On the higher-order derivatives of spectral functions

Hristo S. Sendov^{*}

May 21, 2006

Abstract

In this paper we are interested in the higher-order derivatives of functions of the eigenvalues of symmetric matrices with respect to the matrix argument. We describe the formula for the k-th derivative of such functions in two general cases.

The first case concerns the derivatives of the composition of an arbitrary (not necessarily symmetric) k-times differentiable function with the eigenvalues of symmetric matrices at a symmetric matrix with distinct eigenvalues.

The second case describes the derivatives of the composition of a k-times differentiable *separable* symmetric function with the eigenvalues of symmetric matrices at an arbitrary symmetric matrix. We show that the formula significantly simplifies when the separable symmetric function is k-times continuously differentiable.

As an application of the developed techniques, we re-derive the formula for the Hessian of a general *spectral* function at an arbitrary symmetric matrix. The new tools lead to a shorter, cleaner derivation than the original one in [16].

To make the exposition as self contained as possible, we have included the necessary background results and definitions. The proofs of the intermediate technical results are collected in the appendices.

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

1 Introduction

We say that a real-valued function F of a real symmetric matrix argument is *spectral* if

$$F(UXU^T) = F(X)$$

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

for every real symmetric matrix X in its domain and every orthogonal matrix U. That is, F(X) = F(Y) if X and Y are symmetric and if X is similar to Y. The restriction of F to the subspace of diagonal matrices defines a function f(x) = F(Diag x) on a vector argument $x \in \mathbb{R}^n$. It is easy to see that the function $f : \mathbb{R}^n \to \mathbb{R}$ is symmetric, that is, has the property

f(x) = f(Px) for any permutation matrix P and any x in the domain of f,

and in addition, $F(X) = (f \circ \lambda)(X)$, in which the *eigenvalue map* $\lambda(X) = (\lambda_1(X), ..., \lambda_n(X))$ is the vector of eigenvalues of X ordered in nonincreasing order.

One of the main questions in the theory of spectral functions is what smoothness properties of the symmetric function f are inherited by F. The difficulties arise from the fact that the eigenvalue map $\lambda(X)$ is continuous but not always differentiable with respect to X. In domains, where $\lambda(X)$ is differentiable, it is difficult to organize the differentiation process so that one arrives at an elegant formula for the higher-order derivatives of $(f \circ \lambda)(X)$.

An important subclass of spectral functions is obtained when $f(x) = g(x_1) + \cdots + g(x_n)$ for some function g of one real variable. We call such symmetric functions *separable*; their corresponding spectral functions are called *separable spectral* functions.

In [12] an explicit formulae for the gradient of the spectral function F in terms of the derivatives of the symmetric function f was given:

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

where V is any orthogonal matrix such that $X = V(\text{Diag }\lambda(X))V^T$ is the ordered spectral decomposition of X. In [16] a formula for the Hessian of F was given, whose structure appeared quite different from the one for the gradient. Calculating the third and higher-order derivatives of F becomes unmanageable without an appropriate language for describing them.

In this work we generalize the work in [12] and [16] by proving, in two general cases, the following formula for the k-th derivative of a spectral function

(2)
$$\nabla^k (f \circ \lambda)(X) = V \Big(\sum_{\sigma \in P^k} \operatorname{Diag}{}^{\sigma} \mathcal{A}_{\sigma}(\lambda(X)) \Big) V^T,$$

where again $X = V(\text{Diag }\lambda(X))V^T$. The sum is taken over all permutations on k elements, which are a convenient tool for enumerating the maps $\mathcal{A}_{\sigma}(x)$. The precise meanings of the operators Diag^{σ} and the conjugation by the orthogonal matrix V are explained in the next section; see (6) and (9) respectively. The maps $\mathcal{A}_{\sigma}(x)$ depend only on the partial derivatives of f(x), up to order k, and do not depend on the eigenvalues. Thus, it is easy to see how the higher-order derivatives depend on the eigenvalue map $\lambda(X)$. Formula (2) depends on the eigenvalues only through the compositions $\mathcal{A}_{\sigma}(\lambda(X))$ and the conjugation by the orthogonal matrix V.

We show that (2) holds in two general cases. It holds when f is a k-times (continuously) differentiable function, not necessarily symmetric, and X is a matrix with distinct eigenvalues. It also holds when f is a k-times (continuously) differentiable separable symmetric function and X is an arbitrary symmetric matrix. We give an easy recipe for computing the maps $\mathcal{A}_{\sigma}(x)$ in these two cases.

Our results for separable spectral functions imply those of [5] and [4] for one-parameter families of symmetric matrices; see also the monographs [8] and [9]. Our results also generalize and extend those in [20] when the considerations there are restricted to the space of symmetric matrices. (Notice that the gradients of separable spectral functions, see (1), are the functions considered in [20] when restricted to the space of symmetric matrices.) For example, Theorem 4.1 in [20] assumes that the function f is k-times continuously differentiable to conclude that $(f \circ \lambda)(X(t))$ is k-times differentiable, where X(t) is a k-times differentiable path of symmetric matrices depending on the scalar parameter t. In Theorem 6.1 we only assume that f is k-times differentiable and obtain that $(f \circ \lambda)(X)$ is k-times differentiable with respect to the free symmetric matrix variable X. In that case, one can obtain the derivatives of $(f \circ \lambda)(X(t))$ by using the Chain Rule. Finally, Theorem 6.9 shows that if f is k-times continuously differentiable then $(f \circ \lambda)(X)$ is k-times continuously differentiable then $(f \circ \lambda)(X)$ is k-times continuously differentiable with respect to the variable X.

In addition, we show that if f is a k-times continuously differentiable, separable symmetric function, (2) can be significantly simplified. In that case, if σ_1 and σ_2 are two permutations on k elements with one cycle in their cycle decomposition then $\mathcal{A}_{\sigma_1}(x) = \mathcal{A}_{\sigma_2}(x)$ and these maps allow a simple determinant description. If σ has more than one cycle, then $\mathcal{A}_{\sigma}(x) \equiv 0$.

In Section 7, we re-derive the formula for the Hessian of a general spectral function at an arbitrary symmetric matrix. The techniques developed here lead to a shorter, more streamlined derivation than the original derivation in [16].

The language that we use, based on the generalized Hadamard product, allows us to differentiate (2) just as one would expect: writing the differential quotient and taking the limit as the perturbation goes to zero. This gives a clear view of where the different pieces in the differential come from and gives the process a routine Calculus-like flavour.

In the next section, we give the necessary notation, definitions, and background results to facilitate the reading of this work. The proofs of the technical tools are in the appendices.

2 Notation and background results

By \mathbb{R}^n we denote the standard *n*-dimensional Euclidean space of *n*-tuples of real numbers with standard inner product and norm. By S^n , O^n , and P^n we denote the sets of all $n \times n$ real symmetric, orthogonal, and permutation matrices, respectively. By M^n we denote the real Euclidean space of all $n \times n$ matrices with inner product $\langle X, Y \rangle = \text{tr}(XY^T)$ and corresponding norm $||X|| = \sqrt{\langle X, X \rangle}$. For $A \in S^n$, $\lambda(A) = (\lambda_1(A), ..., \lambda_n(A))$ is the vector of its eigenvalues ordered in nonincreasing order:

$$\lambda_1(A) \ge \lambda_2(A) \ge \cdots \ge \lambda_n(A).$$

By \mathbb{N}_k we denote the set $\{1, 2, ..., k\}$. For any vector x in \mathbb{R}^n , Diag x denotes the diagonal matrix with the vector x on the main diagonal, and diag: $M^n \to \mathbb{R}^n$ denotes its adjoint operator, defined by diag $(X) = (x_{11}, ..., x_{nn})$. By $\mathbb{R}^n_{\downarrow}$ we denote the cone of all vectors x in \mathbb{R}^n such that $x_1 \ge x_2 \ge$ $\cdots \ge x_n$. Denote the standard orthonormal basis in \mathbb{R}^n by $e^1, e^2, ..., e^n$. For a permutation matrix $P \in P^n$ we say that $\sigma : \mathbb{N}_n \to \mathbb{N}_n$ is its corresponding permutation map if for any $h \in \mathbb{R}^n$ we have $Ph = (h_{\sigma(1)}, ..., h_{\sigma(n)})^T$, that is, $P^T e^i = e^{\sigma(i)}$ for all i = 1, ..., n. The symbol δ_{ij} denotes the Kroneker delta. It is equal to one if i = j and zero otherwise.

Any vector $\mu \in \mathbb{R}^n$ defines a partition of \mathbb{N}_n into disjoint *blocks*, where integers *i* and *j* are in the same block if and only if $\mu_i = \mu_j$. In general, the blocks that μ determines need not contain consecutive integers. We agree that the block containing the integer 1 is the first block, I_1 , the block containing the smallest integer that is not in I_1 is the second block, I_2 , and so on. By *r* we denote the number of blocks in the partition. For any two integers, $i, j \in \mathbb{N}_n$ we say that they are *equivalent (with respect to* μ) and write $i \sim j$ (or $i \sim_{\mu} j$) if $\mu_i = \mu_j$, that is, if they are in the same block. Two *k*-indexes $(i_1, ..., i_k)$ and $(j_1, ..., j_k)$ are called *equivalent* if $i_l \sim j_l$ for all l = 1, 2, ..., k, and we write $(i_1, ..., i_k) \sim (j_1, ..., j_k)$ (or $(i_1, ..., i_k) \sim_{\mu} (j_1, ..., j_k)$).

A *k*-tensor on a linear space is a real-valued function of *k* arguments from the linear space, that is linear in each argument separately. Denote the set of all *k*-tensors on \mathbb{R}^n by $T^{k,n}$. The value of the *k*-tensor at $(h_1, ..., h_k)$ is denoted by $T[h_1, ..., h_k]$. For any $(i_1, ..., i_k)$, a *k*-tuple of integers from \mathbb{N}_n , we denote by $T^{i_1...i_k}$ the value $T[e^{i_1}, ..., e^{i_k}]$. Matrices from M^n are viewed as 2-tensors on \mathbb{R}^n , with respect to the fixed basis, and for an $M \in M^n$ we have $M^{ij} = M[e^i, e^j] := \langle e^i, Me^j \rangle$.

The next elementary lemma motivates the following definitions. It is a simple application of the chain rule to the equality $f(\mu) = f(P\mu)$.

Lemma 2.1 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a symmetric function, k times differentiable at the point $\mu \in \mathbb{R}^n$, and let P be a permutation matrix such that $P\mu = \mu$. Then

- (i) $\nabla f(\mu) = P^T \nabla f(\mu)$,
- (ii) $\nabla^2 f(\mu) = P^T \nabla^2 f(\mu) P$, and in general
- (*iii*) $\nabla^s f(\mu)[h_1, ..., h_s] = \nabla^s f(\mu)[Ph_1, ..., Ph_s]$, for any $h_1, ..., h_s \in \mathbb{R}^n$, and $s \in \mathbb{N}_k$.

Definition 2.2 A tensor $T \in T^{k,n}$ is called *symmetric* if for any permutation σ on \mathbb{N}_k it satisfies

$$T[h_{\sigma(1)}, ..., h_{\sigma(k)}] = T[h_1, ..., h_k],$$

for any $h_1, ..., h_k \in \mathbb{R}^n$.

Definition 2.3 (i) Given a vector $\mu \in \mathbb{R}^n$, a tensor $T \in T^{k,n}$ is called *point-symmetric with* respect to μ if for any permutation $P \in P^n$ such that $P\mu = \mu$ we have

$$T[Ph_1, ..., Ph_k] = T[h_1, ..., h_k],$$

for any $h_1, ..., h_k \in \mathbb{R}^n$.

(ii) A k-tensor-valued map $\mu \in \mathbb{R}^n \to \mathcal{F}(\mu) \in T^{k,n}$ is *point-symmetric* if for every $\mu \in \mathbb{R}^n$ and every permutation matrix $P \in P^n$ we have

$$\mathcal{F}(P\mu)[Ph_1, \dots, Ph_k] = \mathcal{F}(\mu)[h_1, \dots, h_k],$$

for any $h_1, ..., h_k \in \mathbb{R}^n$.

Note that if the map $\mu \in \mathbb{R}^n \to \mathcal{F}(\mu) \in T^{k,n}$ is point-symmetric then the tensor $\mathcal{F}(\mu)$ is point-symmetric with respect to μ , for every $\mu \in \mathbb{R}^n$.

- **Definition 2.4** (i) A tensor $T \in T^{k,n}$ is called *block-constant with respect to* μ if $T^{i_1...i_k} = T^{j_1...j_k}$ whenever $(i_1, ..., i_k) \sim_{\mu} (j_1, ..., j_k)$.
 - (ii) A k-tensor-valued map $\mu \in \mathbb{R}^n \to \mathcal{F}(\mu) \in T^{k,n}$ is block-constant if $\mathcal{F}(\mu)$ is block-constant with respect to μ for every $\mu \in \mathbb{R}^n$.

Every block-constant with respect to μ tensor is point-symmetric with respect to μ . By Lemma 2.1, for any differentiable symmetric function $f : \mathbb{R}^n \to \mathbb{R}$ the mapping $\mu \in \mathbb{R}^n \to \nabla f(\mu) \in \mathbb{R}^n$ is a point-symmetric, block-constant, 1-tensor-valued mapping. In general, for every $s \in \mathbb{N}_k$ the mapping (when exists) $\mu \in \mathbb{R}^n \to \nabla^s f(\mu)$ is a point-symmetric, s-tensor-valued map, and if continuous, then the tensor $\nabla^s f(\mu)$ is also symmetric.

By T[h] we denote the (k-1)-tensor on \mathbb{R}^n given by $T[\cdot, ..., \cdot, h]$.

Lemma 2.5 If a k-tensor-valued map $\mu \in \mathbb{R}^n \to T(\mu) \in T^{k,n}$ is point-symmetric and differentiable, then its derivative $\mu \in \mathbb{R}^n \to \nabla T(\mu) \in T^{k+1,n}$ is a point-symmetric map.

Proof. We use the first-order Taylor expansion formula. Let $\{v_m\}$ be a sequence of vectors in \mathbb{R}^n approaching zero such that $v_m/||v_m||$ approaches h as $m \to \infty$.

$$T(\mu + v_m)[h_1, ..., h_k] = T(\mu)[h_1, ..., h_k] + \nabla T(\mu)[h_1, ..., h_k, v_m] + o(||v_m||).$$

On the other hand, for any permutation P we have

$$\begin{aligned} T(\mu + v_m)[h_1, ..., h_k] &= T(P\mu + Pv_m)[Ph_1, ..., Ph_k] \\ &= T(P\mu)[Ph_1, ..., Ph_k] + \nabla T(P\mu)[Ph_1, ..., Ph_k, Pv_m] + o(\|Pv_m\|) \\ &= T(\mu)[h_1, ..., h_k] + \nabla T(P\mu)[Ph_1, ..., Ph_k, Pv_m] + o(\|v_m\|). \end{aligned}$$

Subtracting the two equalities, dividing by $||v_m||$ and letting m go to infinity, we get

$$\nabla T(P\mu)[Ph_1, ..., Ph_k, Ph] = T(\mu)[h_1, ..., h_k, h].$$

Since the vectors $h_1, ..., h_k$, and h are arbitrary, the result follows.

For a fixed vector $\mu \in \mathbb{R}^n$, which is to be understood from the context, we define a linear operation on matrices: $M \in M^n \to M_{\text{in}} \in M^n$, as follows

(3)
$$M_{\rm in}^{ij} = \begin{cases} M^{ij}, & \text{if } i \sim_{\mu} j \\ 0, & \text{otherwise,} \end{cases}$$

and

(4)
$$M_{\rm out} = M - M_{\rm in}$$

2.1 Generalized Hadamard product

In this section we quote briefly several definitions and results from [21] that are crucial for the development in this work. Recall that the Hadamard product of two matrices $H_1 = [H_1^{ij}]$ and $H = [H_2^{ij}]$ of the same size is the matrix of their element-wise product $H_1 \circ H_2 = [H_1^{ij}H_2^{ij}]$. The standard basis on the space M^n is given by the set $\{H_{pq} \in M^n \mid H_{pq}^{ij} = \delta_{ip}\delta_{jq} \text{ for all } i, j \in \mathbb{N}_n\}$, where δ_{ij} is the Kronecker delta function, equal to one if i = j, and zero otherwise.

For each 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 $H_{p_1q_1}, H_{p_2q_2}, \dots, H_{p_kq_k}$

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

Extend this product to a multi-linear map on k matrix arguments:

(5)
$$(H_1 \circ_{\sigma} H_2 \circ_{\sigma} \cdots \circ_{\sigma} H_k)^{i_1 i_2 \dots i_k} = H_1^{i_1 i_{\sigma^{-1}(1)}} \cdots H_k^{i_k i_{\sigma^{-1}(k)}}.$$

For example, when k = 1 there is just one permutation on \mathbb{N}_1 , namely the identity $\sigma = (1)$, and $\circ_{(1)} H = \operatorname{diag} H$. When k = 2 there are two permutations on \mathbb{N}_2 : the identity (1)(2) and the transposition (12). The two corresponding σ -Hadamard products between two matrices are

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

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

Let T be an arbitrary k-tensor on \mathbb{R}^n and let σ be a permutation on \mathbb{N}_k . Let $\text{Diag}^{\sigma}T$ be the 2k-tensor on \mathbb{R}^n defined by

(6)
$$(\operatorname{Diag}^{\sigma}T)^{i_1\dots i_k}_{j_1\dots j_k} = \begin{cases} T^{i_1\dots i_k}, & \text{if } i_s = j_{\sigma(s)}, \forall s = 1, \dots, k\\ 0, & \text{otherwise.} \end{cases}$$

When k = 1 we have $\text{Diag}^{(1)}x = \text{Diag} x$ for any $x \in \mathbb{R}^n$. Any 2k-tensor T on \mathbb{R}^n can be viewed naturally as a k-tensor on the linear space of 2-tensors in the following way

(7)
$$T[H_1, ..., H_k] := \sum_{p_1, q_1=1}^n \cdots \sum_{p_k, q_k=1}^n T^{p_1...p_k}_{q_1...q_k} H_1^{p_1q_1} \cdots H_k^{p_kq_k}.$$

It can be shown that the right-hand side of (7) is invariant under orthonormal changes of the basis in \mathbb{R}^n . If T is a 2k-tensor on \mathbb{R}^n and $H \in M^n$ then by T[H] we denote the 2(k-1)-tensor on \mathbb{R}^n defined by

(8)
$$(T[H])^{i_1...i_{k-1}}_{j_1...j_{k-1}} := \sum_{p,q=1}^{n,n} T^{i_1...i_{k-1}p}_{j_1...j_{k-1}q} H^{pq}.$$

Define dot product between two tensors in $T^{k,n}$ in the usual way

$$\langle T_1, T_2 \rangle = \sum_{p_1, \dots, p_k=1}^n T_1^{p_1 \dots p_k} T_2^{p_1 \dots p_k},$$

and corresponding norm $||T|| = \sqrt{\langle T, T \rangle}$. We define an action (called *conjugation*) of the orthogonal group O^n on the space of all k-tensors on \mathbb{R}^n . For any k-tensor, T, and $U \in O^n$ this action is denoted by $UTU^T \in T^{k,n}$:

(9)
$$(UTU^T)^{i_1\dots i_k} = \sum_{p_1=1}^n \dots \sum_{p_k=1}^n \left(T^{p_1\dots p_k} U^{i_1 p_1} \dots U^{i_k p_k} \right).$$

It is not difficult to show that this action is norm preserving and associative. That is $||VXV^T|| = ||X||$ and $V(UTU^T)V^T = (VU)T(VU)^T$ for all $U, V \in O^n$, see [21].

The Diag^{σ} operator, the σ -Hadamard product, and conjugation by an orthogonal matrix are connected by the following multi-linear duality relation, see [21].

Theorem 2.6 For any k-tensor $T \in T^{k,n}$, any matrices H_1, \ldots, H_k , any orthogonal matrix V, and any permutation σ in P^k we have

(10)
$$\langle T, \tilde{H}_1 \circ_{\sigma} \cdots \circ_{\sigma} \tilde{H}_k \rangle = (V(\operatorname{Diag}^{\sigma} T)V^T)[H_1, ..., H_k],$$

where $\tilde{H}_i = V^T H_i V$, i = 1, ..., k.

We also need the following two lemmas from [21].

Lemma 2.7 Let T be a k-tensor on \mathbb{R}^n , and H be a matrix in M^n . Let $H_{i_1j_1}, \ldots, H_{i_{k-1}j_{k-1}}$ 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

$$\langle T, H_{i_1j_1} \circ_{\sigma} \cdots \circ_{\sigma} H_{i_{k-1}j_{k-1}} \circ_{\sigma} H \rangle = \left(\prod_{t=1}^{k-1} \delta_{i_t j_{\sigma(t)}}\right) \sum_{t=1}^n T^{i_1 \dots i_{k-1}t} H^{tt}.$$

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

$$\langle T, H_{i_1j_1} \circ_{\sigma} \cdots \circ_{\sigma} H_{i_{k-1}j_{k-1}} \circ_{\sigma} H \rangle = \Big(\prod_{\substack{t=1\\t\neq l}}^{k-1} \delta_{i_t j_{\sigma(t)}} \Big) T^{i_1 \dots i_{k-1}j_{\sigma(k)}} H^{j_{\sigma(k)}i_{\sigma^{-1}(k)}}.$$

Lemma 2.8 Let T be any 2k-tensor on \mathbb{R}^n , $V \in O^n$, and let H be any matrix. Then

$$V(T[V^THV])V^T = (VTV^T)[H].$$

2.2 Operations with tensors

For a fixed vector $\mu \in \mathbb{R}^n$ and any $l \in N_k$ define the linear map

$$T \in T^{k,n} \to T_{\text{out}}^{(l)} \in T^{k+1,n}$$

as follows:

(11)
$$(T_{\text{out}}^{(l)})^{i_1\dots i_k i_{k+1}} = \begin{cases} 0, & \text{if } i_l \sim_\mu i_{k+1} \\ \frac{T^{i_1\dots i_{l-1}i_{k+1}i_{l+1}\dots i_k}}{\mu_{i_{k+1}} - \mu_{i_l}}, & \text{if } i_l \not\sim_\mu i_{k+1}. \end{cases}$$

Notice that if T is a block-constant tensor with respect to μ , then so is $T_{\text{out}}^{(l)}$ for each $l \in \mathbb{N}_k$. If $x \in \mathbb{R}^n \to T(x) \in T^{k,n}$ is a k-tensor-valued map, then $x \in \mathbb{R}^n \to T(x)_{\text{out}}^{(l)} \in T^{k+1,n}$ is a (k+1)-tensor-valued map, defined, for each x, by (11) with $\mu := x$. The easy-to-check claim that these maps are linear means that for any two tensors $T_1, T_2 \in T^{k,n}$ and $\alpha, \beta \in \mathbb{R}$ we have

(12)
$$(\alpha T_1 + \beta T_2)_{\text{out}}^{(l)} = \alpha (T_1)_{\text{out}}^{(l)} + \beta (T_2)_{\text{out}}^{(l)}, \text{ for all } l = 1, ..., k.$$

One can iterate this definition: on the space $T^{k+1,n}$ define k+1 linear maps into $T^{k+2,n}$, and so on. A good enumerating tool to keep track of that chain process are the permutations on \mathbb{N}_k , \mathbb{N}_{k+1} , and so on. We make that more clear in the following paragraph.

Given a permutation σ on \mathbb{N}_k we can naturally view it as a permutation on \mathbb{N}_{k+1} fixing the last element. Let τ_l be the transposition (l, k+1), for all l = 1, ..., k, k+1. Define k+1 permutations, $\sigma_{(l)}$, on \mathbb{N}_{k+1} , as follows:

(13)
$$\sigma_{(l)} = \sigma \tau_l, \text{ for } l = 1, ..., k, k+1.$$

Informally speaking, given the cycle decomposition of σ , we obtain $\sigma_{(l)}$, for each l = 1, ..., k, by inserting the element k + 1 immediately after the element l, and when l = k + 1, the permutation $\sigma_{(k+1)}$ fixes the element k + 1. Notice that $\sigma_{(l)}^{-1}(k+1) = l$ for all l, and that the map

(14)
$$(\sigma, l) \in P^k \times \mathbb{N}_{k+1} \to \sigma_{(l)} \in P^{k+1},$$

is one-to-one and onto.

We are now ready to formulate the next theorem. It is the first Calculus-type rule that we need for differentiating spectral functions. It is proved in Appendix B.

Theorem 2.9 Let $\{M_m\}_{m=1}^{\infty}$ be a sequence of symmetric matrices converging to 0, such that the normalized sequence $M_m/||M_m||$ converges to M. Let μ be in $\mathbb{R}^n_{\downarrow}$ and $U_m \to U \in O^n$ be a sequence of orthogonal matrices such that

$$\operatorname{Diag} \mu + M_m = U_m \left(\operatorname{Diag} \lambda (\operatorname{Diag} \mu + M_m)\right) U_m^T, \text{ for all } m = 1, 2, \dots$$

Then for any block-constant k-tensor T on \mathbb{R}^n , and any permutation σ on \mathbb{N}_k we have

(15)
$$\lim_{m \to \infty} \frac{U_m(\operatorname{Diag}^{\sigma} T)U_m^T - \operatorname{Diag}^{\sigma} T}{\|M_m\|} = \sum_{l=1}^k \left(\operatorname{Diag}^{\sigma_{(l)}} T_{\operatorname{out}}^{(l)}\right)[M].$$

Next, for a fixed vector $\mu \in \mathbb{R}^n$ and any $l \in N_k$ define the linear map

$$T \in T^{k,n} \to T_{\mathrm{in}}^{(l)} \in T^{k+1,n},$$

as follows:

(16)
$$(T_{in}^{(l)})^{i_1\dots i_k i_{k+1}} = \begin{cases} T^{i_1\dots i_{l-1}i_{k+1}i_{l+1}\dots i_k}, & \text{if } i_l \sim_\mu i_{k+1} \\ 0, & \text{if } i_l \not\sim_\mu i_{k+1}. \end{cases}$$

Notice that if T is a block-constant tensor with respect to μ , then so is $T_{\text{in}}^{(l)}$ for each l = 1, ..., k. If $x \in \mathbb{R}^n \to T(x) \in T^{k,n}$ is a k-tensor-valued map, then $x \in \mathbb{R}^n \to T(x)_{\text{in}}^{(l)} \in T^{k+1,n}$ is a (k+1)-tensor-valued map defined, for each x, by (16) with $\mu := x$. It is easy to check that these maps are linear, that is, for any two tensors $T_1, T_2 \in T^{k,n}$ and $\alpha, \beta \in \mathbb{R}$ we have

$$(\alpha T_1 + \beta T_2)_{\text{in}}^{(l)} = \alpha (T_1)_{\text{in}}^{(l)} + \beta (T_2)_{\text{in}}^{(l)}, \text{ for all } l = 1, ..., k.$$

Finally, for any $T \in T^{k,n}$ and any $l \in \mathbb{N}_k$ define $T^{\tau_l} \in T^{k+1,n}$ as follows:

(17)
$$(T^{\tau_l})^{i_1 \dots i_k i_{k+1}} = \begin{cases} T^{i_1 \dots i_{l-1} i_l i_{l+1} \dots i_k}, & \text{if } i_l = i_{k+1} \\ 0, & \text{if } i_l \neq i_{k+1} \end{cases}$$

In other words, T^{τ_l} is a (k+1)-tensor with entries off the "hyper plane" $i_l = i_{k+1}$ equal to zero. On the "hyper plane" $i_l = i_{k+1}$ we place the original tensor T.

Notice that when vector μ has distinct coordinates then $i_l \sim_{\mu} i_{k+1}$ if and only if $i_l = i_{k+1}$ and therefore $T_{\text{in}}^{(l)} = T^{\tau_l}$ for every $l \in \mathbb{N}_k$.

The next theorem is the second and last Calculus-type rule that we need. It is proved in Appendix B.

Theorem 2.10 Fix a vector $\mu \in \mathbb{R}^n$. Let $U \in O^n$ be a block-diagonal (with respect to μ) orthogonal matrix and let σ be a permutation on \mathbb{N}_k . Let M be an arbitrary symmetric matrix, and let $h \in \mathbb{R}^n$ be a vector, such that $U^T M_{in} U = \text{Diag } h$. Then

(i) for any block-constant (k+1)-tensor T on \mathbb{R}^n

$$U(\operatorname{Diag}^{\sigma}(T[h]))U^{T} = (\operatorname{Diag}^{\sigma_{(k+1)}}T)[M];$$

(ii) for any block-constant k-tensor T on \mathbb{R}^n

$$U\left(\operatorname{Diag}^{\sigma}(T^{\tau_l}[h])\right)U^T = \left(\operatorname{Diag}^{\sigma_{(l)}}T^{(l)}_{\mathrm{in}}\right)[M], \text{ for all } l = 1, \dots, k,$$

where the permutations $\sigma_{(l)}$, for $l \in \mathbb{N}_k$, are defined by (13).

3 Several standing assumptions

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a k-times differentiable symmetric function. For any integer $s \in [1, k)$, in order to obtain the (s + 1)-th derivative $\nabla^{s+1}(f \circ \lambda)(X)$ of the composition $f \circ \lambda$, we differentiate $\nabla^s(f \circ \lambda)(X)$ and use the tensorial language presented in Section 2 to simplify the calculation. More precisely, for each $\sigma \in P^s$ we define a s-tensor-valued map $\mathcal{A}_{\sigma} : \mathbb{R}^n \to T^{s,n}$, depending only on the function f and its partial derivatives, such that

(18)
$$\nabla^{s}(f \circ \lambda)(X) = V\Big(\sum_{\sigma \in P^{s}} \operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(\lambda(X))\Big) V^{T},$$

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

By [21, Section 5] it is enough to prove (18) only in the case when X is ordered diagonal matrix. That is, $X = \text{Diag } \mu$ for some vector $\mu \in \mathbb{R}^n_{\perp}$.

That (18) holds when s = 1 was shown in [12], see also Subsection 5.2 below.

Let $\{M_m\}_{m=1}^{\infty}$ be any sequence of symmetric matrices converging to 0. In order to show that

$$\lim_{m \to \infty} \frac{\nabla^s (f \circ \lambda) (X + M_m) - \nabla^s (f \circ \lambda) (X) - \nabla^{s+1} (f \circ \lambda) (X) [M_m]}{\|M_m\|} = 0, \quad \text{for } s = 1, \dots, k-1$$

we may assume without loss of generality that $M_m/||M_m||$ converges to a symmetric matrix M. Thus, we assume throughout that $\{M_m\}_{m=1}^{\infty}$ is any sequence of symmetric matrices converging to 0 with $M_m/||M_m||$ converging to matrix $M \in S^n$ and show inductively that

(19)
$$\lim_{m \to \infty} \frac{\nabla^s (f \circ \lambda) (X + M_m) - \nabla^s (f \circ \lambda) (X)}{\|M_m\|} = \nabla^{s+1} (f \circ \lambda) (X) [M], \quad \text{for } s = 1, \dots, k-1.$$

Finally, throughout the rest by $\{U_m\}_{m=1}^{\infty}$ we denote a sequence of orthogonal matrices in O^n , converging to $U \in O^n$ and such that

(20)
$$\operatorname{Diag} \mu + M_m = U_m \left(\operatorname{Diag} \lambda (\operatorname{Diag} \mu + M_m)\right) U_m^T, \text{ for all } m = 1, 2, \dots$$

The next lemma is a simple combination of [13, Lemma 5.10] and [6, Theorem 3.12].

Lemma 3.1 For any $\mu \in \mathbb{R}^n_+$ and any sequence of symmetric matrices $M_m \to 0$ we have that

(21)
$$\lambda(\operatorname{Diag} \mu + M_m)^T = \mu^T + \left(\lambda(X_1^T M_m X_1)^T, ..., \lambda(X_r^T M_m X_r)^T\right)^T + o(\|M_m\|),$$

where $X_l := [e^i | i \in I_l]$, for all l = 1, ..., r.

We denote

(22)
$$h_m := \left(\lambda (X_1^T M_m X_1)^T, ..., \lambda (X_r^T M_m X_r)^T\right)^T.$$

Since $M_m/||M_m||$ converges to M as m goes to infinity and the eigenvalues are continuous functions, we define

(23)
$$h := \lim_{m \to \infty} \frac{h_m}{\|M_m\|} = \left(\lambda (X_1^T M X_1)^T, ..., \lambda (X_r^T M X_r)^T\right)^T.$$

We reserve the symbols h_m and h to denote the above two vectors throughout the paper. With this notation Lemma 3.1 says that

(24)
$$\lambda(\text{Diag}\,\mu + M_m)^T = \mu^T + h_m + o(\|M_m\|).$$

Taking the limit in (20) as m goes to infinity we see, by Theorem 8.1, that U is block-diagonal with respect to μ and

(25)
$$U^T M_{\rm in} U = {\rm Diag} h,$$

where $M_{\rm in}$ is defined by (3).

4 Analyticity of isolated eigenvalues

Let A be in S^n and suppose that the *j*-th largest eigenvalue is isolated, that is

$$\lambda_{j-1}(A) > \lambda_j(A) > \lambda_{j+1}(A).$$

The goal of this section is to give two justifications of the known fact that $\lambda_j(\cdot)$ is an analytic function in a neighbourhood of A. We call a function of several real variables *analytic* at a point if in a neighbourhood of this point it has an power series expansion. The corresponding complex variable notion is called *holomorphic*.

The first justification below is from [23, Theorem 2.1].

Theorem 4.1 Suppose $A \in S^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ is analytic at $\lambda(A)$. Suppose f(Px) = f(x) for every permutation matrix P for which $P\lambda(A) = \lambda(A)$. Then $f \circ \lambda$ is analytic at A.

To see how this theorem implies the analyticity of $\lambda_j(\cdot)$ take

(26)
$$f(x_1, ..., x_n) = \text{the } j^{\text{th}} \text{ largest element of } \{x_1, ..., x_n\}.$$

The function f is a piece-wise affine function. Moreover, for any $x \in \mathbb{R}^n$ in a neighbourhood of the vector $\lambda(A)$ it is given by

$$f(x) = x_j$$

Thus, f is analytic in that neighbourhood. Next, f is a symmetric function and thus by definition f(Px) = f(x) for every $x \in \mathbb{R}^n$ and every permutation matrix P. Therefore by the theorem $\lambda_i = f \circ \lambda$ is an analytic function.

For the second justification we use the following result from [1]. (In the theorem below, $\lambda_i(X)$ denotes an arbitrary eigenvalue of a matrix X, not necessarily the *i*-th largest one.)

Theorem 4.2 (Arnold 1971) Suppose that $A \in \mathbb{C}^{n \times n}$ has q eigenvalues $\lambda_1(A), ..., \lambda_q(A)$ (counting multiplicities) in an open set $\Omega \subset \mathbb{C}$, and the remaining n - q eigenvalues are not in the closure of Ω . Then there is a neighbourhood Δ of A and holomorphic mappings $S : \Delta \to \mathbb{C}^{q \times q}$ and $T : \Delta \to \mathbb{C}^{(n-q) \times (n-q)}$ such that for all $X \in \Delta$

X is similar to
$$\begin{pmatrix} S(X) & 0\\ 0 & T(X) \end{pmatrix}$$
.

and S(A) has eigenvalues $\lambda_1(A), ..., \lambda_q(A)$.

To deduce the result we need, since the j^{th} largest eigenvalue is isolated, we can find an open set $\Omega \subset \mathbb{C}$, such that only that eigenvalue is in Ω and the remaining n-1 are not in the closure of Ω . By the theorem, there is a neighbourhood Δ of A and holomorphic mapping $S : \Delta \to \mathbb{C}$ such that S(X) is equal to the j^{th} largest eigenvalue of X for all X in Δ .

If A is a real symmetric matrix, then the intersection of Δ with S^n is a neighbourhood of A in S^n . Let $\tilde{S}(X)$ denote the restriction of S(X) to $\Delta \cap S^n$. Clearly, $\tilde{S}(X)$ is a holomorphic, real-valued function. Therefore, the coefficients in the power series expansion of $\tilde{S}(X)$ must be real numbers. Thus, the j^{th} largest eigenvalue is a real analytic function in the neighbourhood $\Delta \cap S^n$ or A.

All these considerations make the following observation clear.

Theorem 4.3 Suppose that $A \in S^n$ has distinct eigenvalues and $f : \mathbb{R}^n \to \mathbb{R}$ is k-times (continuously) differentiable in a neighbourhood of $\lambda(A)$. Then $f \circ \lambda$ is k-times (continuously) differentiable in a neighbourhood of A.

5 The k^{th} derivative of functions of eigenvalues at a matrix with distinct eigenvalues

Let $f : \mathbb{R}^n \to \mathbb{R}$ be an arbitrary k-times (continuously) differentiable function. In this section, we do not assume that f is a symmetric function. Our goal is to derive a formula for the k^{th} derivative of $f \circ \lambda$ on the set of symmetric matrices with distinct eigenvalues. The set $\{x \in \mathbb{R}^n | x_i \neq x_j \text{ for every } i \neq j\}$ is dense open set in \mathbb{R}^n . Similarly the set of symmetric matrices with distinct eigenvalues is a dense open set in S^n . (For a simple, convex analysis proof of the last fact, see [19, Corollary 1.6].)

One can obtain the k-th derivative of $f \circ \lambda$ at a matrix with distinct eigenvalues by applying the Chain Rule to the composition $F = f \circ \lambda$. For example, the following formulae are the first three derivatives of F, (see [2, Section X.4]) for any symmetric matrices H_1, H_2, H_3 :

$$\nabla F(X)[H_1] = \nabla f(\lambda(x))[\nabla\lambda(x)[H_1]],$$

$$\nabla^2 F(x)[H_1, H_2] = \nabla^2 f(\lambda(x))[\nabla\lambda(x)[H_1], \nabla\lambda(x)[H_2]] + \nabla f(\lambda(x))[\nabla^2\lambda(x)[H_1, H_2]]$$

$$\nabla^3 F(x)[H_1, H_2, H_3] = \nabla^3 f(\lambda(x))[\nabla\lambda(x)[H_1], \nabla\lambda(x)[H_2], \nabla\lambda(x)[H_3]]$$

$$+ \nabla^2 f(\lambda(x))[\nabla\lambda(x)[H_1], \nabla^2\lambda(x)[H_2, H_3]]$$

$$+ \nabla^2 f(\lambda(x)) [\nabla \lambda(x) [H_2], \nabla^2 \lambda(x) [H_1, H_3]] + \nabla^2 f(\lambda(x)) [\nabla \lambda(x) [H_3], \nabla^2 \lambda(x) [H_1, H_2]] + \nabla f(\lambda(x)) [\nabla^3 \lambda(x) [H_1, H_2, H_3]].$$

This approach to the k-derivative requires every derivative of λ up to the k^{th} . Even if one knows all these derivatives it is not clear how the resulting expression can be simplified. Our goal in this section is to derive a formula for the k^{th} derivative of $f \circ \lambda$ that doesn't require explicit knowledge of the derivatives of λ . Of course the latter can be obtained as a particular case since if f is defined by (26) then $\lambda_j = f \circ \lambda$.

Fix a vector $\mu \in \mathbb{R}^n_{\downarrow}$ with distinct coordinates. Since μ has distinct entries, every block in the partition that it defines has exactly one element. This means that for any $j, i \in \mathbb{N}_n$, $i \sim j \Leftrightarrow i = j$, and that makes any tensor block-constant. In particular, for the matrices X_l , defined in Lemma 3.1, we have $X_l = [e^l]$, l = 1, ..., n. This implies that $h_m = \text{diag } M_m$ and that h = diag M. Notice how the definition of $T_{\text{out}}^{(l)}$, given in (11), changes:

(27)
$$(T_{\text{out}}^{(l)})^{i_1\dots i_k i_{k+1}} = \begin{cases} 0, & \text{if } i_l = i_{k+1} \\ \frac{T^{i_1\dots i_{l-1}i_{k+1}i_{l+1}\dots i_k} - T^{i_1\dots i_{l-1}i_li_{l+1}\dots i_k}}{\mu_{i_{k+1}} - \mu_{i_l}}, & \text{if } i_l \neq i_{k+1} \end{cases}$$

We derive (18) by induction on the order of the derivative.

5.1 Description of the k^{th} derivative

Let $f: \mathbb{R}^n \to \mathbb{R}$ be k-times (continuously) differentiable function defined on the set

$$\Omega := \{ x \in \mathbb{R}^n \, | \, x_i \neq x_j \text{ for every } i \neq j \}.$$

For every natural $s \in \mathbb{N}_k$ and every permutation $\sigma \in P^s$ we define an s-tensor-valued map $\tilde{\mathcal{A}}_{\sigma}$: $\Omega \subset \mathbb{R}^n \to T^{s,n}$ inductively, as follows. For s = 1 and $\sigma = (1)$ we define

$$\hat{\mathcal{A}}_{(1)}(x) := \nabla f(x)$$

Assuming that the maps $\tilde{\mathcal{A}}_{\sigma}(x)$ have been defined for each $\sigma \in P^s$ where the integer s is in [1, k) we define

(28)
$$\tilde{\mathcal{A}}_{\sigma_{(l)}}(x) := (\tilde{\mathcal{A}}_{\sigma}(x))_{\text{out}}^{(l)}, \text{ for all } l \in \mathbb{N}_s, \text{ and}$$
$$\tilde{\mathcal{A}}_{\sigma_{(s+1)}}(x) := \nabla \tilde{\mathcal{A}}_{\sigma}(x).$$

We are now ready to formulate the first main result of this work

Theorem 5.1 Let X be a symmetric matrix with distinct eigenvalues. Let f be a function defined on a neighbourhood of the vector $\lambda(X)$. Then the spectral function $F = f \circ \lambda$ is k-times (continuously) differentiable at X if and only if f is k-times (continuously) differentiable at $\lambda(X)$. Moreover, the formula for the k-th derivative of F at X is given by

(29)
$$\nabla^k F(X) = V\Big(\sum_{\sigma \in P^k} \operatorname{Diag}{}^{\sigma} \tilde{\mathcal{A}}_{\sigma}(\lambda(X))\Big) V^T,$$

where V is any orthogonal matrix such that $X = V(\text{Diag }\lambda(X))V^T$.

The proof proceeds by induction and is presented in the next two subsections.

5.2 Proof of Theorem 5.1: the gradient

Using (24) we compute

$$\lim_{m \to \infty} \frac{(f \circ \lambda)(\operatorname{Diag} \mu + M_m) - (f \circ \lambda)(\operatorname{Diag} \mu)}{\|M_m\|} = \lim_{m \to \infty} \frac{f(\mu + h_m + o(\|M_m\|)) - f(\mu)}{\|M_m\|}$$
$$= \lim_{m \to \infty} \frac{f(\mu) + \nabla f(\mu)[h_m] + o(\|M_m\|) - f(\mu)}{\|M_m\|}$$
$$= \nabla f(\mu)[h]$$
$$= \langle \nabla f(\mu), \operatorname{diag} M \rangle$$
$$= (\operatorname{Diag} \nabla f(\mu))[M].$$

This shows that $\nabla(f \circ \lambda)(\operatorname{Diag} \mu) = \operatorname{Diag}^{(1)} \nabla f(\mu)$. One can see now that

(30)
$$\nabla(f \circ \lambda)(X) = V\left(\operatorname{Diag}^{(1)} \nabla f(\lambda(X))\right) V^{T} = V\left(\sum_{\sigma \in P^{1}} \operatorname{Diag}^{\sigma} \tilde{\mathcal{A}}_{\sigma}(\lambda(X))\right) V^{T},$$

where $X = V(\text{Diag }\lambda(X))V^T$ and $\tilde{\mathcal{A}}_{(1)}(x) = \nabla f(x)$. Trivially, if f is k-times (continuously) differentiable, then $\tilde{\mathcal{A}}_{(1)}(x) = \nabla f(x)$ is (k-1)-times (continuously) differentiable.

If the eigenvalues of X are not distinct and f is a symmetric function, the calculation of the gradient of $f \circ \lambda$ is almost identical and leads to the same final formula. Indeed, using (25), we get

$$\nabla f(\mu)[h] = \langle \nabla f(\mu), \operatorname{diag} (U^T M_{\operatorname{in}} U) \rangle = \left(U(\operatorname{Diag} \nabla f(\mu)) U^T \right) [M] = \left(\operatorname{Diag} \nabla f(\mu) \right) [M].$$

In the last equality we used that U is block-diagonal, orthogonal and the fact that f is symmetric implies that vector $\nabla f(\mu)$ is block-constant, see Lemma 2.1 (i).

5.3 Proof of Theorem 5.1: the induction step

Suppose now that for some $1 \le s < k$

$$\nabla^{s}(f \circ \lambda)(X) = V\big(\sum_{\sigma \in P^{s}} \operatorname{Diag}{}^{\sigma} \tilde{\mathcal{A}}_{\sigma}(\lambda(X))\big) V^{T},$$

where $X = V(\text{Diag }\lambda(X))V^T$. Suppose also that for every $\sigma \in P^s$, the s-tensor-valued map $\tilde{\mathcal{A}}_{\sigma}$: $\mathbb{R}^n \to T^{s,n}$, is (k-s)-times (continuously) differentiable.

Using (24), we differentiate $\nabla^s(f \circ \lambda)$ at the matrix Diag μ :

 $\nabla^{s+1}(f \circ \lambda)(\operatorname{Diag} \mu)[M]$

$$= \lim_{m \to \infty} \frac{\nabla^{s}(f \circ \lambda)(\operatorname{Diag} \mu + M_{m}) - \nabla^{s}(f \circ \lambda)(\operatorname{Diag} \mu)}{\|M_{m}\|}$$

$$= \lim_{m \to \infty} \frac{U_{m}\left(\sum_{\sigma \in P^{s}} \operatorname{Diag}^{\sigma} \tilde{\mathcal{A}}_{\sigma}(\lambda(\operatorname{Diag} \mu + M_{m}))\right)U_{m}^{T} - \sum_{\sigma \in P^{s}} \operatorname{Diag}^{\sigma} \tilde{\mathcal{A}}_{\sigma}(\mu)}{\|M_{m}\|}$$

$$= \lim_{m \to \infty} \frac{\sum_{\sigma \in P^{s}} \left(U_{m}\left(\operatorname{Diag}^{\sigma} \tilde{\mathcal{A}}_{\sigma}(\lambda(\operatorname{Diag} \mu + M_{m}))\right)U_{m}^{T} - \operatorname{Diag}^{\sigma} \tilde{\mathcal{A}}_{\sigma}(\mu)\right)}{\|M_{m}\|}$$

$$= \lim_{m \to \infty} \frac{\sum_{\sigma \in P^{s}} \left(U_{m}\left(\operatorname{Diag}^{\sigma} \tilde{\mathcal{A}}_{\sigma}(\mu + h_{m} + o(\|M_{m}\|))\right)U_{m}^{T} - \operatorname{Diag}^{\sigma} \tilde{\mathcal{A}}_{\sigma}(\mu)\right)}{\|M_{m}\|}$$

$$= \lim_{m \to \infty} \frac{\sum_{\sigma \in P^{s}} \left(U_{m}\left(\operatorname{Diag}^{\sigma} (\tilde{\mathcal{A}}_{\sigma}(\mu) + \nabla \tilde{\mathcal{A}}_{\sigma}(\mu)[h_{m}] + o(\|M_{m}\|))\right)U_{m}^{T} - \operatorname{Diag}^{\sigma} \tilde{\mathcal{A}}_{\sigma}(\mu)\right)}{\|M_{m}\|}$$

$$= \lim_{m \to \infty} \sum_{\sigma \in P^{s}} \frac{U_{m}\left(\operatorname{Diag}^{\sigma} \tilde{\mathcal{A}}_{\sigma}(\mu)\right)U_{m}^{T} - \operatorname{Diag}^{\sigma} \tilde{\mathcal{A}}_{\sigma}(\mu)}{\|M_{m}\|} + \sum_{\sigma \in P^{s}} U\left(\operatorname{Diag}^{\sigma}\left(\nabla \tilde{\mathcal{A}}_{\sigma}(\mu)[h_{l}]\right)\right)U^{T}.$$

By Theorem 2.9, since for every $\sigma \in P^s$ the tensor $\tilde{\mathcal{A}}_{\sigma}(\mu)$ is block-constant, we have

$$\lim_{m \to \infty} \frac{U_m \left(\operatorname{Diag}^{\sigma} \tilde{\mathcal{A}}_{\sigma}(\mu) \right) U_m^T - \operatorname{Diag}^{\sigma} \tilde{\mathcal{A}}_{\sigma}(\mu)}{\|M_m\|} = \sum_{l=1}^s \left(\operatorname{Diag}^{\sigma_{(l)}} \left(\tilde{\mathcal{A}}_{\sigma}(\mu) \right)_{\operatorname{out}}^{(l)} \right) [M]$$
$$= \sum_{l=1}^s \left(\operatorname{Diag}^{\sigma_{(l)}} \tilde{\mathcal{A}}_{\sigma_{(l)}}(\mu) \right) [M].$$

By Theorem 2.10, since for every $\sigma \in P^s \nabla \tilde{\mathcal{A}}_{\sigma}(\mu)$ is a block-constant (s+1)-tensor, we have

$$U\left(\operatorname{Diag}^{\sigma}\left(\nabla\tilde{\mathcal{A}}_{\sigma}(\mu)[h]\right)\right)U^{T} = \left(\operatorname{Diag}^{\sigma_{(s+1)}}\nabla\tilde{\mathcal{A}}_{\sigma}(\mu)\right)[M] = \left(\operatorname{Diag}^{\sigma_{(s+1)}}\tilde{\mathcal{A}}_{\sigma_{(s+1)}}(\mu)\right)[M].$$

Putting everything together we conclude that for every symmetric matrix M:

$$\nabla^{s+1}(f \circ \lambda)(\operatorname{Diag} \mu)[M] = \Big(\sum_{\substack{\sigma \in P^s \\ l \in \mathbb{N}_{s+1}}} \operatorname{Diag}^{\sigma_{(l)}} \tilde{\mathcal{A}}_{\sigma_{(l)}}(\mu)\Big)[M].$$

Notice the parameters of the summation in the above formula and recall that (14) is a one-to-one and onto map. Thus, the comments in Section 3 show that

$$\nabla^{s+1}(f \circ \lambda)(X) = V\Big(\sum_{\sigma \in P^{s+1}} \operatorname{Diag}^{\sigma} \tilde{\mathcal{A}}_{\sigma}(\lambda(X))\Big) V^{T},$$

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

Finally, we show that the (s + 1)-tensor-valued maps $\tilde{\mathcal{A}}_{\sigma_{(l)}}(\cdot)$ are at least (k - s - 1)-times (continuously) differentiable. This is clear when l = s + 1 and $\sigma \in P^s$, since $\tilde{\mathcal{A}}_{\sigma}(\cdot)$ is (k - s)-times

(continuously) differentiable for every $\sigma \in P^s$. For the rest of the maps this is also easy to see. Every entry in $\tilde{\mathcal{A}}_{\sigma_{(l)}}$ is the difference of two entries of $\tilde{\mathcal{A}}_{\sigma}$ divided by a quantity that never becomes zero over the set Ω . This shows that over the set Ω , $\tilde{\mathcal{A}}_{\sigma_{(l)}}(\cdot)$ is (k-s)-times (continuously) differentiable for every $\sigma \in P^s$ and every $l \in \mathbb{N}_s$.

This is the end of the proof of Theorem 5.1.

6 The k^{th} derivative of separable spectral functions

In this section we show that (18) holds at an arbitrary symmetric matrix X (not necessarily with distinct eigenvalues) for the class of separable spectral functions that we now describe.

Let g be a real-valued function on the real interval I, and let X be a symmetric matrix with eigenvalues in I. Define the separable symmetric function

(31)
$$f(x_1, ..., x_n) = g(x_1) + \dots + g(x_n)$$

and the corresponding separable spectral function

(32)
$$F(X) = (f \circ \lambda)(X).$$

Choose an orthogonal matrix V such that $X = V(\text{Diag }\lambda(X))V^T$. Using (1) it is easy to see that if g is differentiable at the points $\{\lambda_i(X) \mid i \in \mathbb{N}_n\}$ then so is F at X and

(33)
$$\nabla F(X) = V(\operatorname{Diag}\left(g'(\lambda_1(X)), ..., g'(\lambda_n(X))\right))V^T.$$

Separable spectral functions and their derivatives are of great importance for modern optimization, for example [3], [11], [22]. For the role of general spectral functions see the two survey papers [14] and [15].

The original interest in the class of matrix-valued functions (33) was started by Löner with his paper [18], where he established the connection between the differentiability of g' and the monotonicity of the map (33) with respect to the semidefinite order. Later in [10], Löwner's student Kraus, investigated the conditions on g' that make the map (33) convex with respect to the semidifinite order. For more information, related and recent results one should refer to [2, Chapter V]. The matrix-valued maps (33) also arise as a particular case of the so called *primary matrix functions* investigated extensively in [7, Chapter 6]. The first two derivatives of (33) can be found in [2, Chapter V].

6.1 Description of the k^{th} derivative

Let $g: I \to \mathbb{R}$ be k-times differentiable. We begin by defining the function $g^{[(1)]}(x): I \to \mathbb{R}$ as

$$g^{[(1)]}(x) := g'(x).$$

Next, define the symmetric function $g^{[(12)]}(x,y): I \times I \to \mathbb{R}$ as

(34)
$$g^{[(12)]}(x,y) := \begin{cases} g''(x), & \text{if } x = y \\ \frac{g^{[(1)]}(x) - g^{[(1)]}(y)}{x - y}, & \text{if } x \neq y. \end{cases}$$

The integral representation $g^{[(12)]}(x,y) = \int_0^1 g''(y+t(x-y)) dt$ shows that $g^{[(12)]}(x,y)$ is as smooth, in both arguments, as g''.

Denote by \tilde{P}^s the set of all permutations from P^s that have one cycle in their cycle decomposition. Clearly $|\tilde{P}^s| = (s-1)!$. Notice that for every $\sigma \in \tilde{P}^s$ and every $l \in \mathbb{N}_s$ we have $\sigma_{(l)} \in \tilde{P}^{s+1}$. Moreover, as σ varies over \tilde{P}^s and l varies over \mathbb{N}_s , the permutation $\sigma_{(l)}$ varies over \tilde{P}^{s+1} in a one-to-one and onto fashion.

Suppose that for every $\sigma \in \tilde{P}^s$, where $1 \leq s < k$, we have defined the function $g^{[\sigma]}(x_1, ..., x_s)$ on the set $I \times I \times \cdots \times I$, s-times, and suppose that these functions are as smooth as $g^{(s)}$ (the s-th derivative of g). For every $\sigma \in \tilde{P}^s$ and every $l \in \mathbb{N}_s$ we define the function $g^{[\sigma_{(l)}]}(x_1, ..., x_s, x_{s+1})$ as follows:

(35)
$$g^{[\sigma_{(l)}]}(x_1, ..., x_{s+1}) := \begin{cases} \nabla_l g^{[\sigma]}(x_1, ..., x_s), & \text{if } x_l = x_{s+1} \\ \frac{g^{[\sigma]}(x_1, ..., x_l, ..., x_s) - g^{[\sigma]}(x_1, ..., x_{s+1}, ..., x_s)}{x_l - x_{s+1}}, & \text{if } x_l \neq x_{s+1}, \end{cases}$$

where in the second case of the definition, both x_l and x_{s+1} are in *l*-th position, and ∇_l denotes the partial derivative with respect to the *l*-th argument. Using the integral formula

$$g^{[\sigma_{(l)}]}(x_1, ..., x_{s+1}) = \int_0^1 \nabla_l g^{[\sigma]}(x_1, ..., x_{l-1}, x_{s+1} + t(x_l - x_{s+1}), x_{l+1}, ..., x_s) dt,$$

for every $l \in \mathbb{N}_s$, we see that $g^{[\sigma_{(l)}]}(x_1, ..., x_{s+1})$ is as smooth as $g^{(s+1)}$, the (s+1)-th derivative of g. We continue inductively in this way until we define the functions $\{g^{[\sigma]}(x_1, ..., x_k) \mid \sigma \in \tilde{P}^k\}$.

Finally, for every $s \in \mathbb{N}_k$ and every $\sigma \in \tilde{P}^s$, we define a s-tensor-valued map

(36)
$$\begin{aligned} \mathcal{A}_{\sigma} : \mathbb{R}^{n} \to T^{s,n}, \text{ by} \\ \left(\mathcal{A}_{\sigma}(x)\right)^{i_{1}\dots i_{s}} := g^{[\sigma]}(x_{i_{1}},\dots,x_{i_{s}}). \end{aligned}$$

Clearly, if $(i_1, ..., i_s) \sim_x (j_1, ..., j_s)$, then $(\mathcal{A}_{\sigma}(x))^{i_1...i_s} = (\mathcal{A}_{\sigma}(x))^{j_1...j_s}$, which shows that (36) defines a block-constant map, moreover, it is as smooth as $g^{(s)}$ for every $s \in \mathbb{N}_k$.

We are now ready to formulate the second main result of this work.

Theorem 6.1 Let g be a k-times differentiable function defined on an interval I. Let X be a symmetric matrix with eigenvalues in the interval I, and let V be an orthogonal matrix such that

 $X = V(\text{Diag }\lambda(X))V^T$. Then the separable spectral function F defined by (31) and (32) is k-times differentiable at X, and its k-th derivative is

(37)
$$\nabla^k F(X) = V\Big(\sum_{\sigma \in P^k} \operatorname{Diag}{}^{\sigma} \mathcal{A}_{\sigma}(\lambda(X))\Big) V^T,$$

where $\mathcal{A}_{\sigma}(x) \equiv 0$ if $\sigma \notin \tilde{P}^k$.

The proof is given in the next subsection. We proceed by induction — consecutively differentiating F(X). The base case for the induction is clear. Indeed, if k = 1 then (37) reduces to the formula for the gradient (33).

6.2 Proof of Theorem 6.1: the induction step

Suppose that $g: I \to \mathbb{R}$ is k-times differentiable and the formula for the s-th derivative $(1 \le s < k)$ of F at the matrix X is given by

$$\nabla^{s} F(X) = V\Big(\sum_{\sigma \in P^{s}} \operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(\lambda(X))\Big) V^{T} = V\Big(\sum_{\sigma \in \tilde{P}^{s}} \operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(\lambda(X))\Big) V^{T}.$$

For each $\sigma \in P^s$, the s-tensor-valued map $\mathcal{A}_{\sigma} : \mathbb{R} \to T^{s,n}$ is (k-s)-times differentiable. Recall Section 3 for the simplifying assumptions and notation that we use below. We now differentiate:

$$\begin{split} \nabla^{(s+1)} F(\operatorname{Diag} \mu)[M] &= \lim_{m \to \infty} \frac{\nabla^s F(\operatorname{Diag} \mu + M_m) - \nabla^s F(\operatorname{Diag} \mu)}{\|M_m\|} \\ &= \lim_{m \to \infty} \frac{U_m \Big(\sum_{\sigma \in \tilde{P}^s} \operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(\lambda(\operatorname{Diag} \mu + M_m)) \Big) U_m^T - \sum_{\sigma \in \tilde{P}^s} \operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(\mu)}{\|M_m\|} \\ &= \lim_{m \to \infty} \frac{U_m \Big(\sum_{\sigma \in \tilde{P}^s} \operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(\mu + h_m + o(\|M_m\|)) \Big) U_m^T - \sum_{\sigma \in \tilde{P}^s} \operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(\mu)}{\|M_m\|} \\ &= \lim_{m \to \infty} \frac{U_m \Big(\sum_{\sigma \in \tilde{P}^s} \operatorname{Diag}^{\sigma} \Big(\mathcal{A}_{\sigma}(\mu) + \nabla \mathcal{A}_{\sigma}(\mu) [h_m] + o(\|M_m\|) \Big) \Big) U_m^T - \sum_{\sigma \in \tilde{P}^s} \operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(\mu)}{\|M_m\|} \\ &= \lim_{m \to \infty} \frac{U_m \Big(\sum_{\sigma \in \tilde{P}^s} \operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(\mu) \Big) U_m^T - \sum_{\sigma \in \tilde{P}^s} \operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(\mu)}{\|M_m\|} + U \Big(\sum_{\sigma \in \tilde{P}^s} \operatorname{Diag}^{\sigma} \big(\nabla \mathcal{A}_{\sigma}(\mu) [h] \big) \Big) U^T. \end{split}$$

Using Theorem 2.9, we wrap up the first summand in the last expression:

$$(38) \quad \lim_{m \to \infty} \frac{U_m \Big(\sum_{\sigma \in \tilde{P}^s} \operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(\mu) \Big) U_m^T - \sum_{\sigma \in \tilde{P}^s} \operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(\mu)}{\|M_m\|} = \sum_{\substack{\sigma \in \tilde{P}^s \\ l \in \mathbb{N}_s}} \Big(\operatorname{Diag}^{\sigma_{(l)}} (\mathcal{A}_{\sigma}(\mu))_{\operatorname{out}}^{(l)} \Big) [M].$$

Next, we focus our attention on the gradient $\nabla \mathcal{A}_{\sigma}(\mu)$. Using the definition, (36), and the Chain Rule, we get

(39)
$$\nabla \left[(\mathcal{A}_{\sigma}(\mu))^{i_1...i_s} \right] = \sum_{l=1}^s \nabla_l g^{[\sigma]}(\mu_{i_1}, ..., \mu_{i_s}) e^{i_l} = \sum_{l=1}^s g^{[\sigma_{(l)}]}(\mu_{i_1}, ..., \mu_{i_s}, \mu_{i_l}) e^{i_l},$$

where for the second equality we used (35). For convenience, for every $\sigma \in \tilde{P}^s$ and every $l \in \mathbb{N}_s$, we define the map

(40)
$$\begin{aligned} T^{l}_{\sigma} : \mathbb{R}^{n} \to T^{s,n}, \text{ by} \\ (T^{l}_{\sigma}(\mu))^{i_{1}...i_{s}} := g^{[\sigma_{(l)}]}(\mu_{i_{1}},...,\mu_{i_{s}},\mu_{i_{l}}) \end{aligned}$$

Notice that each one of these maps is block-constant.

Lemma 6.2 The gradient of $\mathcal{A}_{\sigma}(\mu)$ can be decomposed as

(41)
$$\nabla \mathcal{A}_{\sigma}(\mu) = \sum_{l=1}^{s} \left(T_{\sigma}^{l}(\mu) \right)^{\tau_{l}},$$

where the "lifting" $\left(T_{\sigma}^{l}(\mu)\right)^{\tau_{l}}$ is defined by (17).

Proof. Fix a multi index $(i_1, ..., i_s)$. By definition of the gradient $\nabla \mathcal{A}_{\sigma}(\mu)$ we have

$$\nabla \left[(\mathcal{A}_{\sigma}(\mu))^{i_1 \dots i_s} \right] = \left((\nabla \mathcal{A}_{\sigma}(\mu))^{i_1 \dots i_s, 1}, (\nabla \mathcal{A}_{\sigma}(\mu))^{i_1 \dots i_s, 2}, \dots, (\nabla \mathcal{A}_{\sigma}(\mu))^{i_1 \dots i_s, n} \right)^T$$

We compute the p-th entry in the last vector. On one hand, using (39), we get:

$$(\nabla \mathcal{A}_{\sigma}(\mu))^{i_1...i_s,p} = \sum_{\substack{l=1\\i_l=p}}^{s} g^{[\sigma_{(l)}]}(\mu_{i_1},...,\mu_{i_s},\mu_{i_l}).$$

On the other, using (17) and (40), we evaluate the right-hand side of (41):

$$\left(\sum_{l=1}^{s} \left(T_{\sigma}^{l}(\mu)\right)^{\tau_{l}}\right)^{i_{1}...i_{s},p} = \sum_{l=1}^{s} \left(\left(T_{\sigma}^{l}(\mu)\right)^{\tau_{l}}\right)^{i_{1}...i_{s},p}$$
$$= \sum_{l=1}^{s} \left(T_{\sigma}^{l}(\mu)\right)^{i_{1}...i_{s}} \delta_{i_{l}p}$$
$$= \sum_{\substack{l=1\\i_{l}=p}}^{s} \left(T_{\sigma}^{l}(\mu)\right)^{i_{1}...i_{s}}$$
$$= \sum_{\substack{l=1\\i_{l}=p}}^{s} g^{[\sigma_{(l)}]}(\mu_{i_{1}},...,\mu_{i_{s}},\mu_{i_{l}})$$

The lemma follows.

(42)

Using (41) we return to the second term left from the differentiation of $\nabla^s F(X)$:

$$U\Big(\sum_{\sigma\in\tilde{P}^{s}}\operatorname{Diag}^{\sigma}\big(\nabla\mathcal{A}_{\sigma}(\mu)[h]\big)\Big)U^{T} = U\Big(\sum_{\sigma\in\tilde{P}^{s}}\operatorname{Diag}^{\sigma}\Big(\Big(\sum_{l=1}^{s}\big(T_{\sigma}^{l}(\mu)\big)^{\tau_{l}}\big)[h]\big)\Big)U^{T}$$
$$= U\Big(\sum_{\sigma\in\tilde{P}^{s}}\sum_{l=1}^{s}\operatorname{Diag}^{\sigma}((T_{\sigma}^{l}(\mu))^{\tau_{l}}[h])\Big)U^{T}$$
$$= \sum_{\substack{\sigma\in\tilde{P}^{s}\\l\in\mathbb{N}_{s}}}U\Big(\operatorname{Diag}^{\sigma}((T_{\sigma}^{l}(\mu))^{\tau_{l}}[h])\Big)U^{T}$$
$$= \sum_{\substack{\sigma\in\tilde{P}^{s}\\l\in\mathbb{N}_{s}}}\Big(\operatorname{Diag}^{\sigma_{(l)}}(T_{\sigma}^{l}(\mu))^{(l)}_{\mathrm{in}}\Big)[M],$$

where in the last equality we used Theorem 2.10. Putting (38) and (42) together we obtain

$$\nabla^{(s+1)} F(\operatorname{Diag} \mu)[M] = \sum_{\substack{\sigma \in \tilde{P}^s \\ l \in \mathbb{N}_s}} \left(\operatorname{Diag}^{\sigma_{(l)}} (\mathcal{A}_{\sigma}(\mu))_{\operatorname{out}}^{(l)} \right)[M] + \sum_{\substack{\sigma \in \tilde{P}^s \\ l \in \mathbb{N}_s}} \left(\operatorname{Diag}^{\sigma_{(l)}} (T_{\sigma}^l(\mu))_{\operatorname{in}}^{(l)} \right)[M].$$

We group the two sums into one and notice that since M is an arbitrary symmetric matrix we can remove it from both sides of the equation:

$$\nabla^{(s+1)} F(\operatorname{Diag} \mu) = \sum_{\substack{\sigma \in \tilde{P}^s \\ l \in \mathbb{N}_s}} \operatorname{Diag}^{\sigma_{(l)}} \left((\mathcal{A}_{\sigma}(\mu))_{\operatorname{out}}^{(l)} + (T_{\sigma}^l(\mu))_{\operatorname{in}}^{(l)} \right).$$

This already shows that $\nabla^s F(\text{Diag }\mu)$ is differentiable. We show now that $\nabla^{(s+1)}F(\text{Diag }\mu)$ has the form (37). This last step is the subject of the next lemma.

Lemma 6.3 For every $\sigma \in \tilde{P}^s$ and every $l \in \mathbb{N}_s$ we have

(43)
$$\mathcal{A}_{\sigma_{(l)}}(\mu) = \left(T^l_{\sigma}(\mu)\right)^{(l)}_{\text{in}} + \left(\mathcal{A}_{\sigma}(\mu)\right)^{(l)}_{\text{out}}$$

Proof. Fix a number $l \in \mathbb{N}_s$ and a multi index $(i_1, ..., i_s, i_{s+1})$. We consider two cases depending on whether $\mu_{i_{s+1}}$ equals μ_{i_l} .

Case I. Suppose $i_l \sim_{\mu} i_{s+1}$. Using (36) and (35) the entry on the left-hand side of (43) corresponding to the multi index $(i_1, ..., i_s, i_{s+1})$ is

$$\left(\mathcal{A}_{\sigma_{(l)}}(\mu)\right)^{i_1\dots i_s i_{s+1}} = g^{[\sigma_{(l)}]}(\mu_{i_1},\dots,\mu_{i_s},\mu_{i_{s+1}}) = \nabla_l g^{[\sigma]}(\mu_{i_1},\dots,\mu_{i_s}).$$

On the other hand, the right-hand side evaluates to

$$\left(\left(T_{\sigma}^{l}(\mu) \right)_{\text{in}}^{(l)} + \left(\mathcal{A}_{\sigma}(\mu) \right)_{\text{out}}^{(l)} \right)^{i_{1}\dots i_{s}i_{s+1}} = \left(\left(T_{\sigma}^{l}(\mu) \right)_{\text{in}}^{(l)} \right)^{i_{1}\dots i_{s}i_{s+1}} + \left(\left(\mathcal{A}_{\sigma}(\mu) \right)_{\text{out}}^{(l)} \right)^{i_{1}\dots i_{s}i_{s+1}}$$

$$= \left(\left(T_{\sigma}^{l}(\mu) \right)_{in}^{(l)} \right)^{i_{1}...i_{s}i_{s+1}} + 0$$

$$= \left(T_{\sigma}^{l}(\mu) \right)^{i_{1}...i_{s}}$$

$$= g^{[\sigma_{(l)}]}(\mu_{i_{1}},...,\mu_{i_{s}},\mu_{i_{l}})$$

$$= \nabla_{l}g^{[\sigma]}(\mu_{i_{1}},...,\mu_{i_{s}}),$$

where in the third equality we used (16) and the fact that $T_l(\mu)$ is block-constant.

Case II. Suppose $i_l \not\sim_{\mu} i_{s+1}$. Using (36) and (35) the entry on the left-hand side corresponding to the multi index $(i_1, ..., i_s, i_{s+1})$ is

$$\left(\mathcal{A}_{\sigma_{(l)}}(\mu) \right)^{i_1 \dots i_s i_{s+1}} = g^{[\sigma_{(l)}]}(\mu_{i_1}, \dots, \mu_{i_s}, \mu_{i_{s+1}})$$

=
$$\frac{g^{[\sigma]}(\mu_{i_1}, \dots, \mu_{i_l}, \dots, \mu_{i_s}) - g^{[\sigma]}(\mu_{i_1}, \dots, \mu_{i_{s+1}}, \dots, \mu_{i_s})}{\mu_{i_l} - \mu_{i_{s+1}}},$$

where both μ_{i_l} and $\mu_{i_{s+1}}$ are in the *l*-th position. On the other hand, the right-hand side evaluates to

$$\begin{split} \left(\left(T_{\sigma}^{l}(\mu) \right)_{\mathrm{in}}^{(l)} + \left(\mathcal{A}_{\sigma}(\mu) \right)_{\mathrm{out}}^{(l)} \right)^{i_{1}\dots i_{s}i_{s+1}} &= \left(\left(T_{\sigma}^{l}(\mu) \right)_{\mathrm{in}}^{(l)} \right)^{i_{1}\dots i_{s}i_{s+1}} + \left(\left(\mathcal{A}_{\sigma}(\mu) \right)_{\mathrm{out}}^{(l)} \right)^{i_{1}\dots i_{s}i_{s+1}} \\ &= 0 + \left(\left(\mathcal{A}_{\sigma}(\mu) \right)_{\mathrm{out}}^{(l)} \right)^{i_{1}\dots i_{s}i_{s+1}} \\ &= \frac{\left(\mathcal{A}_{\sigma}(\mu) \right)^{i_{1}\dots i_{l-1}i_{s+1}i_{l+1}\dots i_{s}} - \left(\mathcal{A}_{\sigma}(\mu) \right)^{i_{1}\dots i_{l-1}i_{l}i_{l+1}\dots i_{s}} \\ &= \frac{\left(\mathcal{A}_{\sigma}(\mu) \right)^{i_{1}\dots i_{l-1}i_{s+1}i_{l+1}\dots i_{s}} - \left(\mathcal{A}_{\sigma}(\mu) \right)^{i_{1}\dots i_{l-1}i_{l}i_{l+1}\dots i_{s}} \\ &= \frac{g^{[\sigma]}(\mu_{i_{1}},\dots,\mu_{i_{s+1}},\dots,\mu_{i_{s}}) - g^{[\sigma]}(\mu_{i_{1}},\dots,\mu_{i_{s}})}{\mu_{i_{s+1}} - \mu_{i_{l}}}. \end{split}$$

In both cases, the two sides are equal.

This concludes the inductive step and the proof of Theorem 6.1.

The two separate developments in Section 5 and Section 6 have to be reconciled in their common case. This is done by the following theorem proved in Appendix C.

Theorem 6.4 Suppose that matrix X has distinct eigenvalues, and the spectral function is separable and k-times differentiable at X. Then the two formulae for the k-th derivative of the spectral function at X, namely, the one given in Theorem 5.1 where the operators $\tilde{\mathcal{A}}_{\sigma}$ are defined by the inductive equations (28), and the one in Theorem 6.1 where the operators \mathcal{A}_{σ} are defined by equations (36), are the same. More precisely we have

$$\sum_{\sigma \in P^s} \operatorname{Diag}{}^{\sigma} \tilde{\mathcal{A}}_{\sigma}(x) = \sum_{\sigma \in P^s} \operatorname{Diag}{}^{\sigma} \mathcal{A}_{\sigma}(x), \quad \text{for every } s = 1, 2, ..., k,$$

where $x = \lambda(X)$.

It is worth presenting a particular case of Theorem 6.1. More specializations of Theorem 6.1, in the case when g is C^k , are given in Subsubsection 6.3.1.

Corollary 6.5 Let g be twice differentiable in I and let X be a symmetric matrix with all eigenvalues in I, such that $X = V(\text{Diag }\lambda(X))V^T$ for some orthogonal matrix V. Then

(44)
$$\nabla^2 F(X) = V \left(\operatorname{Diag}^{(12)} \mathcal{A}_{(12)}(\lambda(X)) \right) V^T,$$

where $\mathcal{A}_{\scriptscriptstyle (12)}(\cdot)$ is defined by

$$\mathcal{A}_{(12)}^{ij}(x) = \begin{cases} g''(x_i), & \text{if } x_i = x_j \\ \frac{g'(x_i) - g'(x_j)}{x_i - x_j}, & \text{if } x_i \neq x_j. \end{cases}$$

Using approximation techniques, it was shown in [2, Theorem V.3.3] that for any two symmetric matrices H_1 and H_2

(45)
$$\nabla^2 F(X)[H_1, H_2] = \langle V \big(\mathcal{A}_{(12)}(\lambda(X)) \circ (V^T H_1 V) \big) V^T, H_2 \rangle,$$

where ' \circ ' stands for the usual Hadamard product. For completeness we now show that (44) is the same as (45). This is the content of the next proposition.

Proposition 6.6 For any $n \times n$ matrix A, any orthogonal V, and any symmetric H_1 and H_2 , we have the equality

$$\left(V(\operatorname{Diag}^{(12)}A)V^{T}\right)[H_{1},H_{2}] = \langle V\left(A \circ (V^{T}H_{1}V)\right)V^{T},H_{2}\rangle,$$

where 'o' stands for the ordinary Hadamard product.

Proof. We develop the two sides of the stated equality and compare the results. By Theorem 2.6, the left-hand side is equal to

$$V\left(\operatorname{Diag}^{(12)}A\right)V^{T}[H_{1},H_{2}] = \langle A, \tilde{H}_{1} \circ_{(12)} \tilde{H}_{2} \rangle$$

On the other hand

$$\langle V \big(A \circ (V^T H_1 V) \big) V^T, H_2 \rangle = \langle A \circ \tilde{H}_1, \tilde{H}_2 \rangle = \langle A, \tilde{H}_1 \circ \tilde{H}_2 \rangle.$$

Finally one can check directly from the definitions that $\tilde{H}_1 \circ_{(12)} \tilde{H}_2 = \tilde{H}_1 \circ \tilde{H}_2^T = \tilde{H}_1 \circ \tilde{H}_2$, using the symmetry of \tilde{H}_2 .

6.3 C^k separable spectral functions

Theorem 6.1 holds for every k-times differentiable functions g. If in addition g is k-times continuously differentiable, then (37) can be significantly simplified. That is what we describe in this section. In particular, we show three properties of the functions $g^{[\sigma]}(x_1, ..., x_s)$, for every $1 \leq s \leq k$ and every $\sigma \in \tilde{P}^s$. First, we express $g^{[\sigma]}(x_1, ..., x_s)$ as a ratio of two determinants whenever the numbers $x_1, ..., x_s$ are distinct. Second, as a consequence of the determinant formula, it will become evident that $g^{[\sigma]}(x_1, ..., x_s)$ is a symmetric function of its s arguments. Finally, third, we show that $g^{[\sigma_1]}(x_1, ..., x_s) = g^{[\sigma_2]}(x_1, ..., x_s)$ for all σ_1 and σ_2 in \tilde{P}^s . Thus, all tensors $\{\mathcal{A}_{\sigma}(x) \mid \sigma \in \tilde{P}^k\}$, in (37) are equal to each other, but are lifted onto different k-dimensional "diagonal planes" in the 2k-dimensional tensor.

Recall the Vandermonde determinant:

$$V(x_1, ..., x_s) := \begin{vmatrix} x_1^{s-1} & x_2^{s-1} & \cdots & x_s^{s-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_1 & x_2 & \cdots & x_s \\ 1 & 1 & \cdots & 1 \end{vmatrix} = \prod_{j < i} (x_j - x_i),$$

and for any $y \in \mathbb{R}^s$ consider its variation:

$$V\begin{pmatrix}y_1, \dots, y_s\\x_1, \dots, x_s\end{pmatrix} := \begin{vmatrix} y_1 & y_2 & \cdots & y_s\\x_1^{s-2} & x_2^{s-2} & \cdots & x_s^{s-2}\\ \vdots & \vdots & \ddots & \vdots\\x_1 & x_2 & \cdots & x_s\\1 & 1 & \cdots & 1 \end{vmatrix},$$

with the agreement that when s = 1: $V(x_1) = 1$ and $V\begin{pmatrix}y_1\\x_1\end{pmatrix} = y_1$.

Lemma 6.7 For any vector $(x_1, ..., x_s, x_{s+1})$ with distinct coordinates, any $y \in \mathbb{R}^{s+1}$, and $l \in \mathbb{N}_s$

$$(46) \quad \frac{V\begin{pmatrix}y_1, \dots, y_s\\x_1, \dots, x_s\end{pmatrix}}{V(x_1, \dots, x_s)} - \frac{V\begin{pmatrix}y_1, \dots, y_{l-1}, y_{s+1}, y_{l+1}, \dots, y_s\\x_1, \dots, x_{l-1}, x_{s+1}, x_{l+1}, \dots, x_s\end{pmatrix}}{V(x_1, \dots, x_{l-1}, x_{s+1}, x_{l+1}, \dots, x_s)} = (x_l - x_{s+1}) \frac{V\begin{pmatrix}y_1, \dots, y_l, y_{s+1}, y_{l+1}, \dots, y_s\\x_1, \dots, x_l, x_{s+1}, x_{l+1}, \dots, x_s\end{pmatrix}}{V(x_1, \dots, x_l, x_{s+1}, x_{l+1}, \dots, x_s)}$$

Proof. When s = 1 the lemma is easy to check directly, indeed:

$$\frac{V\binom{y_1}{x_1}}{V(x_1)} - \frac{V\binom{y_2}{x_2}}{V(x_2)} = (x_1 - x_2) \frac{V\binom{y_1, y_2}{x_1, x_2}}{V(x_1, x_2)}$$

For the rest of the proof we assume $s \ge 2$. Consider both sides of the above identity as a multivariate polynomial (of degree one) in the variables $y_1, ..., y_s, y_{s+1}$. We show that the coefficients in front of y_k on both sides are equal for all $k \in \mathbb{N}_{s+1}$. Notice first that

$$V(x_1, ..., x_{l-1}, x_{s+1}, x_{l+1}, ..., x_s) = (-1)^{s-l} V(x_1, ..., x_{l-1}, x_{l+1}, ..., x_s, x_{s+1}),$$

$$V\begin{pmatrix}y_1, \dots, y_{l-1}, y_{s+1}, y_{l+1}, \dots, y_s\\x_1, \dots, x_{l-1}, x_{s+1}, x_{l+1}, \dots, x_s\end{pmatrix} = (-1)^{s-l}V\begin{pmatrix}y_1, \dots, y_{l-1}, y_{l+1}, \dots, y_s, y_{s+1}\\x_1, \dots, x_{l-1}, x_{l+1}, \dots, x_s, x_{s+1}\end{pmatrix}.$$

We consider four cases according to the position of the index k in the partition $\mathbb{N}_{s+1} = \{1, ..., l-1\} \cup \{l\} \cup \{l+1, ..., s\} \cup \{s+1\}$. (In all product formulae below, it is assumed that the index j < i. This condition is omitted for typographical reasons. Also a circumflex above a multiple in a product denotes that the multiple is missing.) First, let $k \in \{1, ..., l-1\}$. The coefficient of y_k in the left-hand side of (46) is equal to

$$\begin{split} (-1)^{k+1} \frac{\prod_{i,j\in\mathbb{N}_{s+h}\{k,s+1\}} (x_j - x_i)}{\prod_{i,j\in\mathbb{N}_{s+h}\{s+1\}} (x_j - x_i)} &- (-1)^{k+1} \frac{\prod_{i,j\in\mathbb{N}_{s+h}\{k,l\}} (x_j - x_i)}{\prod_{i,j\in\mathbb{N}_{s+h}\{l\}} (x_j - x_i)} \\ &= \frac{(-1)^{k+1}}{(x_1 - x_k) \cdots (x_{k-1} - x_k) (x_k - x_{k+1}) \cdots (x_k - x_s)} \\ &- \frac{(-1)^{k+1}}{(x_1 - x_k) \cdots (x_{k-1} - x_k) (x_k - x_{k+1}) \cdots (x_k - x_{s+1})} \\ &= \frac{(-1)^{k+1}}{(x_1 - x_k) \cdots (x_{k-1} - x_k) (x_k - x_{k+1}) \cdots (x_k - x_s)} \left(\frac{1}{x_k - x_l} - \frac{1}{x_k - x_{s+1}}\right) \\ &= \frac{(-1)^{k+1} (x_l - x_{s+1})}{(x_1 - x_k) \cdots (x_{k-1} - x_k) (x_k - x_{k+1}) \cdots (x_k - x_{s+1})} \\ &= (-1)^{k+1} (x_l - x_{s+1}) \frac{\prod_{i,j\in\mathbb{N}_{s+h}\{k\}} (x_j - x_i)}{\prod_{i,j\in\mathbb{N}_{s+h}\{k\}} (x_j - x_i)}, \end{split}$$

which is the coefficient of y_k in the right-hand side of (46).

Suppose now, k = l. Then the coefficient of y_k in the left-hand side of (46) is equal to

$$(-1)^{l+1} \frac{\prod_{i,j \in \mathbb{N}_{s+1} \setminus \{l,s+1\}} (x_j - x_i)}{\prod_{i,j \in \mathbb{N}_{s+1} \setminus \{s+1\}} (x_j - x_i)} - 0 = \frac{(-1)^{l+1}}{(x_1 - x_l) \cdots (x_{l-1} - x_l)(x_l - x_{l+1}) \cdots (x_l - x_s)}$$
$$= \frac{(-1)^{l+1} (x_l - x_{s+1})}{(x_1 - x_l) \cdots (x_{l-1} - x_l)(x_l - x_{l+1}) \cdots (x_l - x_{s+1})}$$
$$= (-1)^{l+1} (x_l - x_{s+1}) \frac{\prod_{i,j \in \mathbb{N}_{s+1} \setminus \{l\}} (x_j - x_i)}{\prod_{i,j \in \mathbb{N}_{s+1}} (x_j - x_i)},$$

which is the corresponding coefficient in the right-hand side of (46).

When $k \in \{l+1, ..., s\}$, the coefficient of y_k on the left-hand side of (46) is:

$$(-1)^{k+1} \frac{\prod_{i,j \in \mathbb{N}_{s+1} \setminus \{k,s+1\}} (x_j - x_i)}{\prod_{i,j \in \mathbb{N}_{s+1} \setminus \{s+1\}} (x_j - x_i)} - (-1)^k \frac{\prod_{i,j \in \mathbb{N}_{s+1} \setminus \{k,l\}} (x_j - x_i)}{\prod_{i,j \in \mathbb{N}_{s+1} \setminus \{l\}} (x_j - x_i)}$$
$$= \frac{(-1)^{k+1}}{(x_1 - x_k) \cdots (x_{k-1} - x_k) (x_k - x_{k+1}) \cdots (x_k - x_s)}$$

$$=\frac{(-1)^{k}}{(x_{1}-x_{k})\cdots(x_{l}-x_{k})\cdots(x_{k-1}-x_{k})(x_{k}-x_{k+1})\cdots(x_{k}-x_{s+1})}$$

$$=\frac{(-1)^{k+1}}{(x_{1}-x_{k})\cdots(x_{l}-x_{k})\cdots(x_{k-1}-x_{k})(x_{k}-x_{k+1})\cdots(x_{k}-x_{s})}\left(\frac{1}{x_{l}-x_{k}}+\frac{1}{x_{k}-x_{s+1}}\right)$$

$$=\frac{(-1)^{k+1}(x_{l}-x_{s+1})}{(x_{1}-x_{k})\cdots(x_{k-1}-x_{k})(x_{k}-x_{k+1})\cdots(x_{k}-x_{s+1})}$$

$$=(-1)^{k+1}(x_{l}-x_{s+1})\frac{\prod_{i,j\in\mathbb{N}_{s+1}\{k\}}(x_{j}-x_{i})}{\prod_{i,j\in\mathbb{N}_{s+1}}(x_{j}-x_{i})},$$

which is the coefficient of y_k on the right-hand side.

Finally, when k = s + 1 the coefficient of y_{s+1} on the left-hand side of (46) is

$$0 - (-1)^{l+1} (-1)^{s-l} \frac{\prod_{i,j \in \mathbb{N}_{s+1} \setminus \{l,s+1\}} (x_j - x_i)}{\prod_{i,j \in \mathbb{N}_{s+1} \setminus \{l\}} (x_j - x_i)} = \frac{(-1)^{s+2}}{(x_1 - x_{s+1}) \cdots (x_l - x_{s+1}) \cdots (x_s - x_{s+1})}$$
$$= \frac{(-1)^{s+2} (x_l - x_{s+1})}{(x_1 - x_{s+1}) \cdots (x_s - x_{s+1})}$$
$$= (-1)^{s+2} (x_l - x_{s+1}) \frac{\prod_{i,j \in \mathbb{N}_{s+1} \setminus \{s+1\}} (x_j - x_i)}{\prod_{i,j \in \mathbb{N}_{s+1}} (x_j - x_i)}$$

which is again the coefficient of y_{s+1} on the right.

Theorem 6.8 Suppose $g \in C^k(I)$. Then for every permutation $\sigma \in \tilde{P}^s$, where $1 \leq s \leq k$, and every vector $(x_1, ..., x_s)$ with distinct coordinates, we have the formula

(47)
$$g^{[\sigma]}(x_1, ..., x_s) = \frac{V\binom{g'(x_1), ..., g'(x_s)}{x_1, ..., x_s}}{V(x_1, ..., x_s)}.$$

In particular, $g^{[\sigma]}(x_1,...,x_s)$ is a symmetric function.

Proof. The proof is by induction on s. When s = 1, then from the definitions we have

$$g^{[(1)]}(x_1) = g'(x_1) = \frac{V\binom{g'(x_1)}{x_1}}{V(x_1)}$$

Suppose (47) holds for s, where $1 \leq s < k$. Let $(x_1, ..., x_s, x_{s+1})$ be a vector with distinct coordinates and let $y = (g'(x_1), ..., g'(x_s), g'(x_{s+1}))$. Fix a permutation $\sigma \in \tilde{P}^s$ and an index $l \in \mathbb{N}_s$. Using (35) together with Lemma 6.7 and the induction hypothesis we get

$$g^{[\sigma_{(l)}]}(x_1, ..., x_s, x_{s+1}) = \frac{g^{[\sigma]}(x_1, ..., x_s) - g^{[\sigma]}(x_1, ..., x_{l-1}, x_{s+1}, x_{l+1}, ..., x_s)}{x_l - x_{s+1}}$$

$$= \frac{1}{(x_l - x_{s+1})} \left(\frac{V(x_{1}^{y_1, \dots, y_s})}{V(x_{1}, \dots, x_s)} - \frac{V(x_{1}^{y_1, \dots, y_{l-1}, y_{s+1}, y_{l+1}, \dots, y_s})}{V(x_{1}, \dots, x_{l-1}, x_{s+1}, x_{l+1}, \dots, x_s)} \right)$$
$$= \frac{V(x_{1}^{y_1, \dots, y_l, y_{s+1}, y_{l+1}, \dots, y_s})}{V(x_{1}, \dots, x_{l}, x_{s+1}, x_{l+1}, \dots, x_s)}$$
$$= \frac{V(x_{1}^{y_1, \dots, y_{l-1}, x_{s+1}, x_{l+1}, \dots, x_s)}}{V(x_{1}, \dots, x_{l-1}, x_{s+1}, x_{l+1}, \dots, x_s)}$$
$$= \frac{V(x_{1}^{y_1, \dots, y_{s+1}, x_{l+1}, \dots, x_s)}}{V(x_{1}, \dots, x_{l-1}, x_{s+1}, x_{l+1}, \dots, x_s)}$$

Since $\tilde{P}^{s+1} = \{\sigma_{(l)} \mid \sigma \in \tilde{P}^s, l \in \mathbb{N}_s\}$ the induction step is completed. Finally, since $g^{[\sigma]}(x_1, ..., x_s)$ is continuous, (47) shows that it is symmetric everywhere on its domain.

We can now significantly simplify Theorem 6.1. Define the k-tensor-valued map

(48)
$$\mathcal{A} : \mathbb{R}^n \to T^{\kappa,n}, \text{ by}$$
$$\left(\mathcal{A}(x)\right)^{i_1\dots i_k} := \frac{V\binom{g'(x_{i_1}), \dots, g'(x_{i_k})}{x_{i_1}, \dots, x_{i_k}}}{V(x_{i_1}, \dots, x_{i_k})}$$

Technically, this definition is good only when the numbers $x_{i_1}, ..., x_{i_k}$ are distinct, but Lemma 6.8 shows that it can be extended continuously everywhere. Clearly, if $(i_1, ..., i_{k+1}) \sim_x (j_1, ..., j_{k+1})$, then $(\mathcal{A}(x))^{i_1...i_{k+1}} = (\mathcal{A}(x))^{j_1...j_{k+1}}$, which shows that (48) defines a block-constant map. Moreover, $\mathcal{A}(x)$ is a symmetric tensor, continuous with respect to x.

Theorem 6.9 Let g be a C^k function defined on an interval I. Let X be a symmetric matrix with eigenvalues in the interval I, and let V be an orthogonal matrix such that $X = V(\text{Diag }\lambda(X))V^T$. Then the separable spectral function F defined by (31) and (32) is k-times continuously differentiable at X, and its k-th derivative is

(49)
$$\nabla^k F(X) = V\Big(\sum_{\sigma \in \tilde{P}^k} \operatorname{Diag}{}^{\sigma} \mathcal{A}(\lambda(X))\Big) V^T,$$

where $\mathcal{A}(x)$ is defined by (48). (\tilde{P}^k is the set of all permutations from P^k with exactly one cycle in their cycle decomposition.)

For most practical applications of derivatives, it is important to know what is the result when they are viewed as multi-linear maps and applied to vectors from the underlying space.

The last part of this subsection is devoted to the representations of the formula for the k-th derivative at X of a C^k separable spectral function, applied at k symmetric matrices.

6.3.1 The derivatives as multi-linear operators

The next corollary is a specialization of Theorem 6.9 to the case when k = 3. It should be compared with [2, Formula (V.22)]. One should keep in mind that we are differentiating separable spectral functions, whose gradients are the class of functions considered in [2, Chapter V].

Corollary 6.10 For $g \in C^3(I)$ and any $n \times n$ symmetric matrices H_1 , H_2 , H_3 we have

$$\nabla^3 F(X)[H_1, H_2, H_3] = 2 \sum_{p_1, p_2, p_3 = 1}^{n, n, n} \mathcal{A}(\lambda(X))^{p_1 p_2 p_3} \tilde{H}_1^{p_1 p_2} \tilde{H}_2^{p_2 p_3} \tilde{H}_3^{p_3 p_1}$$

where $X = V(\text{Diag }\lambda(X))V^T$, and $\tilde{H}_i = V^T H_i V$ for i = 1, 2, 3.

Proof. Without loss of generality suppose that $X = \text{Diag } \mu$ for some $\mu \in \mathbb{R}^n_{\downarrow}$. Then

$$\begin{aligned} \nabla^{3} F(\text{Diag } \mu)[H_{1}, H_{2}, H_{3}] &= \Big(\sum_{\sigma \in \tilde{P}^{3}} \text{Diag}^{\sigma} \mathcal{A}(\mu)\Big)[H_{1}, H_{2}, H_{3}] \\ &= \sum_{\sigma \in \tilde{P}^{3}} \langle \mathcal{A}(\mu), H_{1} \circ_{\sigma} H_{2} \circ_{\sigma} H_{3} \rangle \\ &= \langle \mathcal{A}(\mu), H_{1} \circ_{(123)} H_{2} \circ_{(123)} H_{3} \rangle + \langle \mathcal{A}(\mu), H_{1} \circ_{(132)} H_{2} \circ_{(132)} H_{3} \rangle \\ &= \sum_{q_{1}, q_{2}, q_{3} = 1}^{n, n, n} \mathcal{A}(\mu)^{q_{1}q_{2}q_{3}} H_{1}^{q_{1}q_{3}} H_{2}^{q_{2}q_{1}} H_{3}^{q_{3}q_{2}} + \sum_{p_{1}, p_{2}, p_{3} = 1}^{n, n, n} \mathcal{A}(\mu)^{p_{1}p_{2}p_{3}} H_{1}^{p_{1}p_{2}} H_{3}^{p_{3}p_{1}} \end{aligned}$$

After re-parametrization of the first sum $(q_1 = p_2, q_2 = p_3, q_3 = p_1)$, using the fact that $\mathcal{A}(\mu)$ is a symmetric tensor, and that matrices H_1 , H_2 , H_3 are symmetric, we continue

$$=\sum_{p_1,p_2,p_3=1}^{n,n,n} (\mathcal{A}(\mu)^{p_2p_3p_1} + \mathcal{A}(\mu)^{p_1p_2p_3})H_1^{p_1p_2}H_2^{p_2p_3}H_3^{p_3p_1} = 2\sum_{p_1,p_2,p_3=1}^{n,n,n} \mathcal{A}(\mu)^{p_1p_2p_3}H_1^{p_1p_2}H_2^{p_2p_3}H_3^{p_3p_1},$$

which is what we wanted to show.

In the general case when $H_1,...,H_k$ are distinct symmetric matrices, we cannot simplify the formula for $\nabla^k F(X)[H_1,...,H_k]$ much more than the example in Corollary 6.10.

To show that we can do at least that much, let σ and θ be in \tilde{P}^k , that is, permutations in P^k with one cycle in their cycle decomposition. Suppose that $\sigma = \theta^{-1}$, that \mathcal{A} is a symmetric k-tensor on \mathbb{R}^n , and that H_1, \ldots, H_k are distinct symmetric matrices. Then, re-parameterizing the sum

$$\sum_{q_1,\dots,q_k=1}^{n,\dots,n} \mathcal{A}^{q_1\dots q_k} H_1^{q_1 q_{\sigma^{-1}(1)}} \cdots H_k^{q_k q_{\sigma^{-1}(k)}}$$

according to the substitutions $q_i = p_{\sigma(i)}$ for i = 1, 2, ..., k we get the sum

$$\sum_{p_1,\dots,p_k=1}^{n,\dots,n} \mathcal{A}^{p_{\sigma(1)}\dots p_{\sigma(k)}} H_1^{p_{\sigma(1)}p_1} \cdots H_k^{p_{\sigma(k)}p_k} = \sum_{p_1,\dots,p_k=1}^{n,\dots,n} \mathcal{A}^{p_1\dots p_k} H_1^{p_{\sigma(1)}p_1} \cdots H_k^{p_{\sigma(k)}p_k}$$
$$= \sum_{p_1,\dots,p_k=1}^{n,\dots,n} \mathcal{A}^{p_1\dots p_k} H_1^{p_1p_{\theta^{-1}(1)}} \cdots H_k^{p_kp_{\theta^{-1}(k)}}.$$

In the first equality above we used the fact that \mathcal{A} is a symmetric tensor, while in the second we used that H_i is a symmetric matrix for i = 1, 2, ..., k.

We summarize the last paragraph in the following theorem.

Theorem 6.11 Let \tilde{P}_0^k be a subset of \tilde{P}^k , $k \ge 3$, such that if $\sigma \in \tilde{P}_0^k$ then $\sigma^{-1} \notin \tilde{P}_0^k$. For $g \in C^k(I)$ and any $n \times n$ symmetric matrices H_1, \ldots, H_k we have

(50)
$$\nabla^{k} F(X)[H_{1},...,H_{k}] = 2 \sum_{\sigma \in \tilde{P}_{0}^{k}} \sum_{p_{1},...,p_{k}=1}^{n,...,n} \mathcal{A}(\lambda(X))^{p_{1}...p_{k}} \tilde{H}_{1}^{p_{1}p_{\sigma(1)}} \cdots \tilde{H}_{k}^{p_{k}p_{\sigma(k)}},$$

where $X = V(\text{Diag }\lambda(X))V^T$, and $\tilde{H}_i = V^T H_i V$ for i = 1, 2, ..., k.

If, in the above theorem, all matrices $H_1, ..., H_k$ are the same then Formula (50) can be simplified even more.

Theorem 6.12 For $g \in C^k(I)$ and any $n \times n$ symmetric matrix H

(51)
$$\nabla^{k} F(X)[H,...,H] = (k-1)! \sum_{p_{1},...,p_{k}=1}^{n,...,n} \mathcal{A}(\lambda(X))^{p_{1}...p_{k}} \tilde{H}^{p_{1}p_{2}} \tilde{H}^{p_{2}p_{3}} \cdots \tilde{H}^{p_{k}p_{1}},$$

where $X = V(\text{Diag }\lambda(X))V^T$, and $\tilde{H} = V^T H V$.

Proof. Let H be any $n \times n$ symmetric matrix. Using formulae (49), (10), and (5) we find

$$\nabla^{k} F(X)[H,...,H] = V\Big(\sum_{\sigma \in \tilde{P}^{k}} \operatorname{Diag}^{\sigma} \mathcal{A}(\lambda(X))\Big) V^{T}[H,...,H]$$

$$= \sum_{\sigma \in \tilde{P}^{k}} \langle \mathcal{A}(\lambda(X)), \tilde{H} \circ_{\sigma} \tilde{H} \circ_{\sigma} \cdots \circ_{\sigma} \tilde{H} \rangle$$

$$= \sum_{\sigma \in \tilde{P}^{k}} \sum_{p_{1},...,p_{k}=1}^{n,...,n} \mathcal{A}(\lambda(X))^{p_{1}...p_{k}} \tilde{H}^{p_{1}p_{\sigma}-1}{}_{(1)} \cdots \tilde{H}^{p_{k}p_{\sigma}-1}{}_{(k)}.$$

Let $\sigma \in \tilde{P}^k$ be any permutation with one cycle in its cycle decomposition. In order to prove the result we are going to show that

(52)
$$\sum_{q_1,\dots,q_k=1}^{n,\dots,n} \mathcal{A}(\lambda(X))^{q_1\dots q_k} \tilde{H}^{q_1 q_{\sigma^{-1}(1)}} \cdots \tilde{H}^{q_k q_{\sigma^{-1}(k)}} = \sum_{p_1,\dots,p_k=1}^{n,\dots,n} \mathcal{A}(\lambda(X))^{p_1\dots p_k} \tilde{H}^{p_1 p_2} \tilde{H}^{p_2 p_3} \cdots \tilde{H}^{p_k p_1}.$$

In order to do that we find a re-parametrization (that is, we change the order of summation) of the right-hand side sum that will give the left-hand side sum. Since σ has one cycle in its cycle

decomposition, the map $i \in \mathbb{N}_k \mapsto \sigma^{-i}(1) \in \mathbb{N}_k$ is a permutation as well. Change the order of summation in the right-hand side of (52) according to the rule

$$p_i := q_{\sigma^{-i}(1)}$$
 for all $i = 1, 2, ..., k$.

Notice that we have $p_{i+1} = q_{\sigma^{-(i+1)}(1)} = q_{\sigma^{-1}(\sigma^{-i}(1))}$. The product $\tilde{H}^{p_1p_2}\tilde{H}^{p_2p_3}\cdots\tilde{H}^{p_kp_1}$ after the substitution goes into the product

$$\begin{split} \tilde{H}^{q_{\sigma^{-1}(1)}q_{\sigma^{-2}(1)}} \tilde{H}^{q_{\sigma^{-2}(1)}q_{\sigma^{-3}(1)}} \cdots \tilde{H}^{q_{\sigma^{-k}(1)}q_{\sigma^{-1}(1)}} \\ &= \tilde{H}^{q_{\sigma^{-1}(1)}q_{\sigma^{-1}(\sigma^{-1}(1))}} \tilde{H}^{q_{\sigma^{-2}(1)}q_{\sigma^{-1}(\sigma^{-2}(1))}} \cdots \tilde{H}^{q_{\sigma^{-k}(1)}q_{\sigma^{-1}(\sigma^{-k}(1))}} \\ &= \tilde{H}^{q_{1}q_{\sigma^{-1}(1)}} \tilde{H}^{q_{2}q_{\sigma^{-1}(2)}} \cdots \tilde{H}^{q_{k}q_{\sigma^{-1}(k)}}. \end{split}$$

The last equality follows by a reordering of the product since the indexes $\{\sigma^{-1}(1), \sigma^{-2}(1), ..., \sigma^{-k}(1)\}$ are a permutation of the indexes $\{1, 2, ..., k\}$. Finally we have

$$\mathcal{A}(\lambda(X))^{p_1\dots p_k} = \mathcal{A}(\lambda(X))^{q_{\theta}-1} \dots q_{\theta}-k} (1) = \mathcal{A}(\lambda(X))^{q_1\dots q_k}$$

since $\mathcal{A}(\lambda(X))$ is a symmetric tensor and the indexes $\{\sigma^{-1}(1), \sigma^{-2}(1), ..., \sigma^{-k}(1)\}$ are a permutation of the indexes $\{1, 2, ..., k\}$.

7 The Hessian of a general spectral function

In this section we calculate the formula for the Hessian of a general spectral functions at an arbitrary symmetric matrix. The formula was first obtained in [16] but the insight for it came from [17]. Below, as another application of the tools developed so far, we derive it again. The approach it more streamlined and clearly shows where the different pieces of the Hessian come from.

7.1 Two matrix-valued maps

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a symmetric twice (continuously) differentiable function. Let $\mathcal{A}_{(1)(2)} : \mathbb{R}^n \to M^n$ be defined by

$$\mathcal{A}_{(1)(2)}(x) = \nabla^2 f(x),$$

and let $\mathcal{A}_{(12)}$: $\mathbb{R}^n \to M^n$ be defined entry-wise by

$$\mathcal{A}_{(12)}^{i_1 i_2}(x) = \begin{cases} 0, & \text{if } i_1 = i_2 \\ f_{i_1 i_1}''(x) - f_{i_1 i_2}''(x), & \text{if } i_1 \sim_x i_2 \text{ and } i_1 \neq i_2 \\ \frac{f_{i_2}'(x) - f_{i_1}'(x)}{x_{i_2} - x_{i_1}}, & \text{if } i_1 \not\sim_x i_2. \end{cases}$$

Several of the properties of $\mathcal{A}_{(12)}(x)$ are easily seen from the following integral representation.

Lemma 7.1 If f is a C^2 function, then for every $i_1, i_2 \in \mathbb{N}_n$ we have

$$\mathcal{A}_{(12)}^{i_1i_2}(x) = \int_0^1 f_{i_1i_1}''(\dots, x_{i_1} + t(x_{i_2} - x_{i_1}), \dots, x_{i_2} + t(x_{i_1} - x_{i_2}), \dots) - f_{i_1i_2}''(\dots, x_{i_1} + t(x_{i_2} - x_{i_1}), \dots, x_{i_2} + t(x_{i_1} - x_{i_2}), \dots) dt,$$

where the first displayed argument is in position i_1 and the second is in position i_2 . The missing arguments are the corresponding entries of x, unchanged.

Proof. The first case, when $i_1 = i_2$ is immediate. In the second, $i_l \sim_x i_2$ implies that $x_{i_1} = x_{i_2}$ and the integrand doesn't depend on t. In the third case, $i_1 \not\sim_x i_2$, using the Fundamental Theorem of Calculus, the integral is equal to:

$$\frac{1}{x_{i_2} - x_{i_1}} \int_0^1 \frac{\partial}{\partial t} f'_{i_1}(\dots, x_{i_1} + t(x_{i_2} - x_{i_1}), \dots, x_{i_2} + t(x_{i_1} - x_{i_2}), \dots) dt$$

$$= \frac{f'_{i_1}(\dots, x_{i_2}, \dots, x_{i_1}, \dots) - f'_{i_1}(\dots, x_{i_1}, \dots, x_{i_2}, \dots)}{x_{i_2} - x_{i_1}}$$

$$= \frac{f'_{i_2}(\dots, x_{i_1}, \dots, x_{i_2}, \dots) - f'_{i_1}(\dots, x_{i_1}, \dots, x_{i_2}, \dots)}{x_{i_2} - x_{i_1}}$$

$$= \mathcal{A}_{(12)}^{i_1 i_2}(x).$$

In the second equality we used that $x \mapsto \nabla f(x)$ is a point-symmetric map.

Lemma 7.2 If f(x) is twice (continuously) differentiable, then both $\mathcal{A}_{(1)(2)}(x)$ and $\mathcal{A}_{(12)}(x)$ are point-symmetric maps.

Proof. The fact that $x \mapsto \mathcal{A}_{(1)(2)}(x)$ is a point-symmetric map is Lemma 2.5. This implies that if $i_1 \sim_x j_1$, then $f''_{i_1i_1}(x) = f''_{j_1j_1}(\mu)$. Also, if $i_1 \sim_x j_1$ and $i_2 \sim_x j_2$ with $i_1 \neq i_2$ and $j_1 \neq j_2$, then $f''_{i_1i_2}(x) = f''_{j_1j_2}(x)$. The fact that $x \mapsto \nabla f(x)$ is a point-symmetric map implies that if $i_1 \sim_x j_1$, then $f'_{i_1}(x) = f''_{j_1}(x)$. Now it is easy to see that $x \mapsto \mathcal{A}_{(12)}(x)$ is a point-symmetric map as well.

It is easy to see, using Lemma 7.1, that if f(x) is twice continuously differentiable function, then $\mathcal{A}_{(1)(2)}(x)$ and $\mathcal{A}_{(12)}(x)$ are symmetric matrices, continuous in x.

7.2 $f \circ \lambda$ is twice (continuously) differentiable if and only if f is

We now show that $f \circ \lambda$ is twice (continuously) differentiable at X if and only if f is such at $\lambda(X)$. The 'only if' direction can be seen by restricting $f \circ \lambda$ to the subspace of diagonal matrices. Below, we show the 'if' direction. Without loss of generality, assume that $X = \text{Diag } \mu$, for some $\mu \in \mathbb{R}^n_{\downarrow}$, that $M_m/||M_m||$ converges to M as m goes to infinity, and that (20) holds. Using (30) together with (24) we compute:

$$\nabla^{2}(f \circ \lambda)(\operatorname{Diag} \mu)[M] = \lim_{m \to \infty} \frac{\nabla(f \circ \lambda)(\operatorname{Diag} \mu + M_{m}) - \nabla(f \circ \lambda)(\operatorname{Diag} \mu)}{\|M_{m}\|}$$

$$= \lim_{m \to \infty} \frac{U_{m}(\operatorname{Diag}^{(1)}\nabla f(\lambda(\operatorname{Diag} \mu + M_{m})))U_{m}^{T} - \operatorname{Diag}^{(1)}\nabla f(\mu)}{\|M_{m}\|}$$

$$= \lim_{m \to \infty} \frac{U_{m}(\operatorname{Diag}^{(1)}\nabla f(\mu + h_{m} + o(\|M_{m}\|)))U_{m}^{T} - \operatorname{Diag}^{(1)}\nabla f(\mu)}{\|M_{m}\|}$$

$$= \lim_{m \to \infty} \frac{U_{m}(\operatorname{Diag}^{(1)}(\nabla f(\mu) + \nabla^{2}f(\mu)[h_{m}] + o(\|M_{m}\|)))U_{m}^{T} - \operatorname{Diag}^{(1)}\nabla f(\mu)}{\|M_{m}\|}$$
(53)
$$= \lim_{m \to \infty} \frac{U_{m}(\operatorname{Diag}^{(1)}(\nabla f(\mu)))U_{m}^{T} - \operatorname{Diag}^{(1)}\nabla f(\mu)}{\|M_{m}\|} + U(\operatorname{Diag}^{(1)}(\nabla^{2}f(\mu)[h]))U^{T}.$$

For brevity let $T = \nabla f(\mu)$, let $\mathcal{A}_{(1)(2)} = \mathcal{A}_{(1)(2)}(\mu)$, and let $\mathcal{A}_{(12)} = \mathcal{A}_{(12)}(\mu)$. Using Corollary 2.9

(54)
$$\lim_{m \to \infty} \frac{U_m (\operatorname{Diag}^{(1)} T) U_m^T - \operatorname{Diag}^{(1)} T}{\|M_m\|} = (\operatorname{Diag}^{(12)} T_{\operatorname{out}}^{(1)}) [M].$$

By Lemma 2.1 part (ii), there is a vector b block-constant with respect to μ such that the matrix $\mathcal{A}_{(1)(2)}$ – Diag b is also block-constant with respect to μ . Then by Corollary 2.10, applied with k = 1,

$$U(\operatorname{Diag}^{(1)}(\nabla^{2} f(\mu)[h]))U^{T} = U(\operatorname{Diag}^{(1)}((\mathcal{A}_{(1)(2)} - \operatorname{Diag} b + \operatorname{Diag} b)[h]))U^{T} = U(\operatorname{Diag}^{(1)}((\mathcal{A}_{(1)(2)} - \operatorname{Diag} b)[h]))U^{T} + U(\operatorname{Diag}^{(1)}((\operatorname{Diag} b)[h]))U^{T} = (\operatorname{Diag}^{(1)(2)}(\mathcal{A}_{(1)(2)} - \operatorname{Diag} b))[M] + (\operatorname{Diag}^{(12)}b_{\mathrm{in}}^{(1)})[M].$$

This shows that $f \circ \lambda$ is twice differentiable.

To prove that $f \circ \lambda$ is twice continuously differentiable we need to reorganize the pieces. Direct verification shows that the sum $\mathcal{A}_{(1)(2)} + \mathcal{A}_{(12)}$ is a block-constant matrix. Then vector b can be chosen in such a way that, in addition, $\mathcal{A}_{(12)} + \text{Diag } b$ is a block-constant constant matrix, and

(56)
$$\mathcal{A}_{(12)} + \text{Diag}\,b = T_{\text{out}}^{(1)} + b_{\text{in}}^{(1)}.$$

Putting (53), (54), (55), and (56) together we obtain:

$$\nabla^{2}(f \circ \lambda)(\operatorname{Diag} \mu) = \operatorname{Diag}^{(12)}T_{\operatorname{out}}^{(1)} + \operatorname{Diag}^{(1)(2)}(\mathcal{A}_{(1)(2)} - \operatorname{Diag} b) + \operatorname{Diag}^{(12)}b_{\operatorname{in}}^{(1)}$$

= $\operatorname{Diag}^{(1)(2)}(\mathcal{A}_{(1)(2)} - \operatorname{Diag} b) + \operatorname{Diag}^{(12)}(\mathcal{A}_{(12)} + \operatorname{Diag} b)$
= $\operatorname{Diag}^{(1)(2)}\mathcal{A}_{(1)(2)} + \operatorname{Diag}^{(12)}\mathcal{A}_{(12)}.$

In the last equality we used the fact that $\operatorname{Diag}^{(1)(2)}(\operatorname{Diag} b) = \operatorname{Diag}^{(12)}(\operatorname{Diag} b)$, which can be verified directly. The formula for the Hessian of $f \circ \lambda$ at an arbitrary X, can now be derived routinely:

(57)
$$\nabla^2 (f \circ \lambda)(X) = V \left(\operatorname{Diag}^{(1)(2)} \mathcal{A}_{(1)(2)}(\lambda(X)) + \operatorname{Diag}^{(12)} \mathcal{A}_{(12)}(\lambda(X)) \right) V^T,$$

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

Finally, when f is C^2 both $\mathcal{A}_{(1)(2)}(x)$ and $\mathcal{A}_{(12)}(x)$ are continuous and by [21, Proposition 6.2] $\nabla^2(f \circ \lambda)(X)$ is continuous as well.

8 Appendix A: A refinement of a perturbation result for eigenvectors

The main tool in the derivation of the formula for the Hessian in [16] was Lemma 2.4. The statement of that lemma was broken down into nine parts, and that lead to the consideration of variety of cases when deriving the Hessian. For the higher-order derivatives such case studies would quickly become unmanageable. That is why the goal of this appendix is to transform Lemma 2.4 from [16] into a form more suitable for computations. Consult with Section 3 for the relevant notation.

Recall that any vector $\mu \in \mathbb{R}^n$ defines a partition of \mathbb{N}_n into disjoint *blocks*, where integers *i* and *j* are in the same block if and only if $\mu_i = \mu_j$. By *r* we denote the number of blocks in the partition. By ι_l we denote the largest integer in I_l for all l = 1, ..., r.

Theorem 8.1 Let $\{M_m\}_{m=1}^{\infty}$ be a sequence of symmetric matrices converging to 0, such that the normalized sequence $M_m/||M_m||$ converges to M. Let μ be in $\mathbb{R}^n_{\downarrow}$ and let $U_m \to U \in O^n$ be a sequence of orthogonal matrices such that

$$\operatorname{Diag} \mu + M_m = U_m (\operatorname{Diag} \lambda (\operatorname{Diag} \mu + M_m)) U_m^T$$
, for all $m = 1, 2, \dots$

Then:

(i) The orthogonal matrix U has the form

$$U = \begin{pmatrix} V_1 & 0 & \cdots & 0 \\ 0 & V_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & V_r \end{pmatrix},$$

where V_l is an orthogonal matrix with dimensions $|I_l| \times |I_l|$ for all l.

(ii) The following identity holds

(58)
$$U^T M_{\rm in} U = {\rm Diag} h,$$

(iii) For any indexes $i \in I_l$, $j \in I_s$, and $t \in \{1, ..., r\}$ we have the (strong) first-order expansion

(59)
$$\sum_{p \in I_t} U_m^{ip} U_m^{jp} = \delta_{ij} \delta_{lt} + \frac{\delta_{lt} - \delta_{st}}{\mu_i - \mu_j} M^{ij} \|M_m\| + o(\|M_m\|),$$

with the understanding that the fraction is zero whenever $\delta_{lt} = \delta_{st}$ no matter what the denominator is.

Proof. This lemma, with some modifications, is essentially Lemma 2.4 in [16]. Indeed, Part (i) is [16, Lemma 2.4 Part (i)]. The equality in Part (ii) is an aggregate version of Parts (iv) and (vii) from Lemma 2.4 in [16]. To prove Part (iii) we consider several cases.

- **Case 1.** If $i = j \in I_l$ and t = l, then (59) becomes $\sum_{p \in I_l} (U_m^{ip})^2 = 1 + o(||M_m||)$, which is exactly Part (ii), Lemma 2.4 in [16].
- **Case 2.** If $i = j \in I_l$ and $t \neq l$, then (59) becomes $\sum_{p \in I_t} (U_m^{ip})^2 = o(||M_m||)$, which is a consequence of Part (iii), Lemma 2.4 in [16].
- **Case 3.** If $i \neq j \in I_l$ and t = l, then (59) becomes $\sum_{p \in I_l} U_m^{ip} U_m^{jp} = o(||M_m||)$, which is exactly Part (vi), Lemma 2.4 in [16].
- **Case 4.** If $i \neq j \in I_l$ and $t \neq l$, then (59) becomes $\sum_{p \in I_t} U_m^{ip} U_m^{jp} = o(||M_m||)$, which is a consequence of Part (v), Lemma 2.4 in [16].
- **Case 5.** If $i \in I_l$, $j \in I_s$, with $l \neq s \neq t \neq l$, then (59) becomes $\sum_{p \in I_t} U_m^{ip} U_m^{jp} = o(||M_m||)$, which is a consequence of Part (viii), Lemma 2.4 in [16].

Case 6. If $i \in I_l$, $j \in I_s$, with $l \neq s$ and t = l, then (59) becomes

$$\sum_{p \in I_t} U_m^{ip} U_m^{jp} = \frac{1}{\mu_i - \mu_j} M^{ij} \|M_m\| + o(\|M_m\|),$$

which we prove in Case 7.

Case 7. If $i \in I_l$, $j \in I_s$, with $l \neq s$ and t = s, then (59) becomes

$$\sum_{p \in I_t} U_m^{ip} U_m^{jp} = -\frac{1}{\mu_i - \mu_j} M^{ij} \|M_m\| + o(\|M_m\|).$$

We now show that the expressions in both Case 6 and Case 7 are valid. Recall that Part (ix) from Lemma 2.4 in [16] says that in case when $i \in I_l$, $j \in I_s$ with $l \neq s$, we have

(60)
$$\lim_{m \to \infty} \left(\mu_{\iota_l} \frac{\sum_{p \in I_l} U_m^{ip} U_m^{jp}}{\|M_m\|} + \mu_{\iota_s} \frac{\sum_{p \in I_s} U_m^{ip} U_m^{jp}}{\|M_m\|} \right) = M^{ij}.$$

Introduce the notation

$$\beta_m^l := \frac{\sum_{p \in I_l} U_m^{ip} U_m^{jp}}{\|M_m\|}, \quad \text{for all} \quad l = 1, 2, ..., r,$$

and notice that

$$\sum_{l=1}^{r} \beta_m^l = 0, \quad \text{for all } m,$$

because U_m is an orthogonal matrix and the numerator of the last sum is the product of its *i*-th and *j*-th row. Next, by Case 5 we have

$$\lim_{m \to \infty} \sum_{t \neq l, s} \beta_m^t = 0,$$

 \mathbf{SO}

$$\lim_{m \to \infty} (\beta_m^l + \beta_m^s) = 0.$$

For arbitrary reals a and b we compute

$$\left(a\beta_m^l + b\beta_m^s\right) - \frac{a-b}{\mu_{\iota_l} - \mu_{\iota_s}} \left(\mu_{\iota_l}\beta_m^l + \mu_{\iota_s}\beta_m^s\right) = \left(\beta_m^l + \beta_m^s\right) \frac{b\mu_{\iota_l} - a\mu_{\iota_s}}{\mu_{\iota_l} - \mu_{\iota_s}} \to 0,$$

as $m \to \infty$. Using (60), this shows that

$$\lim_{m \to \infty} (a\beta_m^l + b\beta_m^s) = \frac{a-b}{\mu_{\iota_l} - \mu_{\iota_s}} M^{ij}$$

When (a, b) = (1, 0) we obtain Case 6, and when (a, b) = (0, 1) we obtain Case 7.

9 Appendix B: Tensor analysis

The aim of this appendix is to provide the proofs of Theorems 2.9 and 2.10.

Recall that any vector $\mu \in \mathbb{R}^n$ defines a partition of \mathbb{N}_n into disjoint *blocks*, where integers *i* and *j* are in the same block if and only if $\mu_i = \mu_j$. By *r* we denote the number of blocks in the partition. By ι_l we denote the largest integer in I_l for all l = 1, ..., r.

Theorem 9.1 Let $\{M_m\}_{m=1}^{\infty}$ be a sequence of symmetric matrices converging to 0, such that the normalized sequence $M_m/||M_m||$ converges to M. Let μ be in $\mathbb{R}^n_{\downarrow}$ and $U_m \to U \in O^n$ be a sequence of orthogonal matrices such that

$$\operatorname{Diag} \mu + M_m = U_m \left(\operatorname{Diag} \lambda (\operatorname{Diag} \mu + M_m) \right) U_m^T, \text{ for all } m = 1, 2, \dots$$

Then for every block-constant k-tensor T on \mathbb{R}^n , any matrices H_1, \ldots, H_k , and any permutation σ on \mathbb{N}_k we have

(61)
$$\lim_{m \to \infty} \left(\frac{U_m(\operatorname{Diag}^{\sigma} T)U_m^T - \operatorname{Diag}^{\sigma} T}{\|M_m\|} \right) [H_1, ..., H_k] = \sum_{l=1}^k \left(\operatorname{Diag}^{\sigma_{(l)}} T_{\operatorname{out}}^{(l)} \right) [H_1, ..., H_k, M_{\operatorname{out}}],$$

where M_{out} is the symmetric matrix of off-diagonal blocks of M as defined by (4).

Proof. The idea of the proof is to evaluate separately the expressions on both sides of (61) and compare the results. Notice that both sides of (61) are linear in each argument H_s . That is why it is enough to prove the result when H_s , for s = 1, ..., k, is an arbitrary matrix, $H_{i_s j_s}$, from the standard basis on M^n . In that case notice that

(62)
$$\left(U_m(\operatorname{Diag}^{\sigma}T)U_m^T - \operatorname{Diag}^{\sigma}T \right) [H_{i_1j_1}, ..., H_{i_kj_k}] = \left(U_m(\operatorname{Diag}^{\sigma}T)U_m^T \right)^{i_1...i_k}_{j_1...j_k} - \left(\operatorname{Diag}^{\sigma}T \right)^{i_1...i_k}_{j_1...j_k}.$$

Using the definition of the conjugate action and the fact that T is block-constant, we develop the first term on the right-hand side of the equality sign in (62):

$$(U_{m}(\operatorname{Diag}{}^{\sigma}T)U_{m}^{T})^{i_{1}...i_{k}}_{j_{1}...j_{k}} = \sum_{\substack{p_{\eta},q_{\eta}=1\\\eta=1,...,k}}^{n,...,n} (\operatorname{Diag}{}^{\sigma}T)^{p_{1}...p_{k}}_{q_{1}...q_{k}} \prod_{\nu=1}^{k} U^{i_{\nu}p_{\nu}}_{m} U^{j_{\nu}q_{\nu}}_{m}$$
$$= \sum_{\substack{p_{\eta}=1\\\eta=1,...,k}}^{n,...,n} T^{p_{1}...p_{k}} \prod_{\nu=1}^{k} U^{i_{\nu}p_{\nu}}_{m} U^{j_{\nu}p_{\sigma-1}(\nu)}_{m}$$
$$= \sum_{\substack{p_{\eta}=1\\\eta=1,...,k}}^{n,...,n} T^{p_{1}...p_{k}} \prod_{\nu=1}^{k} U^{i_{\nu}p_{\nu}}_{m} U^{j_{\sigma}(\nu)p_{\nu}}_{m}$$
$$= \sum_{\substack{q_{\eta}=1\\\eta=1,...,k}}^{r,...,r} T^{\iota_{t_{1}}...\iota_{t_{k}}} \prod_{\nu=1}^{k} \left(\sum_{p_{\nu}\in I_{t_{\nu}}} U^{i_{\nu}p_{\nu}}_{m} U^{j_{\sigma}(\nu)p_{\nu}}_{m}\right).$$

Putting everything together, we see that to evaluate the limit on the left-hand side of (61) we have to compute

(63)
$$\lim_{m \to \infty} \frac{\sum_{t_1, \dots, t_k = 1}^{r, \dots, r} T^{\iota_{t_1} \dots \iota_{t_k}} \prod_{\nu = 1}^k \left(\sum_{p_\nu \in I_{t_\nu}} U_m^{i_\nu p_\nu} U_m^{j_{\sigma(\nu)} p_\nu} \right) - (\text{Diag}\,^{\sigma} T)^{j_1 \dots j_k}}{\|M_m\|}.$$

Assume that $i_l \in I_{v_l}$ and $j_{\sigma(l)} \in I_{s_l}$ for all l = 1, ..., k. We investigate several possibilities. Suppose first that among the pairs

(64)
$$(i_1, j_{\sigma(1)}), (i_2, j_{\sigma(2)}), \dots, (i_k, j_{\sigma(k)})$$

at least two have nonequal entries. Without loss of generality we may assume they are $(i_1, j_{\sigma(1)})$ and $(i_2, j_{\sigma(2)})$, that is, $i_1 \neq j_{\sigma(1)}$ and $i_2 \neq j_{\sigma(2)}$. Using (59), for any t_1, t_2 we observe that:

$$\lim_{m \to \infty} \frac{1}{\|M_m\|} \Big(\sum_{p_1 \in I_{t_1}} U_m^{i_1 p_1} U_m^{j_{\sigma(1)} p_1} \Big) \Big(\sum_{p_2 \in I_{t_2}} U_m^{i_2 p_2} U_m^{j_{\sigma(2)} p_2} \Big)$$

$$= \lim_{m \to \infty} \frac{1}{\|M_m\|} \Big(\delta_{i_1 j_{\sigma(1)}} \delta_{v_1 t_1} + \frac{\delta_{v_1 t_1} - \delta_{s_1 t_1}}{\mu_{i_1} - \mu_{j_{\sigma(1)}}} M^{i_1 j_{\sigma(1)}} \|M_m\| + o(\|M_m\|) \Big) \times \\ \Big(\delta_{i_2 j_{\sigma(2)}} \delta_{v_2 t_2} + \frac{\delta_{v_2 t_2} - \delta_{s_2 t_2}}{\mu_{i_2} - \mu_{j_{\sigma(2)}}} M^{i_2 j_{\sigma(2)}} \|M_m\| + o(\|M_m\|) \Big) \\ = \lim_{m \to \infty} \frac{1}{\|M_m\|} \Big(\frac{\delta_{v_1 t_1} - \delta_{s_1 t_1}}{\mu_{i_1} - \mu_{j_{\sigma(1)}}} M^{i_1 j_{\sigma(1)}} \|M_m\| + o(\|M_m\|) \Big) \Big(\frac{\delta_{v_2 t_2} - \delta_{s_2 t_2}}{\mu_{i_2} - \mu_{j_{\sigma(2)}}} M^{i_2 j_{\sigma(2)}} \|M_m\| + o(\|M_m\|) \Big) \\ = 0.$$

Since in this case by definition $(\text{Diag}^{\sigma}T)^{j_1...j_k} = 0$ we see that (63) in zero.

Suppose now, that exactly one pair has unequal entries and let it be $(i_l, j_{\sigma(l)})$. We consider two subcases depending on whether or not i_l and $j_{\sigma(l)}$ are in the same block.

If both i_l and $j_{\sigma(l)}$ are in one block, that is $v_l = s_l$, then using (59), for arbitrary t, we obtain:

$$\lim_{m \to \infty} \frac{1}{\|M_m\|} \left(\sum_{p \in I_t} U_m^{i_l p} U_m^{j_{\sigma(l)} p} \right) = \lim_{m \to \infty} \frac{1}{\|M_m\|} \left(\delta_{i_l j_{\sigma(l)}} \delta_{v_l t} + \frac{\delta_{v_l t} - \delta_{s_l t}}{\mu_{i_l} - \mu_{j_{\sigma(l)}}} M^{i_l j_{\sigma(l)}} \|M_m\| + o(\|M_m\|) \right)$$
$$= \lim_{m \to \infty} \frac{o(\|M_m\|)}{\|M_m\|}$$
$$= 0.$$

In this subcase we again have $(\text{Diag}^{\sigma}T)_{j_1...j_k}^{i_1...i_k} = 0$, thus (63) is equal to zero.

If i_l and $j_{\sigma(l)}$ are in different blocks, $v_l \neq s_l$, then $(\text{Diag}^{\sigma}T)^{i_1...i_k}_{j_1...j_k} = 0$ and by (59) we obtain:

$$\lim_{m \to \infty} \frac{1}{\|M_m\|} \Big(\sum_{t_1, \dots, t_k=1}^{r, \dots, r} T^{\iota_{t_1} \dots \iota_{t_k}} \prod_{\nu=1}^k \Big(\sum_{p_\nu \in I_{t_\nu}} U_m^{i_\nu p_\nu} U_m^{j_{\sigma(\nu)} p_\nu} \Big) \Big) =$$
(65)
$$\lim_{m \to \infty} \frac{1}{\|M_m\|} \Big(\sum_{t_1, \dots, t_k=1}^{r, \dots, r} T^{\iota_{t_1} \dots \iota_{t_k}} \prod_{\nu=1}^k \Big(\delta_{i_\nu j_{\sigma(\nu)}} \delta_{v_\nu t_\nu} + \frac{\delta_{v_\nu t_\nu} - \delta_{s_\nu t_\nu}}{\mu_{i_\nu} - \mu_{j_{\sigma(\nu)}}} M^{i_\nu j_{\sigma(\nu)}} \|M_m\| + o(\|M_m\|) \Big) \Big).$$

We show that the limit of at most two terms of the big sum in (65) may be non-zero. Indeed, summands corresponding to k-tuples $(t_1, ..., t_k)$ with $t_l \notin \{v_l, s_l\}$ converge to zero, because $\delta_{i_l j_{\sigma(l)}} = 0$, $\delta_{v_l t_l} = \delta_{s_l t_l} = 0$, and therefore

$$\delta_{i_l j_{\sigma(l)}} \delta_{v_l t_l} + \frac{\delta_{v_l t_l} - \delta_{s_l t_l}}{\mu_{i_l} - \mu_{j_{\sigma(l)}}} M^{i_l j_{\sigma(l)}} \|M_m\| + o(\|M_m\|) = o(\|M_m\|).$$

Similarly, summands corresponding to k-tuples $(t_1, ..., t_k)$ with $t_{\nu} \neq v_{\nu}$ for some $\nu \neq l$ converge to zero, since then $\delta_{v_{\nu}t_{\nu}} = \delta_{s_{\nu}t_{\nu}} = 0$ (recall that $v_{\nu} = s_{\nu}$ for all $\nu \neq l$). Thus, there are two summands with possible nonzero limit. The first corresponding to the k-tuple $(v_1, ..., v_{l-1}, v_l, v_{l+1}, ..., v_k)$ and the second corresponding to the k-tuple $(v_1, ..., v_{l-1}, s_l, v_{l+1}, ..., v_k)$. Notice finally that if $t_{\nu} = v_{\nu}(=s_{\nu})$ for some $\nu \neq l$, then

$$\delta_{i_{\nu}j_{\sigma(\nu)}}\delta_{v_{\nu}t_{\nu}} + \frac{\delta_{v_{\nu}t_{\nu}} - \delta_{s_{\nu}t_{\nu}}}{\mu_{i_{\nu}} - \mu_{j_{\sigma(\nu)}}} M^{i_{\nu}j_{\sigma(\nu)}} \|M_m\| + o(\|M_m\|) = 1 + o(\|M_m\|),$$

since $i_{\nu} = j_{\sigma(\nu)}$ for $\nu \neq l$. Thus, the limit of the summand in (65) corresponding to the k-tuple $(v_1, ..., v_{l-1}, v_l, v_{l+1}, ..., v_k)$ is equal to

$$\lim_{m \to \infty} \frac{T^{\iota_{v_1} \dots \iota_{v_{l-1}} \iota_{v_l} \iota_{v_{l+1}} \dots \iota_{v_k}}}{\|M_m\|} \Big(\delta_{i_l j_{\sigma(l)}} \delta_{v_l v_l} + \frac{\delta_{v_l v_l} - \delta_{s_l v_l}}{\mu_{i_l} - \mu_{j_{\sigma(l)}}} M^{i_l j_{\sigma(l)}} \|M_m\| + o(\|M_m\|) \Big) \Big(1 + o(\|M_m\|) \Big) \\ = \frac{T^{\iota_{v_1} \dots \iota_{v_{l-1}} \iota_{v_l} \iota_{v_{l+1}} \dots \iota_{v_k}}}{\mu_{i_l} - \mu_{j_{\sigma(l)}}} M^{i_l j_{\sigma(l)}},$$

while, analogously, the limit corresponding to the k-tuple $(v_1, ..., v_{l-1}, s_l, v_{l+1}, ..., v_k)$ is equal to

$$-\frac{T^{\iota_{v_1}...\iota_{v_{l-1}}\iota_{s_l}\iota_{v_{l+1}}...\iota_{v_k}}{\mu_{i_l}-\mu_{j_{\sigma(l)}}}M^{i_lj_{\sigma(l)}}.$$

Putting these two limits together we see that (65), and therefore (63), is equal to

$$\frac{T^{\iota_{v_1}\ldots\iota_{v_{l-1}}\iota_{v_l}\iota_{v_{l+1}}\ldots\iota_{v_k}} - T^{\iota_{v_1}\ldots\iota_{v_{l-1}}\iota_{s_l}\iota_{v_{l+1}}\ldots\iota_{v_k}}}{\mu_{i_l} - \mu_{j_{\sigma(l)}}} M^{i_lj_{\sigma(l)}} = \frac{T^{i_1\ldots i_{l-1}i_li_{l+1}\ldots i_k} - T^{i_1\ldots i_{l-1}j_{\sigma(l)}i_{l+1}\ldots i_k}}{\mu_{i_l} - \mu_{j_{\sigma(l)}}} M^{i_lj_{\sigma(l)}} = \frac{T^{i_1\ldots i_{l-1}i_li_{l+1}\ldots i_k} - T^{i_1\ldots i_{l-1}j_{\sigma(l)}i_{l+1}\ldots i_k}}{\mu_{i_l} - \mu_{j_{\sigma(l)}}} M^{i_lj_{\sigma(l)}}.$$

The first equality follows from the block-constant structure of T and the second from the premise in this case that i_l and $j_{\sigma(l)}$ are in different blocks.

Consider now the last case when $i_{\nu} = j_{\sigma(\nu)}$ for all $\nu = 1, ..., k$. Using (59) one can see that the only summand that may have non-zero limit in the sum in the numerator of (63) is the one corresponding to the multi-index $(t_1, ..., t_k) = (v_1, ..., v_k)$. Thus, using the block-constant structure of T (recall that $i_{\nu} \in I_{\nu_{\nu}}$ for all $\nu = 1, ..., k$), (63) is equal to

$$\lim_{m \to \infty} \frac{1}{\|M_m\|} \left(T^{i_1 \dots i_k} (1 + o(\|M_m\|)) - T^{i_1 \dots i_k} \right) = 0$$

With that we finished calculating (63).

We now compute the right-hand side of (61) and compare with the results above. Suppose that $\sigma(l) = m$, then by the definition of $\sigma_{(l)}$ we have $\sigma_{(l)}^{-1}(m) = k + 1$, $\sigma_{(l)}^{-1}(k+1) = l$, and for any integer $i \in \mathbb{N}_{k+1} \setminus \{m, k+1\}$ we have $\sigma_{(l)}^{-1}(i) = \sigma^{-1}(i)$. Analogously, we have $\sigma_{(l)}(l) = k+1$, $\sigma_{(l)}(k+1) = \sigma(l)$, and for any integer $i \in \mathbb{N}_{k+1} \setminus \{l, k+1\}$ we have $\sigma_{(l)}(i) = \sigma(i)$.

Below we use the standard notation that a circumflex above a term in a product means that the term is omitted. Since $\sigma_{(l)}^{-1}(k+1) = l \neq k+1$ we use the second part of Lemma 2.7 to compute:

$$\sum_{l=1}^{k} \left(\text{Diag}^{\sigma_{(l)}} T_{\text{out}}^{(l)} \right) [H_{i_1 j_1}, \dots, H_{i_k j_k}, M_{\text{out}}] = \sum_{l=1}^{k} \langle T_{\text{out}}^{(l)}, H_{i_1 j_1} \circ_{\sigma_{(l)}} \cdots \circ_{\sigma_{(l)}} H_{i_k j_k} \circ_{\sigma_{(l)}} M_{\text{out}} \rangle$$
$$= \sum_{l=1}^{k} \left(T_{\text{out}}^{(l)} \right)^{i_1 \dots i_k j_{\sigma_{(l)}^{(k+1)}}} \left(\delta_{i_1 j_{\sigma_{(l)}^{(1)}}} \cdots \delta_{i_l j_{\sigma_{(l)}^{(l)}}} \right) M_{\text{out}}^{j_{\sigma_{(l)}^{(k+1)}} i_{\sigma_{(l)}^{-1}(k+1)}}$$

$$=\sum_{l=1}^{k} \left(T_{\text{out}}^{(l)}\right)^{i_1\dots i_k j_{\sigma(l)}} \left(\delta_{i_1 j_{\sigma_{(l)}(1)}}\cdots \delta_{i_l j_{\sigma_{(l)}(l)}}\cdots \delta_{i_k j_{\sigma_{(l)}(k)}}\right) M_{\text{out}}^{j_{\sigma(l)} i_l}$$
$$=\sum_{l=1}^{k} \left(T_{\text{out}}^{(l)}\right)^{i_1\dots i_k j_{\sigma(l)}} \left(\delta_{i_1 j_{\sigma(1)}}\cdots \widehat{\delta_{i_l j_{\sigma(l)}}}\cdots \delta_{i_k j_{\sigma(k)}}\right) M_{\text{out}}^{j_{\sigma(l)} i_l}.$$

The last equality holds because we changed the missing multiple, under the circumflex, (for each fixed l) while keeping the present multiples the same. It is clear now that if at least two of the pairs $(i_1, j_{\sigma(1)}), (i_2, j_{\sigma(2)}), ..., (i_k, j_{\sigma(k)})$ have different entries, then the last sum is zero. Let now exactly one of the pairs have unequal entries, say $i_l \neq j_{\sigma(l)}$, then the sum is equal to

(66)
$$(T_{\text{out}}^{(l)})^{i_1 \dots i_k j_{\sigma(l)}} (\delta_{i_1 j_{\sigma(1)}} \cdots \widehat{\delta_{i_l j_{\sigma(l)}}} \cdots \delta_{i_k j_{\sigma(k)}}) M_{\text{out}}^{j_{\sigma(l)} i_l}$$

If i_l and $j_{\sigma(l)}$ are in the same block, then $(T_{\text{out}}^{(l)})^{i_1 \dots i_k j_{\sigma(l)}} = 0$ by the definition of $T_{\text{out}}^{(l)}$. If i_l and $j_{\sigma(l)}$ are not in the same block, then (66) is equal to

$$(T_{\text{out}}^{(l)})^{i_1\dots i_k j_{\sigma(l)}} M_{\text{out}}^{j_{\sigma(l)} i_l} = \frac{T^{i_1\dots i_{l-1} i_l i_{l+1}\dots i_k} - T^{i_1\dots i_{l-1} j_{\sigma(l)} i_{l+1}\dots i_k}}{\mu_{i_l} - \mu_{j_{\sigma(l)}}} M_{\text{out}}^{i_l j_{\sigma(l)}},$$

because M is a symmetric matrix. Finally, if $i_{\nu} = j_{\sigma(\nu)}$ for all $\nu = 1, ..., k$, then again $(T_{\text{out}}^{(l)})^{i_1...i_k j_{\sigma(l)}} = 0$ for all l. These outcomes are equal to the results in the corresponding cases in the first part of the proof, the theorem follows.

Proposition 9.2 Let T be any k + 1-tensor on \mathbb{R}^n , let x be any vector in \mathbb{R}^n , let V be any $n \times n$ orthogonal matrix, and let σ be any permutation on \mathbb{N}_k . Then

$$V(\operatorname{Diag}^{\sigma}(T[x]))V^{T} = (V(\operatorname{Diag}^{\sigma_{(k+1)}}T)V^{T})[V(\operatorname{Diag} x)V^{T}]$$

Proof. Let $H_{i_1j_1}, \dots, H_{i_kj_k}$ be any k basic matrices. Recall that $\sigma_{(k+1)}(i) = \sigma(i)$ for all $i \in \mathbb{N}_k$ and $\sigma_{(k+1)}(k+1) = k+1$. Using Theorem 2.6 twice, we compute

$$\left(V \left(\text{Diag}^{\sigma}(T[x]) \right) V^{T} \right)^{i_{1} \dots i_{k}}_{j_{1} \dots j_{k}} = \left(V \left(\text{Diag}^{\sigma}(T[x]) \right) V^{T} \right) [H_{i_{1}j_{1}}, \dots, H_{i_{k}j_{k}}]$$

$$= \langle T[x], \tilde{H}_{i_{1}j_{1}} \circ_{\sigma} \dots \circ_{\sigma} \tilde{H}_{i_{k}j_{k}} \rangle$$

$$= \sum_{p_{1},\dots,p_{k}=1}^{n,\dots,n} (T[x])^{p_{1}\dots p_{k}} \tilde{H}_{i_{1}j_{1}}^{p_{1}p_{\sigma}-1} \cdots \tilde{H}_{i_{k}j_{k}}^{p_{k}p_{\sigma}-1} (k)$$

$$= \sum_{p_{1},\dots,p_{k},p_{k+1}=1}^{n,\dots,n} T^{p_{1}\dots p_{k+1}} x^{p_{k+1}} \tilde{H}_{i_{1}j_{1}}^{p_{1}p_{\sigma}-1} \cdots \tilde{H}_{i_{k}j_{k}}^{p_{k}p_{\sigma}-1} (k)$$

$$= \sum_{p_{1},\dots,p_{k},p_{k+1}=1}^{n,\dots,n} T^{p_{1}\dots p_{k+1}} \tilde{H}_{i_{1}j_{1}}^{p_{1}p_{\sigma}-1} \cdots \tilde{H}_{i_{k}j_{k}}^{p_{k}p_{\sigma}-1} (b)$$

$$(\text{Diag} x)^{p_{k+1}p_{\sigma}-1} (k+1)$$

$$= \langle T, \tilde{H}_{i_1 j_1} \circ_{\sigma_{(k+1)}} \dots \circ_{\sigma_{(k+1)}} \tilde{H}_{i_k j_k} \circ_{\sigma_{(k+1)}} \text{Diag } x \rangle$$

= $(V(\text{Diag }^{\sigma_{(k+1)}}T)V^T)[H_{i_1 j_1}, \dots, H_{i_k j_k}, V(\text{Diag } x)V^T]$
= $((V(\text{Diag }^{\sigma_{(k+1)}}T)V^T)[V(\text{Diag } x)V^T])^{i_1 \dots i_k}_{j_1 \dots j_k}.$

Since these equalities hold for all indexes $i_1,...,i_k$ and $j_1,...,j_k$ we are done.

The next lemma says that for any block-constant tensor T, Diag^{σ}T is invariant under conjugations with a block-diagonal orthogonal matrix.

Lemma 9.3 Let T be a block-constant k-tensor on \mathbb{R}^n and let $U \in O^n$ be a block-diagonal matrix (both with respect to the same partitioning of \mathbb{N}_n). Then for any permutation σ in \mathbb{N}_k

$$U(\operatorname{Diag}^{\sigma}T)U^T = \operatorname{Diag}^{\sigma}T.$$

Proof. Let $\{I_1, ..., I_r\}$ be the partitioning of the integers \mathbb{N}_n that determines the block structure. Notice that $U^{ip}U^{jp} = 0$ whenever $i \not\sim j$ or $i \not\sim p$, and that $\sum_{p \in I_s} U^{ip}U^{jp} = \delta_{ij}$ whenever $i \in I_s$. Let $(i_1, ..., i_k)$ be an arbitrary multi index and suppose that $i_l \in I_{v_l}$ for l = 1, ..., k. We expand the left-hand side of the identity:

$$\begin{pmatrix} U(\text{Diag}\,^{\sigma}T)U^{T} \end{pmatrix}^{i_{1}^{1}\dots i_{k}}_{j_{1}\dots j_{k}} = \sum_{\substack{p_{s},q_{s}=1\\s=1,\dots,k}}^{n,\dots,n} (\text{Diag}\,^{\sigma}T)^{p_{1}\dots p_{k}}_{q_{1}\dots q_{k}}U^{i_{1}p_{1}}U^{j_{1}q_{1}}\dots U^{i_{k}p_{k}}U^{j_{k}q_{k}} \\ = \sum_{p_{1},\dots,p_{k}=1}^{n,\dots,n} T^{p_{1}\dots p_{k}}U^{i_{1}p_{1}}U^{j_{1}p_{\sigma}-1}(1)\dots U^{i_{k}p_{k}}U^{j_{k}p_{\sigma}-1}(k) \\ = \sum_{p_{1},\dots,p_{k}=1}^{n,\dots,n} T^{p_{1}\dots p_{k}}U^{i_{1}p_{1}}U^{j_{\sigma}(1)p_{1}}\dots U^{i_{k}p_{k}}U^{j_{\sigma}(k)p_{k}} \\ = \sum_{p_{1},\dots,p_{k}=1}^{r,\dots,r} T^{v_{1}\dots v_{k}}\sum_{\substack{p_{l}\in I_{t_{l}}\\l=1,\dots,k}} U^{i_{1}p_{1}}U^{j_{\sigma}(1)p_{1}}\dots U^{i_{k}p_{k}}U^{j_{\sigma}(k)p_{k}} \\ = T^{v_{1}\dots v_{k}}\sum_{\substack{p_{l}\in I_{v_{l}}\\l=1,\dots,k}} U^{i_{1}p_{1}}U^{j_{\sigma}(1)p_{1}}\dots U^{i_{k}p_{k}}U^{j_{\sigma}(k)p_{k}} \\ = T^{v_{1}\dots v_{k}}\delta_{i_{1}j_{\sigma}(1)}\cdots \delta_{i_{k}j_{\sigma}(k)} \\ = T^{i_{1}\dots i_{k}}\delta_{i_{1}j_{\sigma}(1)}\cdots \delta_{i_{k}j_{\sigma}(k)} \\ = (\text{Diag}\,^{\sigma}T)^{i_{1}\dots i_{k}}. \end{cases}$$

The penultimate equality follows from the fact that T is block-constant.

Given a block structure on \mathbb{N}_n and any matrix M, by M_{in} we denote the matrix with the same diagonal blocks as M and the rest of the entries set to zero, as defined by (3).

Theorem 9.4 Let $U \in O^n$ be a block-diagonal orthogonal matrix. Let M be an arbitrary symmetric matrix and let $h \in \mathbb{R}^n$ be a vector such that

(67)
$$U^T M_{\rm in} U = {\rm Diag} \, h.$$

Let H_1, \ldots, H_k be arbitrary matrices and let σ be a permutation on \mathbb{N}_k . Then

(i) for any block-constant (k+1)-tensor T on \mathbb{R}^n ,

$$\langle T[h], \tilde{H}_1 \circ_{\sigma} \cdots \circ_{\sigma} \tilde{H}_k \rangle = \langle T, H_1 \circ_{\sigma_{(k+1)}} \cdots \circ_{\sigma_{(k+1)}} H_k \circ_{\sigma_{(k+1)}} M_{\mathrm{in}} \rangle$$

(ii) for any block-constant k-tensor T on \mathbb{R}^n

$$\langle T^{\tau_l}[h], \tilde{H}_1 \circ_{\sigma} \cdots \circ_{\sigma} \tilde{H}_k \rangle = \langle T_{\mathrm{in}}^{(l)}, H_1 \circ_{\sigma_{(l)}} \cdots \circ_{\sigma_{(l)}} H_k \circ_{\sigma_{(l)}} M_{\mathrm{in}} \rangle, \text{ for all } l = 1, \dots, k,$$

where the permutations $\sigma_{(l)}$ for l = 1, ..., k, k+1 are defined by (13), $\tilde{H}_i = U^T H_i U$ for i = 1, ..., k, and the lifting T^{τ_l} is defined by (17).

Proof. To see that the first identity holds we use Theorem 2.6, Proposition 9.2, (67), and Lemma 9.3 in that order, as follows:

$$\langle T[h], \tilde{H}_1 \circ_{\sigma} \cdots \circ_{\sigma} \tilde{H}_k \rangle = \left(U(\operatorname{Diag}{}^{\sigma}T[h])U^T \right) [H_1, ..., H_k]$$

$$= \left(U(\operatorname{Diag}{}^{\sigma_{(k+1)}}T)U^T \right) [H_1, ..., H_k, U(\operatorname{Diag}{h})U^T]$$

$$= \left(U(\operatorname{Diag}{}^{\sigma_{(k+1)}}T)U^T \right) [H_1, ..., H_k, M_{\operatorname{in}}]$$

$$= \left(\operatorname{Diag}{}^{\sigma_{(k+1)}}T \right) [H_1, ..., H_k, M_{\operatorname{in}}]$$

$$= \langle T, H_1 \circ_{\sigma_{(k+1)}} \cdots \circ_{\sigma_{(k+1)}} H_k \circ_{\sigma_{(k+1)}} M_{\operatorname{in}} \rangle.$$

The last equality follows again from Theorem 2.6.

To show the second identity, it suffices to prove it for arbitrary basic matrices $H_{i_s j_s}$ s = 1, ..., k. Fix k basic matrices $H_{i_1 j_1}, ..., H_{i_k j_k}$ and suppose that $i_l \in I_{v_l}$ for l = 1, ..., k. Then

$$\langle T^{\tau_{l}}[h], H_{i_{1}j_{1}} \circ_{\sigma} \cdots \circ_{\sigma} H_{i_{k}j_{k}} \rangle = \left(U(\text{Diag}\,^{\sigma}T^{\tau_{l}}[h])U^{T} \right) [H_{i_{1}j_{1}}, \dots, H_{i_{k}j_{k}}]$$

$$= \left(U(\text{Diag}\,^{\sigma}T^{\tau_{l}}[h])U^{T} \right)^{i_{1}\dots i_{k}}_{j_{1}\dots j_{k}}$$

$$= \sum_{\substack{p_{1},\dots,p_{k}=1\\q_{1},\dots,q_{k}=1}}^{n,\dots,n} (\text{Diag}\,^{\sigma}T^{\tau_{l}}[h])^{p_{1}\dots p_{k}}U^{i_{1}p_{1}}U^{j_{1}q_{1}}\dots U^{i_{k}p_{k}}U^{j_{k}q_{k}}$$

$$= \sum_{\substack{p_{1},\dots,p_{k}=1\\p_{1},\dots,p_{k}=1}}^{n,\dots,n} (T^{\tau_{l}}[h])^{p_{1}\dots p_{k}}U^{i_{1}p_{1}}U^{j_{1}p_{\sigma}-1}(1)}\dots U^{i_{k}p_{k}}U^{j_{k}p_{\sigma}-1}(k)$$

$$= \sum_{p_{1},\dots,p_{k}=1}^{n,\dots,n} (T^{\tau_{l}}[h])^{p_{1}\dots p_{k}}U^{i_{1}p_{1}}U^{j_{\sigma}(1)p_{1}}\dots U^{i_{k}p_{k}}U^{j_{\sigma}(k)p_{k}}$$

$$= \sum_{p_1,\dots,p_k=1}^{n,\dots,n} \sum_{p_{k+1}=1}^{n} (T^{\tau_l})^{p_1\dots p_k p_{k+1}} h^{p_{k+1}} U^{i_1 p_1} U^{j_{\sigma(1)} p_1} \cdots U^{i_k p_k} U^{j_{\sigma(k)} p_k}$$

$$= \sum_{p_1,\dots,p_k=1}^{n,\dots,n} T^{p_1\dots p_k} h^{p_l} U^{i_1 p_1} U^{j_{\sigma(1)} p_1} \cdots U^{i_k p_k} U^{j_{\sigma(k)} p_k}$$

$$= \sum_{t_1,\dots,t_k=1}^{r,\dots,r} T^{\iota_{t_1}\dots\iota_{t_k}} \sum_{\substack{p_\eta \in I_{t_\eta} \\ \eta=1,\dots,k}} h^{p_l} U^{i_1 p_1} U^{j_{\sigma(1)} p_1} \cdots U^{i_k p_k} U^{j_{\sigma(k)} p_k}$$

$$= T^{\iota_{v_1}\dots\iota_{v_k}} \sum_{\substack{p_l \in I_{v_l} \\ l=1,\dots,k}} h^{p_l} U^{i_1 p_1} U^{j_{\sigma(1)} p_1} \cdots U^{i_k p_k} U^{j_{\sigma(k)} p_k}$$

$$= T^{i_1\dots i_v} \delta_{i_1 j_{\sigma(1)}} \cdots \delta_{i_l j_{\sigma(l)}} \cdots \delta_{i_k j_{\sigma(k)}} \sum_{p_l \in I_{v_l}} h^{p_l} U^{i_1 p_l} U^{j_{\sigma(l)} p_l}$$

$$= T^{i_1\dots i_k} \delta_{i_1 j_{\sigma(1)}} \cdots \widehat{\delta_{i_l j_{\sigma(l)}}} \cdots \delta_{i_k j_{\sigma(k)}} M^{i_l j_{\sigma(l)}}_{i_1}.$$

To evaluate the right-hand side of the identity, we use the second part of Lemma 2.7 since $\sigma_{(l)}^{-1}(k+1) = l \neq k+1$. Recall also that $\sigma_{(l)}(s) = \sigma(s)$ for $s \in \mathbb{N}_{k+1} \setminus \{l, k+1\}$ and $\sigma_{(l)}(k+1) = \sigma(l)$ for all l = 1, ..., k.

$$\langle T_{\mathrm{in}}^{(l)}, H_{i_{1}j_{1}} \circ_{\sigma_{(l)}} \cdots \circ_{\sigma_{(l)}} H_{i_{k}j_{k}} \circ_{\sigma_{(l)}} M_{\mathrm{in}} \rangle$$

$$= (T_{\mathrm{in}}^{(l)})^{i_{1} \dots i_{k}j_{\sigma_{(l)}(k+1)}} \delta_{i_{1}j_{\sigma_{(l)}(1)}} \cdots \delta_{i_{l}j_{\sigma_{(l)}(l)}} \cdots \delta_{i_{k}j_{\sigma_{(l)}(k)}} M_{\mathrm{in}}^{j_{\sigma_{(l)}(k+1)}i_{\sigma_{(l)}^{-1}(k+1)}}$$

$$= (T_{\mathrm{in}}^{(l)})^{i_{1} \dots i_{k}j_{\sigma_{(l)}(l)}} \delta_{i_{1}j_{\sigma_{(1)}}} \cdots \delta_{i_{k}j_{\sigma_{(l)}(l)}} M_{\mathrm{in}}^{j_{\sigma_{(l)}(l)}i_{l}}$$

$$= T^{i_{1} \dots i_{k}} \delta_{i_{1}j_{\sigma_{(1)}}} \cdots \delta_{i_{k}j_{\sigma_{(k)}}} M_{\mathrm{in}}^{j_{\sigma_{(l)}(l)}i_{l}}$$

$$= T^{i_{1} \dots i_{k}} \delta_{i_{1}j_{\sigma_{(1)}}} \cdots \delta_{i_{k}j_{\sigma_{(k)}}} M_{\mathrm{in}}^{i_{1}j_{\sigma_{(l)}}}$$

In the third equality above we used the fact that T is block-constant, plus the fact that $M_{in}^{j_{\sigma(l)}i_l} = 0$ if $j_{\sigma(l)} \not\sim i_l$. In the fourth we used the fact that M is a symmetric matrix. The last equality holds because we changed the missing multiple, while keeping the present multiples the same.

Proposition 9.5 Let $U \in O^n$ be a block-diagonal orthogonal matrix, let H be an $n \times n$ matrix, and let σ be an arbitrary permutation on \mathbb{N}_k .

(i) If T is a (k+1)-tensor such that for some fixed $l \in \mathbb{N}_k$ we have $T^{p_1 \dots p_l \dots p_{k+1}} = 0$ whenever $p_l \sim p_{k+1}$, then

$$\left(U(\operatorname{Diag}^{o_{(l)}}T)U^{T}\right)[H_{\operatorname{in}}]=0$$

(ii) If T is a (k+1)-tensor such that for some fixed $l \in \mathbb{N}_k$ we have $T^{p_1 \dots p_l \dots p_{k+1}} = 0$ whenever $p_l \not\sim p_{k+1}$, then

$$(U(\operatorname{Diag}^{\sigma_{(l)}}T)U^T)[H_{\operatorname{out}}] = 0$$

(iii) If T is any (k+1)-tensor, then

$$(U(\operatorname{Diag}^{\sigma_{(k+1)}}T)U^T)[H_{\operatorname{out}}] = 0.$$

Proof. Fix an index l in \mathbb{N}_k . Let $H_{i_1j_1}, \dots, H_{i_kj_k}$ be arbitrary basic matrices, and let H be an arbitrary matrix. Using the definitions we compute:

$$\left(U(\operatorname{Diag}^{\sigma_{(l)}}T)U^{T} \right) [H_{i_{1}j_{1}}, \dots, H_{i_{k}j_{k}}, H] = \sum_{i_{k+1}, j_{k+1}=1}^{n, n} \left(U(\operatorname{Diag}^{\sigma_{(l)}}T)U^{T} \right)^{i_{1}...i_{k+1}} H^{i_{k+1}j_{k+1}}$$

$$= \sum_{i_{k+1}, j_{k+1}=1}^{n, n} \sum_{\substack{p_{s}, q_{s}=1\\s=1, \dots, k+1}}^{n, \dots, n} (\operatorname{Diag}^{\sigma_{(l)}}T)^{p_{1}...p_{k+1}} U^{i_{1}p_{1}} U^{j_{1}q_{1}} \dots U^{i_{k+1}p_{k+1}} U^{j_{k+1}q_{k+1}} H^{i_{k+1}j_{k+1}}$$

$$= \sum_{i_{k+1}, j_{k+1}=1}^{n, n} \sum_{\substack{s=1, \dots, k+1\\s=1, \dots, k+1}}^{n, \dots, n} T^{p_{1}...p_{k+1}} U^{i_{1}p_{1}} U^{j_{1}p_{\sigma_{(l)}^{-1}(1)}} \dots U^{i_{k+1}p_{k+1}} U^{j_{k+1}p_{\sigma_{(l)}^{-1}(k+1)}} H^{i_{k+1}j_{k+1}}$$

$$= \sum_{i_{k+1}, j_{k+1}=1}^{n, n} \sum_{\substack{s=1, \dots, k+1\\s=1, \dots, k+1}}^{n, \dots, n} T^{p_{1}...p_{k+1}} U^{i_{1}p_{1}} U^{j_{\sigma_{(l)}(1)}p_{1}} \dots U^{i_{l}p_{l}} U^{j_{\sigma_{(l)}(1)}p_{l}} \dots U^{i_{k+1}p_{k+1}} U^{j_{\sigma_{(l)}(k+1)}p_{k+1}} H^{i_{k+1}j_{k+1}}$$

$$= \sum_{i_{k+1}, j_{k+1}=1}^{n, n} \sum_{\substack{s=1, \dots, k+1\\s=1, \dots, k+1}}^{n, \dots, n} T^{p_{1}...p_{k+1}} U^{i_{1}p_{1}} U^{j_{\sigma_{(1)}(1)}p_{1}} \dots U^{i_{l}p_{l}} U^{j_{\sigma_{(l)}(1)}p_{l}} \dots U^{i_{k+1}p_{k+1}} U^{j_{\sigma_{(l)}(p_{k+1})}p_{k+1}} H^{i_{k+1}j_{k+1}}$$

Suppose now that T is a (k+1)-tensor with $T^{p_1...p_{l}...p_{k+1}} = 0$ whenever $p_l \sim p_{k+1}$ and that $H = H_{\text{in}}$. Then $H^{i_{k+1}j_{k+1}} \neq 0$ implies that $i_{k+1} \sim j_{k+1}$. In that case, by the fact that U is block-diagonal, $U^{j_{k+1}p_l}U^{i_{k+1}p_{k+1}} \neq 0$ implies that $p_l \sim p_{k+1}$, which implies that $T^{p_1...p_{l}...p_{k+1}} = 0$. Thus every summand in the double sum above is zero.

In the second case, suppose T is a (k + 1)-tensor with $T^{p_1 \dots p_l \dots p_{k+1}} = 0$ whenever $p_l \not\sim p_{k+1}$ and $H = H_{\text{out}}$. Then $H^{i_{k+1}j_{k+1}} \neq 0$ implies that $i_{k+1} \not\sim j_{k+1}$. In that case, by the fact that U is block-diagonal, $U^{j_{k+1}p_l}U^{i_{k+1}p_{k+1}} \neq 0$ implies that $p_l \not\sim p_{k+1}$, which implies that $T^{p_1 \dots p_l \dots p_{k+1}} = 0$. The sum is zero.

In the third case, suppose that T is any (k + 1)-tensor and $H = H_{out}$. A calculation almost identical to the one at the beginning of the proof (it differs only in the last step) shows that

$$(U(\operatorname{Diag}^{\sigma_{(k+1)}}T)U^{T})[H_{i_{1}j_{1}}, ..., H_{i_{k}j_{k}}, H] = \sum_{\substack{n,n\\i_{k+1}, j_{k+1}=1}}^{n,n} \sum_{\substack{p_{s}=1\\s=1,...,k+1}}^{n,...,n} T^{p_{1}...p_{k+1}}U^{i_{1}p_{1}}U^{j_{\sigma(1)}p_{1}} \cdots U^{i_{k}p_{k}}U^{j_{\sigma(k)}p_{k}}U^{i_{k+1}p_{k+1}}U^{j_{k+1}p_{k+1}}H^{i_{k+1}j_{k+1}}.$$

Then $H^{i_{k+1}j_{k+1}} \neq 0$ implies that $i_{k+1} \not \sim j_{k+1}$. In that case, by the fact that U is block-diagonal, $U^{j_{k+1}p_{k+1}}U^{i_{k+1}p_{k+1}} = 0$. Again the sum is zero.

We are finally ready to conclude the proofs of our main two analytical tools.

Proof of Theorem 2.9. A consequence of Theorem 9.1 and Proposition 9.5.

Proof of Theorem 2.10. A consequence of Theorem 2.6, Theorem 9.4, Proposition 9.5, and the fact that $M = M_{\text{in}} + M_{\text{out}}$.

If vector μ , defining the equivalence relation on \mathbb{N}_n , has distinct coordinates, then every tensor from $T^{k,n}$ is block-constant and the block-diagonal orthogonal matrices are precisely the signed identity matrices (those with plus or minus one on the main diagonal and zeros everywhere else). In this case we also have $i \sim j$ if and only if i = j and thus $T_{in}^{(l)} = T^{\tau_l}$. Moreover, since Proposition 9.5 holds for arbitrary matrices (symmetric or not), Theorem 2.10 becomes the next result, valid for an arbitrary matrix H.

Corollary 9.6 Let σ be a permutation on \mathbb{N}_k and let H be an arbitrary matrix. Then

(i) for any (k+1)-tensor T on \mathbb{R}^n ,

 $\operatorname{Diag}^{\sigma}(T[\operatorname{diag} H]) = (\operatorname{Diag}^{\sigma_{(k+1)}}T)[H];$

(ii) for any k-tensor T on \mathbb{R}^n

 $\operatorname{Diag}^{\sigma}(T^{\tau_l}[\operatorname{diag} H]) = (\operatorname{Diag}^{\sigma_{(l)}}T^{\tau_l})[H], \quad for \ all \ l = 1, \dots, k,$

where the permutations $\sigma_{(l)}$, for $l \in \mathbb{N}_k$, are defined by (13).

10 Appendix C: Proof of Theorem 6.4

Let $X \in S^n$ be a symmetric matrix with distinct eigenvalues, and let $x = \lambda(X)$. The proof of Theorem 6.4 is by induction on s. When s = 1 there is nothing to show since by definition $\tilde{\mathcal{A}}_{(1)}(x) = \nabla f(x) = \mathcal{A}_{(1)}(x)$ for every $x \in \mathbb{R}^n$. Suppose that for some integer s in [1, k) we have

$$\sum_{\sigma \in P^s} \operatorname{Diag}{}^{\sigma} \tilde{\mathcal{A}}_{\sigma}(x) = \sum_{\sigma \in P^s} \operatorname{Diag}{}^{\sigma} \mathcal{A}_{\sigma}(x),$$

for every $x \in \mathbb{R}^n$ with distinct coordinates.

Recall that by definition the tensor $\mathcal{A}_{\sigma}(x)$ is equal to zero if the permutation σ has more than one cycle in its cycle decomposition. Then using Lemma 6.3 we get

$$\sum_{\sigma \in P^{s+1}} \operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(x) = \sum_{\substack{\sigma \in P^{s} \\ l \in \mathbb{N}_{s+1}}} \operatorname{Diag}^{\sigma_{(l)}} \mathcal{A}_{\sigma_{(l)}}(x)$$
$$= \sum_{\substack{\sigma \in P^{s} \\ l \in \mathbb{N}_{s}}} \operatorname{Diag}^{\sigma_{(l)}} \mathcal{A}_{\sigma_{(l)}}(x)$$
$$= \sum_{\substack{\sigma \in P^{s} \\ l \in \mathbb{N}_{s}}} \operatorname{Diag}^{\sigma_{(l)}} \left((\mathcal{A}_{\sigma}(x))_{\operatorname{out}}^{(l)} + (T_{\sigma}^{l}(x))_{\operatorname{in}}^{(l)} \right)$$

Let M be an arbitrary symmetric matrix with norm one. Let $\{M_m\}_{m=1}^{\infty}$ be a sequence of symmetric matrices converging to zero and such that $M_m/||M_m||$ converges to M. Finally, let $\{U_m\}_{m=1}^{\infty}$ be a sequence of orthogonal matrices such that

$$\operatorname{Diag} x + M_m = U_m (\operatorname{Diag} \lambda (\operatorname{Diag} x + M_m)) U_m^T.$$

By taking a subsequence if necessary, we may assume that U_m converges to $U \in O^n$ when m goes to infinity. Since the decomposition of the integers \mathbb{N}_n into blocks is determined by the repeated eigenvalues of the matrix X, and the later are all distinct, we have $M_{\text{in}} = \text{Diag}(\text{diag } M)$. (Moreover, every tensor is block-constant.) Thus defining vector $h \in \mathbb{R}^n$ as in (23) we see that h = diag Mand by (25) we have $U^T M_{\text{in}} U = \text{Diag}(\text{diag } M)$. Reversing the steps leading to (42) the induction hypothesis, and then using the first part of Theorem 2.10, on the one hand we get

$$\begin{split} \Big(\sum_{\substack{\sigma \in P^s \\ l \in \mathbb{N}_s}} \operatorname{Diag}^{\sigma_{(l)}}(T^l_{\sigma}(x))^{(l)}_{\mathrm{in}}\Big)[M] &= U\Big(\sum_{\sigma \in P^s} \operatorname{Diag}^{\sigma}\left(\nabla \mathcal{A}_{\sigma}(x)[h]\right)\Big)U^T \\ &= \lim_{t \to 0} U\Big(\sum_{\sigma \in P^s} \operatorname{Diag}^{\sigma}\left(\mathcal{A}_{\sigma}(x+th) - \mathcal{A}_{\sigma}(x)\right)\Big)U^T \\ &= \lim_{t \to 0} U\Big(\sum_{\sigma \in P^s} \operatorname{Diag}^{\sigma}\left(\nabla \tilde{\mathcal{A}}_{\sigma}(x)[h]\right)\Big)U^T \\ &= U\Big(\sum_{\sigma \in P^s} \operatorname{Diag}^{\sigma}(\nabla \tilde{\mathcal{A}}_{\sigma}(x)[h])\Big)U^T \\ &= \Big(\sum_{\sigma \in P^s} \operatorname{Diag}^{\sigma_{(s+1)}} \nabla \tilde{\mathcal{A}}_{\sigma}(x)\Big)[M] \\ &= \Big(\sum_{\sigma \in P^s} \operatorname{Diag}^{\sigma_{(s+1)}} \tilde{\mathcal{A}}_{\sigma_{(s+1)}}(x)\Big)[M]. \end{split}$$

In the last equality we used the second line from (28). On the other hand, using (15), the induction hypothesis, and again (15) we have

$$\Big(\sum_{\substack{\sigma \in P^s \\ l \in \mathbb{N}_s}} \operatorname{Diag}^{\sigma_{(l)}}(\mathcal{A}_{\sigma}(x))_{\operatorname{out}}^{(l)}\Big)[M] = \lim_{m \to \infty} \frac{U_m\Big(\sum_{\sigma \in P^s} \operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(x)\Big)U_m^T - \sum_{\sigma \in P^s} \operatorname{Diag}^{\sigma} \mathcal{A}_{\sigma}(x)}{\|M_m\|}$$

$$= \lim_{m \to \infty} \frac{U_m \left(\sum_{\sigma \in P^s} \operatorname{Diag}^{\sigma} \tilde{\mathcal{A}}_{\sigma}(x)\right) U_m^T - \sum_{\sigma \in P^s} \operatorname{Diag}^{\sigma} \tilde{\mathcal{A}}_{\sigma}(x)}{\|M_m\|}$$
$$= \left(\sum_{\substack{\sigma \in P^s\\l \in \mathbb{N}_s}} \operatorname{Diag}^{\sigma_{(l)}} \left(\tilde{\mathcal{A}}_{\sigma}(x)\right)_{\operatorname{out}}^{(l)}\right) [M]$$
$$= \left(\sum_{\substack{\sigma \in P^s\\l \in \mathbb{N}_s}} \operatorname{Diag}^{\sigma_{(l)}} \tilde{\mathcal{A}}_{\sigma_{(l)}}(x)\right) [M].$$

In the last equality we used the first line in (28). Thus we see that

$$\Big(\sum_{\sigma\in P^{s+1}}\operatorname{Diag}^{\sigma}\mathcal{A}_{\sigma}(x)\Big)[M] = \Big(\sum_{\sigma\in P^{s+1}}\operatorname{Diag}^{\sigma}\tilde{\mathcal{A}}_{\sigma}(x)\Big)[M],$$

and since M was arbitrary, we are done.

Acknowledgements: I thank Jérôme Malick for his interest in this work, and informing me about several typos in the earlier versions of the manuscript. I would like to especially thank the first referee for his thorough reading, invaluable help, and guidance in bringing up the manuscript to its current state. Finally, I would like to thank Heinz Bauschke for his support and encouragement.

References

- V.I. ARNOLD, On matrices depending on parameters, Russian Mathematical Surveys, 26 (1971) 29–43.
- [2] R. BHATIA, Matrix Analysis, Springer-Verlag, NY, 1997.
- [3] X. CHEN, H.-D. QI and P. TSENG, Analysis of nonsmooth symmetric matrix functions with applications to semidifinite programming, *SIAM Journal on Optimization*, 13 (2003) 960–985.
- [4] YU.L. DALECKII, Integration and differentiation of functions of Hermitian operators depending on a parameter, American Mathematical Society Translations, Series 2, 16 (1960) 396–400.
- [5] YU.L. DALECKII and S.G. KREIN, Formulas of differentiation according to a parameter of functions of Hermitian operators, *Doklady Akademii Nauk SSSR (N.S.)*, 76 (1951) 13–16.
- [6] J.-B. HIRIART-URRUTY and D. YE, Sensitivity analysis of all eigenvalues of a symmetric matrix, *Numerische Mathematik*, 70 (1992) 45–72.
- [7] R.A. HORN and C.R. JOHNSON, *Topics in Matrix Analysis*, Cambridge University Press, First paperback edition with corrections, 1994.

- [8] T. KATO, Perturbation Theory for Linear Operators, second ed., Grundlehren der Mathematischen Wissenschaften, Band 132, Springer-Verlag New York, Inc., Berlin, 1976.
- [9] E.C. KEMBLE, The Fundamental Principles of Quantum Mechanics, Dover, New York, 1958.
- [10] F. KRAUS, Uber Konvexe Matrixfunctionen, Mathematische Zeitschrift, 41:1 (1936) 18-42.
- [11] A.S. LEWIS, Convex analysis on the Hermitian matrices, SIAM Journal on Optimization, 6 (1996) 164–177.
- [12] A.S. LEWIS, Derivatives of spectral functions, Mathematics of Operations Research, 21 (1996) 576–588.
- [13] A.S. LEWIS, Nonsmooth analysis of eigenvalues, Mathematical Programming, 84 (1999) 1–24.
- [14] A.S. LEWIS, The mathematics of eigenvalue optimization, Mathematical Programming, Series B, 97:1–2 (2003) 155–176.
- [15] A.S. LEWIS and M.L. OVERTON, Eigenvalue optimization, Acta numerica, 5 (1996) 149–190.
- [16] A.S. LEWIS and H.S. SENDOV, Twice differentiable spectral functions, SIAM Journal of Matrix Analysis and Applications, 23:2 (2001) 368–386.
- [17] A.S. LEWIS and H.S. SENDOV, Quadratic expansions of spectral functions, *Linear Algebra and Its Applications*, 340 (2002) 97–121.
- [18] K. LOWNER, Uber Monotone Matrixfunctionen, Mathematische Zeitschrift, 38:1 (1934) 177-216.
- [19] J. MALICK and H.S. SENDOV, Clarke Generalized Jacobian of the Projection onto the Cone of Positive Semidefinite Matrices, *Set-Valued Analysis*, accepted.
- [20] R. MATHIAS, A chain rule for matrix functions and applications, SIAM Journal on Matrix Analysis and Applications, 17:3 (1996) 610–620.
- [21] H.S. SENDOV, Generalized Hadamard product and the derivatives of spectral functions, *SIAM Journal on Matrix Analysis and Applications*, to appear.
- [22] D. SUN and J. SUN, Semismooth matrix-valued functions, Mathematics of Operations Research, 27:1 (2002) 150–169.
- [23] N.-K. TSING, M.K.H. FAN, and E.I. VERRIEST, On analyticity of functions involving eigenvalues, *Linear Algebra and its Applications*, 207 (1994) 159–180.