Generalized Hadamard Product and the Derivatives of Spectral Functions

Hristo S. Sendov*

October 23, 2018

Abstract

In this work we propose a generalization of the Hadamard product between two matrices to a tensor-valued, multi-linear product between k matrices for any $k \geq 1$. A multi-linear dual operator to the generalized Hadamard product is presented. It is a natural generalization of the $\operatorname{Diag} x$ operator, that maps a vector $x \in \mathbb{R}^n$ into the diagonal matrix with x on its main diagonal. Defining an action of the $n \times n$ orthogonal matrices on the space of k-dimensional tensors, we investigate its interactions with the generalized Hadamard product and its dual. The research is motivated, as illustrated throughout the paper, by the apparent suitability of this language to describe the higher-order derivatives of spectral functions and the tools needed to compute them. For more on the later we refer the reader to [14] and [15], where we use the language and properties developed here to study the higher-order derivatives of spectral functions.

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

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

1 Introduction

Spectral functions, are functions on a symmetric matrix argument invariant under a closed subgroup of the orthogonal group on the space of all $n \times n$ symmetric matrices, S^n. More precisely, $F : S^n \to \mathbb{R}$ is spectral if

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

for all $X \in S^n$ and $U \in O(n)$ — the orthogonal group on $\mathbb{R}^n$. It is not difficult to see that such 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: $\lambda(X) = (\lambda_1(X), ..., \lambda_n(X))$ — all eigenvalues of X. We will 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 smoothness properties of eigenvalues have been studied for some time now and find a lot of applications in areas ranging from matrix perturbation theory [16], and eigenvalue optimization [9], [8], to quantum mechanics [4]. The Taylor expansion (when it exists) of the eigenvalues of symmetric matrices depending on one scalar parameter are described in the monograph by Kato [3]. This naturally raises the questions about the differentiability properties of the spectral functions. Many such questions have already been investigated in the literature (see below) and surprisingly the answers to most of them follow 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 $\lambda(X)$. It is only natural, then to try to describe the differentials of $f \circ \lambda$ in terms of the differentials of the simpler function f.

Here is a list of properties for which $f \circ \lambda$ has that property at (or around) the matrix X if and only if f has the same property at (or around) the vector $\lambda(X)$.

(i) F is lower semicontinuous at A if, and only if, f is at $\lambda(A)$, [5].

(ii) F is lower semicontinuous and convex if, and only if, f is, [2], [5].

(iii) The symmetric function corresponding to the Fenchel conjugate of F is the Fenchel conjugate of f, [13], [5]. (A similar statement holds for the recession function of F, [13].)

(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, [13].

(v) F is Lipschitz around A if, and only if, f is such around $\lambda(A)$, [6]

(vi) F is (continuously) differentiable at A if, and only if, f is at $\lambda(A)$, [6].

(vii) F is strictly differentiable at A if, and only if, f is at $\lambda(A)$, [6], [7].

(viii) $\nabla(f \circ \lambda)$ is semismooth at X if, and only if, ∇f is at $\lambda(X)$, [12].

(ix) If f is l.s.c. and convex, then F is twice epi-differentiable at A relatively to Ω if, and only if, f is twice epi-differentiable at $\lambda(A)$ relative to $\lambda(\Omega)$, [17], 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 $\lambda(X)$, [11].

(xi) F is twice (continuously) differentiable at X if, and only if, f is twice (continuously) differentiable at $\lambda(X)$, [10].

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

(xiii) F is analytic at A if, and only if, f is at $\lambda(A)$, [18].

(xiv) F is a polynomial of the entries of A if, and only if, f is a polynomial. This is a consequence of the Chevalley Restriction Theorem, [19, p. 143].

We want to stress that there are exceptions to the pattern. For example if f is directionally differentiable at $\lambda(X)$ this doesn't imply that $f \circ \lambda$ is such at X, see [6].

In [6] and [10] the authors gave explicit formulae for the gradient and the Hessian of the spectral function F in terms of the derivatives of the symmetric function f. In order to reproduce them here we need a bit more notation. For any vector x in $\mathbb{R}^n$, $\operatorname{Diag} x$ will denote the diagonal matrix with vector x on the main diagonal, and $\operatorname{diag}\colon M^n \to \mathbb{R}^n$ will denote its dual operator defined by $\operatorname{diag}(X) = (x_{11}, ..., x_{nn})$. Recall that the Hadamard product of two matrices $A = [A^{ij}]$ and $B = [B^{ij}]$ of the same size is the matrix of their element-wise product $A \circ B = [A^{ij}B^{ij}]$. Thus we have

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

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

where V is any orthogonal matrix such that $X = V\big(\operatorname{Diag} \lambda(X)\big)V^T$ 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 [10] a conjecture was made that F is k-times (continuously) differentiable at A if, and only if, f is such at $\lambda(A)$. It is conceivable that high-powered analytical methods may give a direct proof of this conjecture, but never the less an interesting question is what the k^{th} differential of F looks like and how to compute it practically. Explicit formula for the k^{th} differential of F will generalize the formula for the k^{th} term in the Taylor expansion (when it exists) of the individual eigenvalues given in [3].

Before attacking the questions in the previous paragraph we need to answer several more basic questions. What are the common features in Formulae (1) and (2), that we expect to generalize when we further differentiate? We propose a language that shows a good promise to simplify the description of the higher order derivatives of spectral functions. It is based on the idea of generalizing the Hadamard product of two matrices to a k-tensor valued product between k matrices. The current paper is the first of three. It defines what we mean by a generalized Hadamard product and investigates some of its multi-linear algebraic properties. In [14] we will formulate calculus-type rules for the interaction between the generalized Hadamard product and the eigenvalues of symmetric matrices. Finally, in [15] we will describe how to compute the derivatives of spectral functions in two important cases. In particular, we will show that Conjecture 4.1 holds for the derivatives of any spectral function at a symmetric matrix with distinct eigenvalues, as well as for the derivatives of separable spectral functions at an arbitrary symmetric matrix. (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.)

2 Generalizations of the Hadamard product

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

Let us look closely at the Hadamard product, $H_1 \circ H_2$, between two matrices H_1 and H_2 from M^n. It is a matrix valued function on two matrix arguments, linear in each argument separately. Therefore 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:

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

Naturally, we may define the *cross Hadamard product* by the rule

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

and then extend this to a bilinear function on all $M^n \times M^n$. The Hadamard product and the cross Hadamard product are essentially the same thing:

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

These observations can be naturally generalized in the following way. Denote the set $\{1, 2, ..., k\}$ of the first k natural numbers by $\mathbb{N}_k$. 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 will index the elements of a tensor just like we index the entries of a matrix. The space of all k-tensors on $\mathbb{R}^n$ will be denoted by $T^{k,n}$.

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

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

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. 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. Using the definition that 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*

$$\begin{aligned}
&(H_{p_1q_1} \circ_\sigma \cdots \circ_\sigma H_{p_{k-1}q_{k-1}} \circ_\sigma H)^{i_1 i_2 \ldots i_k} \\
&= \delta_{i_1 p_1} \delta_{i_1 q_{\sigma(1)}} \cdots \delta_{i_{l-1} p_{l-1}} \delta_{i_{l-1} q_{\sigma(l-1)}} H^{i_{\sigma(l)} i_l} \delta_{i_{l+1} p_{l+1}} \delta_{i_{l+1} q_{\sigma(l+1)}} \cdots \delta_{i_k p_k} \delta_{i_k q_{\sigma(k)}},
\end{aligned}$$

where $l = \sigma^{-1}(k)$.

Proof. Let $l = \sigma^{-1}(k)$, using the result of the previous lemma we calculate.

$$\begin{aligned}
&(H_{p_1q_1} \circ_\sigma \cdots \circ_\sigma H_{p_{k-1}q_{k-1}} \circ_\sigma H)^{i_1 i_2 \ldots i_k} = H_{p_1 q_1}^{i_1 i_{\sigma^{-1}(1)}} \cdots H_{p_{k-1} q_{k-1}}^{i_{k-1} i_{\sigma^{-1}(k-1)}} H^{i_k i_{\sigma^{-1}(k)}} \\
&= \delta_{i_1 p_1} \delta_{i_{\sigma^{-1}(1)} q_1} \cdots \delta_{i_{k-1} p_{k-1}} \delta_{i_{\sigma^{-1}(k-1)} q_{k-1}} H^{i_k i_{\sigma^{-1}(k)}} \\
&= \delta_{i_1 p_1} \delta_{i_1 q_{\sigma(1)}} \cdots \delta_{i_{l-1} p_{l-1}} \delta_{i_{l-1} q_{\sigma(l-1)}} H^{i_{\sigma(l)} i_l} \delta_{i_{l+1} p_{l+1}} \delta_{i_{l+1} q_{\sigma(l+1)}} \cdots \delta_{i_k p_k} \delta_{i_k q_{\sigma(k)}}. \quad \blacksquare
\end{aligned}$$

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

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 The σ-Hadamard product has meaning even when $k = 1$. In that case, there is just one permutation on the set $\mathbb{N}_1$ and the σ-Hadamard product corresponding to it has one matrix argument and returns, by definition, a vector (1-tensor). Since $\sigma = (1)$, extending the notation, the σ-Hadamard product is given by the rule:

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

Extending by linearity we get

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

For any two k-tensors, T_1, and T_2 we define a scalar product between them in the natural way:

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

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.

$$\begin{aligned}
&\langle T, H_{p_1q_1} \circ_\sigma H_{p_2q_2} \circ_\sigma \cdots \circ_\sigma H_{p_{k-1}q_{k-1}} \circ_\sigma H \rangle \\
&= \sum_{p_k,q_k=1}^{n,n} H^{p_kq_k} \langle T, H_{p_1q_1} \circ_\sigma H_{p_2q_2} \circ_\sigma \cdots \circ_\sigma H_{p_{k-1}q_{k-1}} \circ_\sigma H_{p_kq_k} \rangle \\
&= \sum_{p_k,q_k=1}^{n,n} H^{p_kq_k} \sum_{i_1,\ldots,i_k=1}^{n,\ldots,n} T^{i_1\ldots i_k} (H_{p_1q_1} \circ_\sigma H_{p_2q_2} \circ_\sigma \cdots \circ_\sigma H_{p_{k-1}q_{k-1}} \circ_\sigma H_{p_kq_k})^{i_1\ldots i_k} \\
&= \sum_{p_k,q_k=1}^{n,n} H^{p_kq_k} \sum_{i_1,\ldots,i_k=1}^{n,\ldots,n} T^{i_1\ldots i_k} \delta_{i_1p_1} \delta_{p_1q_{\sigma(1)}} \cdots \delta_{i_kp_k} \delta_{p_kq_{\sigma(k)}} \\
&= \sum_{p_k,q_k=1}^{n,n} H^{p_kq_k} T^{p_1\ldots p_k} \delta_{p_1q_{\sigma(1)}} \cdots \delta_{p_kq_{\sigma(k)}}.
\end{aligned}$$

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

3 A partial order on P^k and one 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.

In other words σ *refines* μ if every cycle of σ is contained in a cycle of μ. Clearly the cycles of σ will partition the cycles of μ. If σ refines μ we will denote it by

$$\mu \preceq \sigma.$$

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. Clearly the refinement 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 that 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 $k = 3$ $\mathcal{D}((123)) = \mathcal{D}((132)) = \{x \in \mathbb{R}^3 \,:\, x_1 = x_2 = x_3\}$. Clearly the image of the identity permutation is $\mathbb{R}^k$. The following relationship helps to visualize the partial order on P^k

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

Finally, given a tensor $T \in T^{k,n}$ we may want to preserve the entries lying on a diagonal "subspace" of T and substitute the rest of the entries of T with zeros. In other words, given a permutation $\mu \in P^k$, we introduce the notation $P_\mu(T)$ for the tensor in $T^{k,n}$ defined by

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

After all these preparations, we can formulate the main result in this section. It describes when we 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*

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

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 are going to 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 for any $T \in T^{k,n}$ if, and only if, $\mu \preceq \sigma_2^{-1} \circ \sigma_1$. Direct calculation shows:

$$\begin{aligned} &\langle P_\mu(T), H_{p_1q_1} \circ_{\sigma_1} \cdots \circ_{\sigma_1} H_{p_kq_k} \rangle \\ &\qquad = \sum_{i_1,\ldots,i_k=1}^{n,\ldots,n} (P_\mu(T))^{i_1\ldots i_k} (H_{p_1q_1} \circ_{\sigma_1} \cdots \circ_{\sigma_1} H_{p_kq_k})^{i_1\ldots i_k} \\ &\qquad = \sum_{i_1,\ldots,i_k=1}^{n,\ldots,n} (P_\mu(T))^{i_1\ldots i_k} H_{p_1q_1}^{i_1 i_{\sigma_1^{-1}(1)}} \cdots H_{p_kq_k}^{i_k i_{\sigma_1^{-1}(k)}} \\ &\qquad = \sum_{i_1,\ldots,i_k=1}^{n,\ldots,n} (P_\mu(T))^{i_1\ldots i_k} \delta_{i_1p_1}\delta_{i_1q_{\sigma_1(1)}} \cdots \delta_{i_kp_k}\delta_{i_kq_{\sigma_1(k)}} \\ &\qquad = (P_\mu(T))^{p_1\ldots p_k} \delta_{p_1q_{\sigma_1(1)}} \cdots \delta_{p_kq_{\sigma_1(k)}}. \end{aligned}$$

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

Analogously we have

$$\langle P_\mu(T), H_{p_1q_1} \circ_{\sigma_2} \cdots \circ_{\sigma_2} H_{p_kq_k} \rangle = (P_\mu(T))^{p_1\ldots p_k} \delta_{p_1q_{\sigma_2(1)}} \cdots \delta_{p_kq_{\sigma_2(k)}},$$

which is equal to $T^{p_1\ldots p_k}$ 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 expressions are zero and we trivially have equality.

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 $s \in \mathbb{N}_k$. 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. This is a contradiction. Thus for some $s_1 \in \mathbb{N}_k$ we have $p_{s_1} \neq q_{\sigma_2(s_1)}$ and again we will have that both expressions 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 first expression is equal to $T^{p_1 \dots p_k}$. If we show that $p_s = q_{\sigma_2(s)}$ for every $s \in \mathbb{N}_k$, then we will be done. 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 will have $p_{\sigma_2^{-1}(r_0)} \neq q_{r_0}$, and for $s_1 = \sigma_1^{-1}(r_0)$ we have $p_{\sigma_2^{-1}(\sigma_1(s_1))} \neq q_{\sigma_1(s_1)}$. Again $\mu \preceq \sigma_2^{-1} \circ \sigma_1$ 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

$$(P_\mu(T))^{p_1 \dots p_k} \delta_{p_1 q_{\sigma_1(1)}} \cdots \delta_{p_k q_{\sigma_1(k)}} = (P_\mu(T))^{p_1 \dots p_k} \delta_{p_1 q_{\sigma_2(1)}} \cdots \delta_{p_k q_{\sigma_2(k)}}, \tag{4}$$

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 $i_1, \dots, i_k$ 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 an 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 $p_{\sigma_2^{-1}(\sigma_1(s_0))} \neq p_{s_0}$.

With the choices so made, the left-hand side of Equation (4) will be equal to $T^{i_1 \dots i_k} \neq 0$. We will reach a contradiction if we show that for some r_0, $p_{r_0} \neq q_{\sigma_2(r_0)}$, since then the right-hand side of Equation (4) will be zero. Suppose on the contrary that $p_r = q_{\sigma_2(r)}$ for every $r \in \mathbb{N}_k$. Then,

$$p_{\sigma_2^{-1}(\sigma_1(s))} = q_{\sigma_1(s)} = p_s, \quad \text{for every } s \in \mathbb{N}_k.$$

Substitute above $s = s_0$ to reach a contradiction. Thus, $p_{r_0} \neq q_{\sigma_2(r_0)}$ for some $r_0 \in N_k$ and 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^n such that $\mu \preceq \nu$. Then for arbitrary permutation $\sigma \in P^k$, any matrices $H_1,\ldots,H_k$, 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 will be useful to see what are 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 In this example we demonstrate that Formula (1) for the first derivative of a spectral function, at X, can be rewritten in a different form. Let $X = V(\mathrm{Diag}\,\lambda(X))V^T$ and $\tilde{E} = V^T E V$, where E is a symmetric matrix. Using the definitions and notation in the previous subsection we have:

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

Example 3.5 Let X be a symmetric matrix with ordered spectral decomposition $X = V(\operatorname{Diag}\lambda(X))V^T$. 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 the examples in Section 2 we have:

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

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

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

All these examples support the following conjecture, which describes the structure of the higher-order derivatives of spectral functions.

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

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

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

In [15] we will show that this conjecture holds for the derivatives of any spectral function at a symmetric matrix X with distinct eigenvalues, as well as for the derivatives of separable spectral functions at an arbitrary symmetric matrix. (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 to compute the operators $\mathcal{A}_\sigma$ for every σ in P^k.

There is one major draw-back of the conjectured formula above. On the left hand-side we have the the k-th derivative of the spectral function evaluated at the matrices $E_1, ..., E_k$ while on the right-hand side these matrices are "jumbled" with the orthogonal matrix V into the σ-Hadamard products $\tilde{E}_1 \circ_\sigma \cdots \circ_\sigma \tilde{E}_k$. This is the problem that we address in the next section.

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

$$\langle \mathrm{Diag}\, x, H \rangle = \langle x, \mathrm{diag}\, H \rangle, \tag{5}$$

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

$$\langle U(\mathrm{Diag}\, x)U^T, H \rangle = \langle x, \mathrm{diag}\,(U^T H U) \rangle. \tag{6}$$

Vector x can be viewed as a 1-tensor on $\mathbb{R}^n$ given through the linear isometry $x \to \langle x, \cdot \rangle$ and similarly $\mathrm{Diag}\, x$ can be viewed as a 2-tensor. In this section we will generalize Equations (5) and (6) for an arbitrary k-tensor in place of x and arbitrary σ-Hadamard product in place of diag.

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

$$(\mathrm{Diag}^{\sigma} T)^{i_1 \dots i_k}_{j_1 \dots j_k} = \begin{cases} T^{i_1 \dots i_k}, & \text{if } i_s = j_{\sigma(s)}, \forall s = 1, \dots, k, \\ 0, & \text{otherwise.} \end{cases}$$

When $k = 1$ and σ is the only choice from P^1, namely $\sigma = (1)$, then this definition coincides with the definition of the Diag operator in Equation (5). Equivalent way to define $\mathrm{Diag}^{\sigma} T$ that is useful for calculations is:

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

We now define an action 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:

$$(UTU^T)^{i_1 \dots i_k} = \sum_{p_1 = 1}^{n} \cdots \sum_{p_k = 1}^{n} \Big(T^{p_1 \dots p_k} U^{i_1 p_1} \cdots U^{i_k p_k} \Big). \tag{7}$$

In the case $k = 1$, when T is viewed as an n-dimensional vector, this is exactly the action of the orthogonal group on $\mathbb{R}^n$:

$$(UTU^T)^{i_1} \equiv (UT)^{i_1} = \sum_{p_1 = 1}^{n} U^{i_1 p_1} T^{p_1}.$$

In the case $k = 2$ the definition coincides with the conjugate action of the orthogonal group on the set of all $n \times n$ square matrices:

$$(UTU^T)^{ij} = \sum_{p,q=1}^{n,n} T^{pq} U^{ip} U^{jq},$$

hence the use of the same notation for the general action UTU^T. For future reference we state the formula of the action in the case when the tensor is of even order. That is, if T is a $2k$-tensor, then

$$(UTU^T)^{i_1 \ldots i_k}_{j_1 \ldots j_k} = \sum_{p_1,q_1=1}^{n,n} \cdots \sum_{p_k,q_k=1}^{n,n} \Big(T^{p_1 \ldots p_k}_{q_1 \ldots q_k} \prod_{\nu=1}^{k} U^{i_\nu p_\nu} U^{j_\nu q_\nu} \Big). \tag{8}$$

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 will be given by:

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

That is, the conjugate action of a permutation matrix on a k-tensor is what one expects it to be. We have the following immediate observation.

Lemma 4.1 *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.

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

■

A natural question to ask is whether the action defined above on the space $T^{k,n}$ is associative.

Lemma 4.2 *for any k-tensor, T, on $\mathbb{R}^n$ and any two orthogonal matrices U, V in O^n we have*

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

Proof. The proof is a direct calculation using the definitions. On one hand we have

$$(V(UTU^T)V^T)^{i_1\dots i_k} = \sum_{p_1=1}^{n}\cdots\sum_{p_k=1}^{n}\Big((UTU^T)^{p_1\dots p_k}\prod_{\nu=1}^{k}V^{i_\nu p_\nu}\Big)$$

$$= \sum_{p_1=1}^{n}\cdots\sum_{p_k=1}^{n}\Big(\Big(\sum_{l_1=1}^{n}\cdots\sum_{l_k=1}^{n}T^{l_1\dots l_k}\prod_{\mu=1}^{k}U^{p_\mu l_\mu}\Big)\prod_{\nu=1}^{k}V^{i_\nu p_\nu}\Big).$$

On the other hand we have

$$((VU)T(VU)^T)^{i_1\dots i_k} = \sum_{l_1=1}^{n}\cdots\sum_{l_k=1}^{n}T^{l_1\dots l_k}\prod_{\mu=1}^{k}(VU)^{i_\mu l_\mu}.$$

Using that

$$(VU)^{i_\mu l_\mu} = \sum_{p_\mu=1}^{n}V^{i_\mu p_\mu}U^{p_\mu l_\mu},$$

we get

$$\prod_{\mu=1}^{k}(VU)^{i_\mu l_\mu} = \prod_{\mu=1}^{k}\Big(\sum_{p_\mu=1}^{n}V^{i_\mu p_\mu}U^{p_\mu l_\mu}\Big) = \sum_{p_1=1}^{n}\cdots\sum_{p_k=1}^{n}\Big(\prod_{\mu=1}^{k}V^{i_\mu p_\mu}U^{p_\mu l_\mu}\Big).$$

Putting everything together and observing that we can exchange the multiple sum $\sum_{p_1=1}^{n}\cdots\sum_{p_k=1}^{n}$ with the multiple sum $\sum_{l_1=1}^{n}\cdots\sum_{l_k=1}^{n}$ we finish the proof of the lemma. ■

Let us see now that conjugation with an orthogonal matrix is orthogonal transformation on $T^{k,n}$. That is, it doesn't change the norm of the tensor. In other words, if we define

$$\|T\| := \sqrt{\langle T, T\rangle},$$

then we have the following lemma.

Lemma 4.3 *Let T be a k-tensor on $\mathbb{R}^n$, and U be any orthogonal matrix in O^n, then*

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

Proof. Direct calculation of the quantity $\|UTU^T\|^2$ gives:

$$\begin{aligned}
&\|UTU^T\|^2 = \langle UTU^T, UTU^T \rangle \\
&= \sum_{i_1,\ldots,i_k=1}^{n,\ldots,n} (UTU^T)^{i_1\ldots i_k} (UTU^T)^{i_1\ldots i_k} \\
&= \sum_{i_1,\ldots,i_k=1}^{n,\ldots,n} \Big(\sum_{p_1,\ldots,p_k=1}^{n,\ldots,n} T^{p_1\ldots p_k} U^{i_1 p_1} \cdots U^{i_k p_k} \Big) \Big(\sum_{q_1,\ldots,q_k=1}^{n,\ldots,n} T^{q_1\ldots q_k} U^{i_1 q_1} \cdots U^{i_k q_k} \Big) \\
&= \sum_{p_1,\ldots,p_k=1}^{n,\ldots,n} \sum_{q_1,\ldots,q_k=1}^{n,\ldots,n} T^{p_1\ldots p_k} T^{q_1\ldots q_k} \Big(\sum_{i_1=1}^{n} U^{i_1 p_1} U^{i_1 q_1} \Big) \cdots \Big(\sum_{i_k=1}^{n} U^{i_k p_k} U^{i_k q_k} \Big) \\
&= \sum_{p_1,\ldots,p_k=1}^{n,\ldots,n} \sum_{q_1,\ldots,q_k=1}^{n,\ldots,n} T^{p_1\ldots p_k} T^{q_1\ldots q_k} \delta_{p_1 q_1} \cdots \delta_{p_k q_k} \\
&= \sum_{p_1,\ldots,p_k=1}^{n,\ldots,n} (T^{p_1\ldots p_k})^2 \\
&= \|T\|^2. \qquad \blacksquare
\end{aligned}$$

After all these preparations, we can give the following generalization to Equation (6). (When, $k = 1$ and $\sigma = (1)$ we obtain exactly Equation (6).)

Theorem 4.4 *For any k-tensor T, any matrices $H_1,\ldots,H_k$, any orthogonal matrix U, and any permutation σ on $\mathbb{N}_k$ we have the identity*

$$\langle T, \tilde{H}_1 \circ_\sigma \cdots \circ_\sigma \tilde{H}_k \rangle = \big(U(\mathrm{Diag}\,{}^\sigma T)U^T \big)[H_1, \ldots, H_k], \tag{9}$$

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 $\tilde{H}_{ij}^{pq} = U^{ip} U^{jq}$, we develop the left-hand side of Equation (9):

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

$$= \sum_{p_1,\dots,p_k=1}^{n,\dots,n} T^{p_1\dots p_k} U^{i_1 p_1} U^{j_1 p_{\sigma^{-1}(1)}} \cdots U^{i_k p_k} U^{j_k p_{\sigma^{-1}(k)}}.$$

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

$$\begin{aligned}
&(U(\mathrm{Diag}^{\,\sigma} T)U^T)[H_{i_1 j_1}, ..., H_{i_k j_k}] = \\
&\qquad = \sum_{p_1,q_1=1}^{n,n} \cdots \sum_{p_k,q_k=1}^{n,n} \Big((U(\mathrm{Diag}^{\,\sigma} T)U^T)^{p_1\dots p_k}_{q_1\dots q_k} H^{p_1 q_1}_{i_1 j_1} \cdots H^{p_k q_k}_{i_k j_k} \Big) \\
&\qquad = (U(\mathrm{Diag}^{\,\sigma} T)U^T)^{i_1\dots i_k}_{j_1\dots j_k} \\
&\qquad = \sum_{p_1,q_1=1}^{n,n} \cdots \sum_{p_k,q_k=1}^{n,n} \Big((\mathrm{Diag}^{\,\sigma} T)^{p_1\dots p_k}_{q_1\dots q_k} \prod_{\nu=1}^{k} U^{i_\nu p_\nu} U^{j_\nu q_\nu} \Big) \\
&\qquad = \sum_{p_1=1}^{n} \cdots \sum_{p_k=1}^{n} \Big(T^{p_1\dots p_k} \prod_{\nu=1}^{k} U^{i_\nu p_\nu} U^{j_\nu p_{\sigma^{-1}(\nu)}} \Big).
\end{aligned}$$

This shows that the both sides are equal. ∎

If we take the orthogonal matrix U to be the identity matrix we obtain the following corollary.

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

$$\langle T, H_1 \circ_\sigma ... \circ_\sigma H_k \rangle = (\mathrm{Diag}^{\,\sigma} T)[H_1, ..., H_k]. \tag{10}$$

If in Corollary 4.5 we substitute the matrices H_1,...,H_k with $\tilde{H}_1$,...,$\tilde{H}_k$ and we use Theorem 4.4, we obtain the next result.

Corollary 4.6 *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*

$$(\mathrm{Diag}^{\,\sigma} T)[\tilde{H}_1, ..., \tilde{H}_k] = \big(U(\mathrm{Diag}^{\,\sigma} T)U^T\big)[H_1, ..., H_k]. \tag{11}$$

If in Corollary 4.5 we take σ to be the identity permutation, then we get the next corollary, which generalizes Equation (5).

Corollary 4.7 *For any k-tensor T, any matrices $H_1,\dots,H_k$ we have the identity*

$$T[\operatorname{diag} H_1, \dots, \operatorname{diag} H_k] = (\operatorname{Diag}^{(\mathrm{id})} T)[H_1, \dots, H_k]. \tag{12}$$

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

Example 4.8 As we saw in Example 3.4, the first derivative of the spectral function $f \circ \lambda$ at the point $X = V(\operatorname{Diag} \lambda(X))V^T$, applied to the symmetric matrix E is given by the formula

$$\nabla(f \circ \lambda)(X)[E] = \langle \nabla f(\mu), \circ_{(1)} \tilde{E} \rangle,$$

where $\tilde{E} = V^T E V$. This formula can be rewritten as

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

This was essentially the original form of this formula given in Equation (1).

The usefulness of the new notation becomes more evident below.

Example 4.9 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 $\mathcal{A}_\sigma : \mathbb{R}^n \to T^{k,n}$, $\sigma \in P^k$, such that

$$\nabla^k (f \circ \lambda)(X) = V\Big(\sum_{\sigma \in P^k} \operatorname{Diag}^{\sigma} \mathcal{A}_\sigma(\lambda(X))\Big) V^T, \tag{13}$$

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

In [15] we will show that this conjecture holds for the derivatives of any spectral function at a symmetric matrix X with distinct eigenvalues, as well as for the derivatives of separable spectral functions at an arbitrary symmetric matrix. (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 $\mathcal{A}_\sigma$, that depend only on the symmetric function f.

5 Comments on Conjecture 4.1

In this section we show that once Conjecture 4.1 is established for $k = 1$, then for $k \geq 2$ it is enough to prove it only in the case when the $X = \operatorname{Diag} x$ for some $x \in \mathbb{R}^n$ with $x_1 \geq \cdots \geq x_n$. 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 $T[\cdot, ..., H]$.

Lemma 5.1 *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

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

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

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 will 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

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

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.

Proposition 5.2 *Suppose the k-th derivative of the spectral function $F = f\circ\lambda$ is given by Equation (13) 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.) Clearly, we have $X = V(\mathrm{Diag}\,\lambda(X))V^T$, and by continuity of eigenvalues $\lambda(X_m)$ approaches $\lambda(X)$. Using the formula for the k-th derivative and the continuity of the tensorial maps, the contradiction follows. $\blacksquare$

6 Equivalence relations on $\mathbb{N}_n$

Suppose that $\sim$ is an equivalence relation on the integers $\mathbb{N}_n$ and denote by I_1, I_2,...,I_r the equivalence classes determined by $\sim$. The equivalence classes will be also called *blocks*. One may assume that the blocks are numbered so that I_1 contains the integer 1, I_2 contains that smallest integer not in I_1, I_3 contains the smallest integer not in $I_1 \cup I_2$, and so on.

In this short section, we will be interested in tensors having the following structure.

Definition 6.1 We say that a tensor $T \in T^{k,n}$ is *block-constant* (with respect to the equivalence relation $\sim$) if

$$T^{i_1 \ldots i_k} = T^{j_1 \ldots j_k}, \text{ whenever } i_s \sim j_s \text{ for all } s = 1, 2, ..., k.$$

Let μ be an arbitrary but fixed permutation in P^k. We introduce the linear operator $\tilde{P}_\mu$ on the space $T^{k,n}$, generalizing the operator P_μ defined in Section 3. The definition is element-wise, as follows:

$$\left(\tilde{P}_\mu(T)\right)^{i_1 \ldots i_k} := \begin{cases} T^{i_1 \ldots i_k}, & \text{if } i_s \sim i_{\mu(s)} \; \forall s \in \mathbb{N}_k, \\ 0, & \text{otherwise.} \end{cases}$$

Clearly, when the equivalence relation $\sim$ is such that $i \sim j$ if, and only if, $i = j$, then $\tilde{P}_\mu$ becomes equal to the previously defined P_μ. We would like to conclude this work with a generalization of Theorem 3.1.

Theorem 6.2 *Let σ_1, σ_2, and μ be three permutations in P^k. Then for any block-constant matrices H_1,...,H_k, and any tensor T in $T^{k,n}$ we have the identity:*

$$\langle \tilde{P}_\mu(T), H_1 \circ_{\sigma_1} \cdots \circ_{\sigma_1} H_k \rangle = \langle \tilde{P}_\mu(T), H_1 \circ_{\sigma_2} \cdots \circ_{\sigma_2} H_k \rangle$$

if, and only if, $\mu \preceq \sigma_2^{-1} \circ \sigma_1$.

Proof. The proof is completely analogous to the one of Theorem 3.1. Consider a basis for the space of block constant matrices $\{H_{pq} : 1 \leq p, q \leq n\}$ such that H_{pq}^{ij} is equal to one, if $i \sim j$, and zero otherwise. Then all we have to change in the proof of Theorem 3.1 is the "=" signs between indexes with "$\sim$" signs and all "$\neq$" signs with "$\nsim$". ■

References

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

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

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

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

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

[6] A.S. LEWIS. Derivatives of spectral functions. *Mathematics of Operations Research*, 21:576–588, 1996.

[7] A.S. LEWIS. Nonsmooth analysis of eigenvalues. *Mathematical Programming*, 84:1–24, 1999.

[8] A.S. LEWIS. The mathematics of eigenvalue optimization. *Mathematical Programming, Series B*, 97:155–176, 2003.

[9] A.S. LEWIS and M.L. OVERTON. Eigenvalue optimization. *Acta Numerica*, 5:149–190, 1996.

[10] A.S. LEWIS and H.S. SENDOV. Twice differentiable spectral functions. *SIAM Journal of Matrix Analysis and Applications*, 23(2):368–386, 2001.

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

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

[13] A. SEEGER. Convex analysis of spectrally defined matrix functions. *SIAM Journal on Optimization*, 7(3):679–696, 1997.

[14] 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*, 2004. Mathematical Series 2004-309.

[15] 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*, 2004. Mathematical Series 2004-310.

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

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

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

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