, for any permutation μ.

Lemma 4.2 *For any permutation μ on $\mathbb{N}_k$, any permutation matrix P in P^n and any k-tensor T on $\mathbb{R}^n$, we have*

$$P(\mathrm{Diag}^{\,\mu} T)P^T = \mathrm{Diag}^{\,\mu}(PTP^T).$$

Proof. Let σ be the permutation on $\mathbb{N}_n$ corresponding to P. Fix any multi index $\binom{i_1 \ldots i_k}{j_1 \ldots j_k}$. We begin calculating the right-hand side entry corresponding to that index. In the third equality below, we use the fact that σ is a one-to-one map.

$$\begin{aligned}
\big(P(\operatorname{Diag}{}^{\mu} T)P^T\big)^{i_1 \ldots i_k}_{j_1 \ldots j_k} &= \big(\operatorname{Diag}{}^{\mu} T\big)^{\sigma(i_1) \ldots \sigma(i_k)}_{\sigma(j_1) \ldots \sigma(j_k)} \\
&= T^{\sigma(i_1) \ldots \sigma(i_k)} \delta_{\sigma(i_1)\sigma(j_{\mu(1)})} \cdots \delta_{\sigma(i_k)\sigma(j_{\mu(k)})} \\
&= T^{\sigma(i_1) \ldots \sigma(i_k)} \delta_{i_1 j_{\mu(1)}} \cdots \delta_{i_k j_{\mu(k)}} \\
&= (PTP^T)^{i_1 \ldots i_k} \delta_{i_1 j_{\mu(1)}} \cdots \delta_{i_k j_{\mu(k)}} \\
&= \big(\operatorname{Diag}{}^{\mu}(PTP^T)\big)^{i_1 \ldots i_k}_{j_1 \ldots j_k}. \qquad \blacksquare
\end{aligned}$$

These preparations lead to the following generalization to Equation (9). (When, $k = 1$ and $\sigma = (1)$ we obtain exactly Equation (9).)

Theorem 4.3 *For any k-tensor T on $\mathbb{R}^n$, any matrices $H_1, \ldots, H_k$ in M^n, any orthogonal matrix U in O^n, and any permutation σ on $\mathbb{N}_k$ we have the identity*

$$\langle T, \tilde{H}_1 \circ_\sigma \cdots \circ_\sigma \tilde{H}_k \rangle = \big(U(\operatorname{Diag}{}^{\sigma} T)U^T\big)[H_1, \ldots, H_k], \tag{12}$$

where $\tilde{H}_i = U^T H_i U$, for all $i = 1, 2, \ldots, k$.

Proof. Since both sides are linear in each argument separately, it is enough to show that the equality holds for k-tuples $(H_{i_1 j_1}, \ldots, H_{i_k j_k})$ of basic matrices.

Using Lemma 2.2 and the fact that $\tilde{H}^{pq}_{ij} = U^{ip} U^{jq}$, we develop the left-hand side of Equation (12):

$$\begin{aligned}
\langle T, \tilde{H}_{i_1 j_1} \circ_\sigma \cdots \circ_\sigma \tilde{H}_{i_k j_k} \rangle &= \sum_{p_1, \ldots, p_k = 1}^{n, \ldots, n} T^{p_1 \ldots p_k} \tilde{H}^{p_1 p_{\sigma^{-1}(1)}}_{i_1 j_1} \cdots \tilde{H}^{p_k p_{\sigma^{-1}(k)}}_{i_k j_k} \\
&= \sum_{p_1, \ldots, p_k = 1}^{n, \ldots, n} T^{p_1 \ldots p_k} U^{i_1 p_1} U^{j_1 p_{\sigma^{-1}(1)}} \cdots U^{i_k p_k} U^{j_k p_{\sigma^{-1}(k)}}.
\end{aligned}$$

On the other hand, using the definitions we calculate that the right-hand side is:

$$(U(\operatorname{Diag}{}^{\sigma} T)U^T)[H_{i_1 j_1}, \ldots, H_{i_k j_k}] =$$

$$= \sum_{p_1,q_1=1}^{n,n} \cdots \sum_{p_k,q_k=1}^{n,n} \big((U(\mathrm{Diag}\,^{\sigma}T)U^T)^{p_1\ldots p_k}_{q_1\ldots q_k} H^{p_1 q_1}_{i_1 j_1} \cdots H^{p_k q_k}_{i_k j_k}\big)$$
$$= (U(\mathrm{Diag}\,^{\sigma}T)U^T)^{i_1\ldots i_k}_{j_1\ldots j_k}$$
$$= \sum_{p_1,q_1=1}^{n,n} \cdots \sum_{p_k,q_k=1}^{n,n} \Big((\mathrm{Diag}\,^{\sigma}T)^{p_1\ldots p_k}_{q_1\ldots q_k} \prod_{\nu=1}^{k} U^{i_\nu p_\nu} U^{j_\nu q_\nu}\Big)$$
$$= \sum_{p_1=1}^{n} \cdots \sum_{p_k=1}^{n} \Big(T^{p_1\ldots p_k} \prod_{\nu=1}^{k} U^{i_\nu p_\nu} U^{j_\nu p_{\sigma^{-1}(\nu)}}\Big).$$

This shows that the both sides are equal. ■

Corollary 4.4 *For any k-tensor T, any matrices H_1,...,H_k, and any permutation σ on $\mathbb{N}_k$, we have the identity*

$$\langle T, H_1 \circ_\sigma \ldots \circ_\sigma H_k \rangle = (\mathrm{Diag}\,^{\sigma}T)[H_1, \ldots, H_k]. \tag{13}$$

If in Corollary 4.4 we substitute the matrices H_1,...,H_k with $\tilde{H}_1$,...,$\tilde{H}_k$ and we use Theorem 4.3, we obtain the next result.

Corollary 4.5 *For any k-tensor T, orthogonal matrix $U \in O^n$, permutation σ on $\mathbb{N}_k$, and any matrices H_1,...,H_k we have the identity*

$$(\mathrm{Diag}\,^{\sigma}T)[\tilde{H}_1, \ldots, \tilde{H}_k] = \big(U(\mathrm{Diag}\,^{\sigma}T)U^T\big)[H_1, \ldots, H_k]. \tag{14}$$

If in Corollary 4.4 we take σ to be the identity permutation then, we get the next corollary, which generalizes Equation (8).

Corollary 4.6 *For any k-tensor T, any matrices H_1,...,H_k we have the identity*

$$T[\mathrm{diag}\, H_1, \ldots, \mathrm{diag}\, H_k] = (\mathrm{Diag}\,^{(\mathrm{id})}T)[H_1, \ldots, H_k]. \tag{15}$$

We conclude this section with a second look at the first two derivatives of spectral functions.

Example 4.7 As we saw in Example 3.4, the first derivative of the spectral function $f \circ \lambda$ at the point $X = V(\mathrm{Diag}\,\lambda(X))V^T$, applied to the symmetric matrix E is given by the formula

$$\nabla(f\circ\lambda)(X)[E] = \langle V\big(\mathrm{Diag}\,\nabla f(\lambda(X))\big)V^T, E\rangle = V\big(\mathrm{Diag}\,^{(1)}\nabla f(\lambda(X))\big)V^T[E].$$

The usefulness of the notation becomes more evident below.

Example 4.8 Let X be a symmetric matrix with ordered spectral decomposition $X = V(\operatorname{Diag}\lambda(X))V^T$. Take two symmetric matrices E_1 and E_2 and let $\tilde{E}_i = V^T E_i V$ for $i = 1, 2$. As we saw in Example 3.4, the Hessian of the spectral function $f \circ \lambda$ at the point $X = V(\operatorname{Diag}\lambda(X))V^T$, applied to the symmetric matrices E_1 and E_2 is given by the formula

$$\nabla^2(f \circ \lambda)(X)[E_1, E_2] = \langle \nabla^2 f(\lambda(X)), \tilde{E}_1 \circ_{(1)(2)} \tilde{E}_2 \rangle + \langle \mathcal{A}(\lambda(X)), \tilde{E}_1 \circ_{(12)} \tilde{E}_2 \rangle.$$

With the notation introduced in this section we can rewrite it as

$$\begin{aligned}\nabla^2(f \circ \lambda)(X)[E_1, E_2] = \big(V\big(\operatorname{Diag}^{(1)(2)} \nabla^2 f(\lambda(X))\big)V^T\big)[E_1, E_2] \\ + \big(V\big(\operatorname{Diag}^{(12)} \mathcal{A}(\lambda(X))\big)V^T\big)[E_1, E_2].\end{aligned}$$

Or, in other words

$$\nabla^2(f \circ \lambda)(X) = V\big(\operatorname{Diag}^{(1)(2)} \nabla^2 f(\lambda(X)) + \operatorname{Diag}^{(12)} \mathcal{A}(\lambda(X))\big)V^T.$$

Finally, we express Conjecture 3.1 in the new language.

Conjecture 4.1 The spectral function $f \circ \lambda$ is k times (continuously) differentiable at X if, and only if, $f(x)$ is k times (continuously) differentiable at the vector $\lambda(X)$. Moreover, there are k-tensor valued maps $\mathcal{A}_\sigma : \mathbb{R}^n \to T^{k,n}$, $\sigma \in P^k$, depending only on the symmetric function f, such that

$$\nabla^k(f \circ \lambda)(X) = V\Big(\sum_{\sigma \in P^k} \operatorname{Diag}^{\sigma} \mathcal{A}_\sigma(\lambda(X))\Big)V^T, \tag{16}$$

where $X = V(\operatorname{Diag}\lambda(X))V^T$.

Formula (16) says that the orthogonal matrix V in the ordered spectral decomposition of X also "diagonalizes" the k-th derivative of $f \circ \lambda$ at X. Moreover, the effect of the eigenvalues in the right-hand side of (16) can very clearly be seen: only V and $\lambda(X)$ depend on the eigenvalues. In addition, we can easily evaluate the derivative, as a multi-linear function, at any k symmetric matrices, using Theorem 4.3 and the σ-Hadamard product. Finally, there are precisely $k!$ summands in the right-hand side of (16), this should be compared with the classical Faà de Bruno formula [5, Lemma 1.3.1] for the k-th derivative of the composition of two (smooth) functions, in which the number of summands in highly nontrivial.

In [16] we show that this conjecture holds for the derivatives of any function (not necessarily symmetric) of the eigenvalues of symmetric matrices, at a symmetric matrix X with distinct eigenvalues; as well as for the derivatives of separable spectral functions at an arbitrary symmetric matrix X. (Separable spectral functions are those arising from symmetric functions $f(x) = g(x_1) + \cdots + g(x_n)$ for some function g on a scalar argument.) There we also describe how, for every σ in P^k, to compute the operators $\mathcal{A}_\sigma(x)$, depending only on the symmetric function $f(x)$.

5 Sufficient condition for Conjecture 4.1

Recall that Examples (4.7), and (4.8) show that Conjecture 4.1 holds for $k = 1$ and $k = 2$. The next Theorem summarizes this section.

Theorem 5.1 *Using the notation from Conjecture 4.1 we have.*

- *It is enough to establish Conjecture 4.1 only in the case when the* $X = \operatorname{Diag} x$ *for some* $x \in \mathbb{R}^n$ *with* $x_1 \geq \cdots \geq x_n$.
- *If the maps* $\mathcal{A}_\sigma$ *are continuous at* $\lambda(X)$ *for all* $\sigma \in P^k$, *then* $\nabla^k(f \circ \lambda)$ *is continuous at* X.

We begin with a simple lemma. For brevity, given a k-tensor, T, on M^n by $T[H]$ we denote the $(k-1)$-tensor $T[\cdot, \cdots, H]$.

Lemma 5.2 *Let* T *be any* $2k$*-tensor on* R^n, $U \in O^n$, *and let* H *be any matrix. Then, the following identity holds.*

$$U(T[\tilde{H}])U^T = (UTU^T)[H],$$

where $\tilde{H} = U^T H U$.

Proof. Since both sides are linear with respect to H, it is enough to prove the identity only for basic matrices $H_{i_k j_k}$. By the definition of conjugation, and using the fact that $\tilde{H}^{pq}_{i_k j_k} = U^{i_k p} U^{j_k q}$ we obtain

$$\big(U(T[\tilde{H}_{i_k j_k}])U^T\big)^{i_1 \ldots i_{k-1}}_{j_1 \cdots j_{k-1}}$$

$$= \sum_{\substack{p_s,q_s=1\\ s=1,\dots,k-1}}^{n,\dots,n} (T[\tilde{H}_{i_kj_k}])^{p_1\dots p_{k-1}}_{q_1\dots q_{k-1}} U^{i_1p_1}U^{j_1q_1}\dots U^{i_{k-1}p_{k-1}}U^{j_{k-1}q_{k-1}}$$
$$= \sum_{\substack{p_s,q_s=1\\ s=1,\dots,k}}^{n,\dots,n} T^{p_1\dots p_k}_{q_1\dots q_k} U^{i_1p_1}U^{j_1q_1}\dots U^{i_kp_k}U^{j_kq_k}$$
$$= (UTU^T)^{i_1\dots i_k}_{j_1\dots j_k}$$
$$= \big((UTU^T)[H_{i_kj_k}]\big)^{i_1\dots i_{k-1}}_{j_1\dots j_{k-1}}. \qquad \blacksquare$$

We now establish the first part of Theorem 5.1. Suppose that Conjecture 4.1 holds for all derivatives of order less than k and for the k-th derivative it holds only for ordered diagonal matrices. We show that the conjecture holds for the k-th derivative at an arbitrary matrix. Indeed, let $X = V(\mathrm{Diag}\,\lambda(X))V^T$, let E be arbitrary symmetric matrix and denote $\tilde{E} = V^TEV$. Then,

$$\begin{aligned}
&\nabla^{k-1}F(X+E) = \nabla^{k-1}F\big(V(\mathrm{Diag}\,\lambda(X)+\tilde{E})V^T\big)\\
&= V\big(\nabla^{k-1}F(\mathrm{Diag}\,\lambda(X)+\tilde{E})\big)V^T\\
&= V\big(\nabla^{k-1}F(\mathrm{Diag}\,\lambda(X))\big)V^T + V\big(\nabla^{k}F(\mathrm{Diag}\,\lambda(X))[\tilde{E}]\big)V^T + o(\|E\|)\\
&= \nabla^{k-1}F(X) + \big(V(\nabla^{k}F(\mathrm{Diag}\,\lambda(X)))V^T\big)[E] + o(\|E\|),
\end{aligned}$$

where in the last equality we used Lemma 5.2. This shows that $\nabla^{k-1}F$ is differentiable at X and that $V(\nabla^kF(\mathrm{Diag}\,\lambda(X)))V^T$ is the k-th derivative of F at X.

The second part of Theorem 5.1 is the next proposition.

Proposition 5.3 *Suppose the k-th derivative of the spectral function $F = f\circ\lambda$ is given by Equation (16) for all X. If for every $\sigma \in P^k$ the tensor valued map $x \in \mathbb{R}^n \to \mathcal{A}_\sigma(x) \in T^{k,n}$ is continuous then, $\nabla^kF(X)$ is continuous in X, in other words $F \in C^k$.*

Proof. Suppose that there is a sequence of symmetric matrices X_m approaching X and an $\epsilon > 0$ such that

$$\|\nabla^kF(X_m) - \nabla^kF(X)\| > \epsilon, \text{ for all } m.$$

Let $X_m = V_m(\mathrm{Diag}\,\lambda(X_m))V_m^T$ and suppose without loss of generality that the orthogonal V_m approaches V (otherwise, take a subsequence.) By continuity of the eigenvalues, we have that $X = V(\mathrm{Diag}\,\lambda(X))V^T$ and that $\lambda(X_m)$ approaches $\lambda(X)$. Using the formula for the k-th derivative and the continuity of the maps $\mathcal{A}_\sigma(x)$, the contradiction follows. ■

Acknowledgement: We would like to thank the first referee for their thorough reading of the manuscript, useful comments, and pointing out several typos.

References

[1] J. DADOK. On the C^∞ Chevalley's theorem. *Advances in Mathematics*, 44 (1982), pp. 121–131.

[2] C. DAVIS. All convex invariant functions of hermitian matrices. *Archiv der Mathematik*, 8 (1957), pp. 276–278.

[3] T. KATO. *Perturbation Theory for Linear Operators*, 2nd ed., Grundlehren der Mathematischen Wissenschaften, Band 132. Springer-Verlag New York, Inc., Berlin, 1976.

[4] E.C. KEMBLE. *The Fundamental Principles of Quantum Mechanics*. Dover, New York, 1958.

[5] S.G. KRATNZ and H.R. PARKS. *A Primer of Real Analitic Functions* Basler Lehrbücker, a series of advanced textbooks in mathematics, vol 4, Birkhäuser Verlag, Basel, Switzerland, 1992.

[6] A.S. LEWIS. Convex analysis on the Hermitian matrices. *SIAM Journal on Optimization*, 6 (1996), pp. 164–177.

[7] A.S. LEWIS. Derivatives of spectral functions. *Mathematics of Operations Research*, 21 (1996), pp. 576–588.

[8] A.S. LEWIS. Nonsmooth analysis of eigenvalues. *Mathematical Programming*, 84 (1999), pp. 1–24.

[9] A.S. LEWIS. The mathematics of eigenvalue optimization. *Mathematical Programming, Series B*, 97 (2003), pp. 155–176.

[10] A.S. LEWIS and M.L. OVERTON. Eigenvalue optimization. *Acta Numerica*, 5 (1996), pp. 149–190.

[11] A.S. LEWIS and H.S. SENDOV. Twice differentiable spectral functions. *SIAM Journal of Matrix Analysis and Applications*, 23 (2001), pp. 368–386.

[12] A.S. LEWIS and H.S. SENDOV. Quadratic expansions of spectral functions. *Linear Algebra and Its Applications*, 340 (2002), pp. 97–121.

[13] H. QI and X. YANG. Semismoothness of spectral functions. *SIAM Journal on Matrix Analysis and its Applications*, 2003. To appear.

[14] A. SEEGER. Convex analysis of spectrally defined matrix functions. *SIAM Journal on Optimization*, 7 (1997), pp. 679–696.

[15] H.S. SENDOV. Interactions between the generalized Hadamard product and the eigenvalues of symmetric matrices. *Research Report, Department of Mathematics and Statistics, University of Guelph, Ontario.* Mathematical Series 2004-309, , 2004.

[16] H.S. SENDOV. On the higher-order derivatives of spectral functions: Two special cases. *Research Report, Department of Mathematics and Statistics, University of Guelph, Ontario.* Mathematical Series 2004-310, 2004.

[17] G. W. STEWARD and J. SUN. *Matrix Perturbation Theory.* Academic Press, 1990.

[18] M. TORKI. Second-order epi-differentiability and subdifferentiability of spectral convex functions. *Personal Communication*, 1999.

[19] N.-K. TSING, M. K. H. FAN, and E. I. VERRIEST. On analyticity of functions involving eigenvalues. *Linear Algebra and its Applications*, 207 (1994), pp. 159–180.

[20] G. WARNER. *Harmonic Analysis on Semi-Simple Lie Groups.* Springer-Verlag, Berlin-Heidelberg-New York, 1972.