Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Linear Algebra and its Applications 424 (2007) 240–281
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

The higher-order derivatives of spectral functions ☆

Hristo S. Sendov

Department of Mathematics and Statistics, 50 Stone Road East, University of Guelph, Guelph, Ont., Canada N1G 2W1

Received 14 July 2006; accepted 20 December 2006
Available online 19 January 2007
Submitted by R.A. Brualdi

Dedicated to Professor Roger Horn on the occasion of his 65th birthday

Abstract

We are interested in higher-order derivatives of functions of the eigenvalues of real symmetric matrices with respect to the matrix argument. We describe a formula for the kth 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.

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

AMS classification: 49R50; 47A75

Keywords: Spectral function; Differentiable; Twice differentiable; Higher-order derivative; Eigenvalue optimization; Symmetric function; Perturbation theory; Tensor analysis; Hadamard product

☆ Research supported by NSERC.
E-mail address: hssendov@uoguelph.ca

0024-3795/$ - see front matter
doi:10.1016/j.laa.2006.12.013

1. Introduction

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

$$F(UXU^{\mathrm{T}}) = F(X)$$

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 similar. The restriction of F to the subspace of diagonal matrices defines a function $f(x) = F(\mathrm{Diag}\, x)$ on a vector argument $x \in \mathbb{R}^n$. The function $f : \mathbb{R}^n \to \mathbb{R}$ is *symmetric*, that is, has the property

$$f(x) = f(Px) \quad \text{for any permutation matrix } P \text{ and any } x \text{ in the domain of } f,$$

and in addition, $F(X) = (f \circ \lambda)(X)$, in which the *eigenvalue map*

$$\lambda(X) = (\lambda_1(X), \ldots, \lambda_n(X))$$

is the vector of eigenvalues of X arranged in non-increasing order.

What smoothness properties of the symmetric function f are inherited by F? The eigenvalue map $\lambda(X)$ is continuous but not always differentiable with respect to X. Even 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 [13] there is an explicit formulae for the gradient of the spectral function F in terms of the derivatives of the symmetric function f:

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

where V is any orthogonal matrix such that $X = V(\mathrm{Diag}\, \lambda(X))V^{\mathrm{T}}$ is the *ordered* spectral decomposition of X. In [17] 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 [13,17] by proving, in two general cases, the following formula for the kth derivative of a spectral function

$$\nabla^k(f \circ \lambda)(X) = V\left(\sum_{\sigma \in P^k} \mathrm{Diag}^{\sigma} \mathscr{A}_{\sigma}(\lambda(X))\right)V^{\mathrm{T}}, \tag{2}$$

where again $X = V(\mathrm{Diag}\, \lambda(X))V^{\mathrm{T}}$. The sum is taken over all permutations on k elements, which are a convenient tool for enumerating the maps $\mathscr{A}_{\sigma}(x)$. 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 $\mathscr{A}_{\sigma}(x)$ depend only on the partial derivatives of $f(x)$ up to order k, and do not depend on the eigenvalues; they reveal how the higher-order derivatives depend on the eigenvalue map $\lambda(X)$. Formula (2) depends on the eigenvalues only through the compositions $\mathscr{A}_{\sigma}(\lambda(X))$ and conjugation by the orthogonal matrix V.

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

Our results for separable spectral functions imply those of [5,4] for one-parameter families of symmetric matrices; see also the monographs [9,10]. More precisely, when restricted to the space of real symmetric matrices, the Daleckiĭ–Kreĭn formulae describe the kth order derivative of the function $t \to F(X(t))$, where $X(t)$ is a k-times continuously differentiable curve of symmetric matrices and $F = f \circ \lambda$ is a separable spectral function with f being k-times continuously differentiable. We describe the higher-order derivatives of $X \to F(X)$ from which one can obtain the derivatives of $t \to F(X(t))$ by applying the chain rule.

Our results also capture and extend those in [21] when specialized to symmetric matrices. (The gradients of separable spectral functions are the functions considered in [21] when restricted to the space of symmetric matrices.) For example, Theorem 4.1 in [21] shows that if the separable function f is k-times continuously differentiable and $t \in \mathbb{R} \mapsto X(t)$ is a k-times differentiable curve of symmetric matrices, then $F(X(t))$ is k-times differentiable, where $F = f \circ \lambda$. Thus, Theorem 4.1 in [21] strengthens the Daleckiĭ–Kreĭn result by dispensing with the requirement that $\frac{\partial^k}{\partial t^k} X(t)$ be continuous. In Theorem 6.1 we assume only that f is k-times differentiable to obtain that $F(X)$ is k-times differentiable with respect to the symmetric matrix variable X. In that case, one can again use the chain rule to obtain the derivatives of $F(X(t))$. In addition, Theorem 6.9 shows that if f is k-times continuously differentiable then $F(X)$ is also k-times continuously differentiable with respect to the variable X.

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 $\mathscr{A}_{\sigma_1}(x) = \mathscr{A}_{\sigma_2}(x)$ and these maps allow a simple determinant description. If σ has more than one cycle, then $\mathscr{A}_{\sigma}(x) \equiv 0$.

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 [17].

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. 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 $\langle X, Y\rangle = \mathrm{tr}(XY^{\mathrm{T}})$ and corresponding norm $\|X\| = \sqrt{\langle X, X\rangle}$. For $A \in S^n$, $\lambda(A) = (\lambda_1(A), \dots, \lambda_n(A))$ is the vector of its eigenvalues arranged in non-increasing order:

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

By $\mathbb{N}_k$ we denote the set $\{1, 2, \dots, k\}$. For any vector x in $\mathbb{R}^n$, $\mathrm{Diag}x$ denotes the diagonal matrix with the entries of vector x on the main diagonal, and $\mathrm{diag}\colon M^n \to \mathbb{R}^n$ denotes its adjoint operator, defined by $\mathrm{diag}(X) = (x_{11}, \dots, x_{nn})$. By $\mathbb{R}^n_{\downarrow}$ we denote the cone of all vectors x in $\mathbb{R}^n$ such that $x_1 \geqslant x_2 \geqslant \cdots \geqslant x_n$. Denote the standard orthonormal basis in $\mathbb{R}^n$ by $e^1, e^2, \dots, e^n$. 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 $Ph = (h_{\sigma(1)}, \dots, h_{\sigma(n)})^{\mathrm{T}}$, that is, $P^{\mathrm{T}}e^i = e^{\sigma(i)}$ for all $i = 1, \dots, n$. 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-indices $(i_1, \dots, i_k)$ and $(j_1, \dots, j_k)$ are called *equivalent* if $i_l \sim j_l$ for all $l = 1, 2, \dots, k$, and we write $(i_1, \dots, i_k) \sim (j_1, \dots, j_k)$ (or $(i_1, \dots, i_k) \sim_\mu (j_1, \dots, j_k)$).

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 $(h_1, \dots, h_k)$ is denoted by $T[h_1, \dots, h_k]$. For any $(i_1, \dots, i_k)$, a k-tuple of integers from $\mathbb{N}_n$, we denote by $T^{i_1 \cdots i_k}$ the value $T[e^{i_1}, \dots, e^{i_k}]$. 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 following lemma motivates the following definitions. It is an 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 that is 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) $\nabla f(\mu) = P^{\mathrm{T}} \nabla f(\mu)$,
(ii) $\nabla^2 f(\mu) = P^{\mathrm{T}} \nabla^2 f(\mu) P$, *and*
(iii) $\nabla^s f(\mu)[h_1, \dots, h_s] = \nabla^s f(\mu)[Ph_1, \dots, Ph_s]$, *for any $h_1, \dots, h_s \in \mathbb{R}^n$, and $s \in \mathbb{N}_k$.*

Definition 2.2. A tensor $T \in T^{k,n}$ is called *symmetric* if

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

for any permutation σ on $\mathbb{N}_k$ and any $h_1, \dots, h_k \in \mathbb{R}^n$.

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

$$T[Ph_1, \dots, Ph_k] = T[h_1, \dots, h_k]$$

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

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

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

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

If the map $\mu \in \mathbb{R}^n \to \mathscr{F}(\mu) \in T^{k,n}$ is point symmetric then the tensor $\mathscr{F}(\mu)$ 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 $T^{i_1 \cdots i_k} = T^{j_1 \cdots j_k}$ whenever $(i_1, \dots, i_k) \sim_\mu (j_1, \dots, j_k)$.

(ii) A k-tensor-valued map $\mu \in \mathbb{R}^n \to \mathscr{F}(\mu) \in T^{k,n}$ is *block constant* if $\mathscr{F}(\mu)$ is block constant with respect to μ for every $\mu \in \mathbb{R}^n$.

Every tensor that is block constant with respect to μ 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 it exists) $\mu \in \mathbb{R}^n \to \nabla^s f(\mu)$ is a point-symmetric, s-tensor-valued map. In addition, if the mapping $\mu \in \mathbb{R}^n \to \nabla^s f(\mu)$ is continuous, then the tensor $\nabla^s f(\mu)$ is symmetric.

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

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 formula for the first-order Taylor expansion. Let vectors $h_1, \dots, h_k, h$ be given and let $\{v_m\}$ be any sequence of vectors in $\mathbb{R}^n$ approaching zero such that $v_m/\|v_m\|$ approaches h as $m \to \infty$

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

On the other hand, for any permutation P we have

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

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

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

The result follows. $\square$

For any given fixed vector $\mu \in \mathbb{R}^n$ we define a linear operation on matrices: $M \in M^n \to M_{\mathrm{in}} \in M^n$, as follows

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

and

$$M_{\mathrm{out}} = M - M_{\mathrm{in}}. \tag{4}$$

Even though in M_{in} and M_{out} we omit the dependence on μ, no confusion will arise since the μ will be clear from the context.

2.1. Generalized Hadamard product

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 $\{H_{pq} \in M^n | H_{pq}^{ij} = \delta_{ip}\delta_{jq} \quad \text{for all } i, j \in \mathbb{N}_n\}$.

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

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

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

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

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 of two matrices are

$$H_1 \circ_{(1)(2)} H_2 = (\operatorname{diag} H_1)(\operatorname{diag} H_2)^{\mathrm{T}},$$
$$H_1 \circ_{(12)} H_2 = H_1 \circ H_2^{\mathrm{T}}.$$

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

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

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

$$T[H_1, \ldots, H_k] := \sum_{p_1, q_1 = 1}^{n} \cdots \sum_{p_k, q_k = 1}^{n} T^{\;p_1 \cdots p_k}_{\;q_1 \cdots q_k} H_1^{p_1 q_1} \cdots H_k^{p_k q_k}. \tag{7}$$

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

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

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

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

the corresponding norm is $\|T\| = \sqrt{\langle T, T \rangle}$. 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 any $U \in O^n$ this action is denoted by $UTU^{\mathrm{T}} \in T^{k,n}$:

$$(UTU^{\mathrm{T}})^{i_1 \ldots i_k} = \sum_{p_1=1}^{n} \cdots \sum_{p_k=1}^{n} (T^{p_1 \cdots p_k} U^{i_1 p_1} \cdots U^{i_k p_k}). \tag{9}$$

This action is norm preserving and associative, that is,

$$\|VXV^{\mathrm{T}}\| = \|X\| \quad \text{and} \quad V(UTU^{\mathrm{T}})V^{\mathrm{T}} = (VU)T(VU)^{\mathrm{T}}$$

for all $U, V \in O^n$; see [22, Lemma 4.1].

The $\operatorname{Diag}^\sigma$ operator, the σ-Hadamard product, and conjugation by an orthogonal matrix are connected by the following multi-linear duality relation; see [22, Theorem 4.3].

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*

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

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

We also need the following two lemmas from [22].

Lemma 2.7. *Let T be a k-tensor on $\mathbb{R}^n$, and H be a matrix in M^n. Let $H_{i_1 j_1}, \dots, H_{i_{k-1} j_{k-1}}$ be basic matrices in M^n, and let σ be a permutation on $\mathbb{N}_k$.*

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

$$\langle T, H_{i_1 j_1} \circ_\sigma \cdots \circ_\sigma H_{i_{k-1} j_{k-1}} \circ_\sigma H \rangle = \left(\prod_{t=1}^{k-1} \delta_{i_t j_{\sigma(t)}} \right) \sum_{t=1}^{n} T^{i_1 \dots i_{k-1} t} H^{tt}.$$

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

$$\langle T, H_{i_1 j_1} \circ_\sigma \cdots \circ_\sigma H_{i_{k-1} j_{k-1}} \circ_\sigma H \rangle = \left(\prod_{\substack{t=1 \\ t \neq l}}^{k-1} \delta_{i_t j_{\sigma(t)}} \right) T^{i_1 \dots i_{k-1} j_{\sigma(k)}} H^{j_{\sigma(k)} i_{\sigma^{-1}(k)}}.$$

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

$$V(T[V^{\mathrm{T}} H V])V^{\mathrm{T}} = (VTV^{\mathrm{T}})[H].$$

2.2. *Operations with tensors*

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

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

as follows:

$$(T^{(l)}_{\mathrm{out}})^{i_1 \dots i_k i_{k+1}} = \begin{cases} 0, & \text{if } i_l \sim_\mu i_{k+1}, \\ \dfrac{T^{i_1 \cdots i_{l-1} i_{k+1} i_{l+1} \cdots i_k} - T^{i_1 \cdots i_{l-1} i_l i_{l+1} \cdots i_k}}{\mu_{i_{k+1}} - \mu_{i_l}}, & \text{if } i_l \nsim_\mu i_{k+1}. \end{cases} \tag{11}$$

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

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

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.

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

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

Given the cycle decomposition of σ, we obtain $\sigma_{(l)}$ for each $l = 1, \dots, k$ by inserting the element $k+1$ immediately after the element l; when $l = k+1$, we obtain the permutation $\sigma_{(k+1)}$ by appending the one-element cycle $(k+1)$ to σ. Notice that $\sigma_{(l)}^{-1}(k+1) = l$ for all l, and that the map

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

is one-to-one and onto.

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

Theorem 2.9. *Let M be a given real symmetric matrix and let $\{M_m\}_{m=1}^{\infty}$ be a sequence of real 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*

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

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

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

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:

$$(T_{\mathrm{in}}^{(l)})^{i_1\ldots i_k i_{k+1}} = \begin{cases} T^{i_1\ldots i_{l-1} i_{k+1} i_{l+1}\ldots i_k}, & \text{if } i_l \sim_{\mu} i_{k+1}, \\ 0, & \text{if } i_l \nsim_{\mu} i_{k+1}. \end{cases} \tag{16}$$

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

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

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:

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

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

When μ has distinct entries, then $i_l \sim_{\mu} i_{k+1}$ if and only if $i_l = i_{k+1}$ and therefore $T_{\mathrm{in}}^{(l)} = T^{\tau_l}$ for every $l \in \mathbb{N}_k$.

The next theorem is the second and final calculus-like 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 in S^n, and let $h \in \mathbb{R}^n$ be a vector such that $U^{\mathrm{T}} M_{\mathrm{in}} U = \operatorname{Diag} h$. Then*

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

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

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

$$U(\mathrm{Diag}^{\sigma}(T^{\tau_l}[h]))U^{\mathrm{T}} = (\mathrm{Diag}^{\sigma_{(1)}} T^{(l)}_{\mathrm{in}})[M] \quad \textit{for all } l = 1, \ldots, k,$$

where the permutations $\sigma_{(1)}$, 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 $\mathscr{A}_\sigma : \mathbb{R}^n \to T^{s,n}$, depending only on the function f and its partial derivatives, such that

$$\nabla^s(f \circ \lambda)(X) = V\left(\sum_{\sigma \in P^s} \mathrm{Diag}^{\sigma} \mathscr{A}_\sigma(\lambda(X))\right) V^{\mathrm{T}}, \tag{18}$$

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

By [22, Section 5] it is enough to prove (18) only when X is an ordered diagonal matrix. That is, $X = \mathrm{Diag}\,\mu$ for some vector $\mu \in \mathbb{R}^n_{\downarrow}$.

That (18) holds when $s = 1$ was shown in [13], see also Subsection 5.2.

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

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

for $s = 1, \ldots, k-1$, 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 real symmetric matrices converging to 0 with $M_m/\|M_m\|$ converging to $M \in S^n$ and show inductively that

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

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

$$\mathrm{Diag}\,\mu + M_m = U_m(\mathrm{Diag}\,\lambda(\mathrm{Diag}\,\mu + M_m))U_m^{\mathrm{T}} \quad \text{for all } m = 1, 2, \ldots \tag{20}$$

The next lemma combines [14, Lemma 5.10] and [7, Theorem 3.12].

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

$$\lambda(\mathrm{Diag}\,\mu + M_m)^{\mathrm{T}} = \mu^{\mathrm{T}} + (\lambda(X_1^{\mathrm{T}} M_m X_1)^{\mathrm{T}}, \ldots, \lambda(X_r^{\mathrm{T}} M_m X_r)^{\mathrm{T}})^{\mathrm{T}} + \mathrm{o}(\|M_m\|), \tag{21}$$

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

We denote

$$h_m := (\lambda(X_1^{\mathrm{T}} M_m X_1)^{\mathrm{T}}, \ldots, \lambda(X_r^{\mathrm{T}} M_m X_r)^{\mathrm{T}})^{\mathrm{T}}. \tag{22}$$

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

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

Throughout the paper, we reserve the symbols h_m and h to denote the vectors in (22) and (23). With this notation Lemma 3.1 says that

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

Taking the limit in (20) as m goes to infinity Theorem A.1 ensures that U is block diagonal with respect to μ and

$$U^{\mathrm{T}} M_{\mathrm{in}} U = \mathrm{Diag}\, h, \tag{25}$$

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

4. Analyticity of isolated eigenvalues

Let A be in S^n and suppose that the jth 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. A function of several real variables is *analytic* at a point if in a neighbourhood of this point it can be expressed as power series. The corresponding complex variable notion is *holomorphic*.

Our first justification is from [24, 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

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

Notice that 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 $x \in R^n$ and every permutation matrix P. Theorem 4.1 ensures that $\lambda_j = f \circ \lambda$ is an analytic function.

Our second justification uses the following result from [1]. (In the theorem below, $\lambda_i(X)$ denotes an arbitrary eigenvalue of a matrix X, not necessarily the ith largest one.)

Theorem 4.2 (Arnold 1971). *Suppose $A \in \mathbb{C}^{n\times n}$ has q eigenvalues $\lambda_1(A), \ldots, \lambda_q(A)$ (counting multiplicities) in an open set $\Omega \subset \mathbb{C}$, and suppose 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 $X \in \Delta$*

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

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

To deduce the result we need, since the jth 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 $n-1$ eigenvalues are not in the closure of Ω. Theorem 4.2 ensures that, there is a neighbourhood Δ of A and a holomorphic mapping $S : \Delta \to \mathbb{C}$ such that $S(X)$ is equal to the jth 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. The restriction $\tilde{S}(X)$ of $S(X)$ to $\Delta \cap S^n$ is a holomorphic real-valued function. Therefore, the coefficients in the power series expansion of $\tilde{S}(X)$ must be real numbers. Thus, the jth 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 kth 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 kth 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 a dense open set in $\mathbb{R}^n$, and the set of symmetric matrices with distinct eigenvalues is a dense open set in S^n. (For a convex analysis proof of the last fact, see [20, Corollary 1.6].)

One can obtain the kth 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:

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

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

$$\begin{aligned}\nabla^3 F(x)[H_1, H_2, H_3] = {} & \nabla^3 f(\lambda(x))[\nabla\lambda(x)[H_1], \nabla\lambda(x)[H_2], \nabla\lambda(x)[H_3]] \\ & + \nabla^2 f(\lambda(x))[\nabla\lambda(x)[H_1], \nabla^2\lambda(x)[H_2, H_3]] \\ & + \nabla^2 f(\lambda(x))[\nabla\lambda(x)[H_2], \nabla^2\lambda(x)[H_1, H_3]] \\ & + \nabla^2 f(\lambda(x))[\nabla\lambda(x)[H_3], \nabla^2\lambda(x)[H_1, H_2]] \\ & + \nabla f(\lambda(x))[\nabla^3\lambda(x)[H_1, H_2, H_3]].\end{aligned}$$

This approach to the kth derivative requires every derivative of λ up to the kth. 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 kth derivative of $f \circ \lambda$ that does not 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 entries. Thus, 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 $X_l = [e^l]$, $l = 1, \ldots, n$. This implies that $h_m = \operatorname{diag} M_m$ and that $h = \operatorname{diag} M$. The definition of $T^{(l)}_{\mathrm{out}}$ given in (11) is now:

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

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

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

$$\tilde{\mathscr{A}}_{(1)}(x) := \nabla f(x).$$

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

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

We are now ready to formulate our first main theorem.

Theorem 5.1. *Let X be a symmetric matrix with distinct eigenvalues. Let f be a function defined on a neighbourhood of $\lambda(X)$. 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,*

$$\nabla^k F(X) = V \left(\sum_{\sigma \in P^k} \operatorname{Diag}^\sigma \tilde{\mathscr{A}}_\sigma(\lambda(X)) \right) V^{\mathrm{T}}, \tag{29}$$

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

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

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

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

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

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

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

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

In the last equality we used Lemma 2.1(i), U is block diagonal and orthogonal, and f is symmetric, so $\nabla f(\mu)$ is block constant.

5.3. *Proof of Theorem 5.1: the induction step*

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

$$\nabla^s(f\circ\lambda)(X)=V\left(\sum_{\sigma\in P^s}\mathrm{Diag}^{\sigma}\tilde{\mathscr{A}}_{\sigma}(\lambda(X))\right)V^{\mathrm{T}},$$

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

Using (24), we differentiate $\nabla^s(f\circ\lambda)$ at $\mathrm{Diag}\,\mu$:

$$\begin{aligned}
&\nabla^{s+1}(f\circ\lambda)(\mathrm{Diag}\,\mu)[M]\\
&\quad=\lim_{m\to\infty}\frac{\nabla^s(f\circ\lambda)(\mathrm{Diag}\,\mu+M_m)-\nabla^s(f\circ\lambda)(\mathrm{Diag}\,\mu)}{\|M_m\|}\\
&\quad=\lim_{m\to\infty}\frac{U_m\left(\sum_{\sigma\in P^s}\mathrm{Diag}^{\sigma}\tilde{\mathscr{A}}_{\sigma}(\lambda(\mathrm{Diag}\,\mu+M_m))\right)U_m^{\mathrm{T}}-\sum_{\sigma\in P^s}\mathrm{Diag}^{\sigma}\tilde{\mathscr{A}}_{\sigma}(\mu)}{\|M_m\|}\\
&\quad=\lim_{m\to\infty}\sum_{\sigma\in P^s}\frac{U_m(\mathrm{Diag}^{\sigma}\tilde{\mathscr{A}}_{\sigma}(\lambda(\mathrm{Diag}\,\mu+M_m)))U_m^{\mathrm{T}}-\mathrm{Diag}^{\sigma}\tilde{\mathscr{A}}_{\sigma}(\mu)}{\|M_m\|}\\
&\quad=\lim_{m\to\infty}\sum_{\sigma\in P^s}\frac{U_m(\mathrm{Diag}^{\sigma}\tilde{\mathscr{A}}_{\sigma}(\mu+h_m+\mathrm{o}(\|M_m\|)))U_m^{\mathrm{T}}-\mathrm{Diag}^{\sigma}\tilde{\mathscr{A}}_{\sigma}(\mu)}{\|M_m\|}\\
&\quad=\lim_{m\to\infty}\sum_{\sigma\in P^s}\frac{U_m(\mathrm{Diag}^{\sigma}(\tilde{\mathscr{A}}_{\sigma}(\mu)+\nabla\tilde{\mathscr{A}}_{\sigma}(\mu)[h_m]+\mathrm{o}(\|M_m\|)))U_m^{\mathrm{T}}-\mathrm{Diag}^{\sigma}\tilde{\mathscr{A}}_{\sigma}(\mu)}{\|M_m\|}\\
&\quad=\lim_{m\to\infty}\sum_{\sigma\in P^s}\frac{U_m(\mathrm{Diag}^{\sigma}\tilde{\mathscr{A}}_{\sigma}(\mu))U_m^{\mathrm{T}}-\mathrm{Diag}^{\sigma}\tilde{\mathscr{A}}_{\sigma}(\mu)}{\|M_m\|}\\
&\qquad+\sum_{\sigma\in P^s}U(\mathrm{Diag}^{\sigma}(\nabla\tilde{\mathscr{A}}_{\sigma}(\mu)[h]))U^{\mathrm{T}}.
\end{aligned}$$

For every $\sigma \in P^s$ the tensor $\tilde{\mathscr{A}}_\sigma(\mu)$ is block constant, so Theorem 2.9 ensures that

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

For every $\sigma \in P^s$ the gradient $\nabla\tilde{\mathscr{A}}_\sigma(\mu)$ is a block-constant $(s+1)$-tensor, so Theorem 2.10 implies that

$$U(\mathrm{Diag}^\sigma(\nabla\tilde{\mathscr{A}}_\sigma(\mu)[h]))U^{\mathrm{T}} = (\mathrm{Diag}^{\sigma^{(s+1)}}\nabla\tilde{\mathscr{A}}_\sigma(\mu))[M] = (\mathrm{Diag}^{\sigma^{(s+1)}}\tilde{\mathscr{A}}_{\sigma^{(s+1)}}(\mu))[M].$$

We conclude that

$$\nabla^{s+1}(f\circ\lambda)(\mathrm{Diag}\,\mu)[M] = \left(\sum_{\substack{\sigma\in P^s\\ l\in\mathbb{N}_{s+1}}} \mathrm{Diag}^{\sigma^{(l)}}\tilde{\mathscr{A}}_{\sigma^{(l)}}(\mu)\right)[M]$$

for every symmetric matrix M. Because (14) is a one-to-one and onto map, we see that

$$\nabla^{s+1}(f\circ\lambda)(X) = V\left(\sum_{\sigma\in P^{s+1}} \mathrm{Diag}^\sigma\tilde{\mathscr{A}}_\sigma(\lambda(X))\right)V^{\mathrm{T}},$$

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

Finally, we show that the $(s+1)$-tensor-valued maps $\tilde{\mathscr{A}}_{\sigma^{(l)}}(\cdot)$ are at least $(k-s-1)$-times (continuously) differentiable. This is clear when $l = s+1$ and $\sigma \in P^s$, since $\tilde{\mathscr{A}}_\sigma(\cdot)$ is $(k-s)$-times (continuously) differentiable for every $\sigma \in P^s$. For the rest of the maps, every entry in $\tilde{\mathscr{A}}_{\sigma^{(l)}}$ is the difference of two entries of $\tilde{\mathscr{A}}_\sigma$ divided by a quantity that never becomes zero over the set Ω. This shows that $\tilde{\mathscr{A}}_{\sigma^{(l)}}(\cdot)$ is $(k-s)$-times (continuously) differentiable on the set Ω for every $\sigma \in P^s$ and every $l \in \mathbb{N}_s$.

This concludes the proof of Theorem 5.1.

6. The kth 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.

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

$$f(x_1,\ldots,x_n) = g(x_1)+\cdots+g(x_n) \tag{31}$$

is the separable spectral function

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

Choose an orthogonal matrix V such that $X = V(\mathrm{Diag}\,\lambda(X))V^{\mathrm{T}}$. It follows from (1) that if g is differentiable at the points $\{\lambda_i(X)|i\in\mathbb{N}_n\}$ then F is differentiable at X and

$$\nabla F(X) = V(\mathrm{Diag}\,(g'(\lambda_1(X)),\ldots,g'(\lambda_n(X))))V^{\mathrm{T}}. \tag{33}$$

Separable spectral functions and their derivatives are of great importance for modern optimization; see [3,12,23]. For the role of general spectral functions see the two survey papers [15,16].

Löner studied the class of matrix-valued functions (33) in [19], where he established a connection between monotonicity of the map (33) with respect to the semidefinite order and differentiability propertied of g'. Löwner's student Kraus, determined conditions on g' that make the map (33) convex with respect to the semidifinite order [11]. For more information, related, and recent results see [2, Chapter V], [8, Section 6.6] and the monograph [6]. The matrix-valued map (33) is the *primary matrix function* $g'(X)$; see [8, Chapter 6] for a general discussion and the Daleckiĭ–Kreĭn formulae for the kth derivative of a primary matrix function along a one dimensional curve. The first two derivatives of (33) can be found in [2, Chapter V].

6.1. Description of the kth 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

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

The integral representation $g^{[(12)]}(x, y) = \int_0^1 g''(y + t(x - y))\,\mathrm{d}t$ 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, so $|\tilde{P}^s| = (s - 1)!$ For every $\sigma \in \tilde{P}^s$ and every $l \in \mathbb{N}_s$ we have $\sigma_{(1)} \in \tilde{P}^{s+1}$. Moreover, as σ varies over $\tilde{P}^s$ and l varies over $\mathbb{N}_s$, the permutation $\sigma_{(1)}$ varies over $\tilde{P}^{s+1}$ in a one-to-one and onto fashion.

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

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

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

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

for every $l \in \mathbb{N}_s$, we see that $g^{[\sigma_{(1)}]}(x_1, \ldots, x_{s+1})$ 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 $\{g^{[\sigma]}(x_1, \ldots, x_k) | \sigma \in \tilde{P}^k\}$.

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

$$\mathscr{A}_\sigma : \mathbb{R}^n \to T^{s,n}, \quad \text{by}$$
$$(\mathscr{A}_\sigma(x))^{i_1 \ldots i_s} := g^{[\sigma]}(x_{i_1}, \ldots, x_{i_s}). \tag{36}$$

If $(i_1,\dots,i_s) \sim_x (j_1,\dots,j_s)$, then $(\mathscr{A}_\sigma(x))^{i_1\dots i_s} = (\mathscr{A}_\sigma(x))^{j_1\dots j_s}$, 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 our second main theorem.

Theorem 6.1. *Let g be a k-times differentiable real-valued function defined on a real interval I. Let $X \in S^n$ have eigenvalues in I, and let V be an orthogonal matrix such that $X = V(\operatorname{Diag}\lambda(X))V^{\mathrm{T}}$. Then the separable spectral function F defined by (31) and (32) is k-times differentiable at X, and its kth derivative is*

$$\nabla^k F(X) = V\left(\sum_{\sigma\in P^k} \operatorname{Diag}^\sigma \mathscr{A}_\sigma(\lambda(X))\right)V^{\mathrm{T}}, \tag{37}$$

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

The proof is given in the next subsection. We proceed by induction—consecutively differentiating $F(X)$. In the base case $k = 1$, (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 sth derivative ($1 \leqslant s < k$) of F at the matrix X is given by

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

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

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

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

$$\lim_{m\to\infty} \sum_{\sigma\in\tilde{P}^s} \frac{U_m(\mathrm{Diag}^{\sigma}\mathscr{A}_\sigma(\mu))U_m^{\mathrm{T}} - \mathrm{Diag}^{\sigma}\mathscr{A}_\sigma(\mu)}{\|M_m\|}$$
$$= \sum_{\substack{\sigma\in\tilde{P}^s\\ l\in\mathbb{N}_s}} (\mathrm{Diag}^{\sigma(1)}(\mathscr{A}_\sigma(\mu))^{(l)}_{\mathrm{out}})[M]. \tag{38}$$

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

$$\nabla[(\mathscr{A}_\sigma(\mu))^{i_1\ldots i_s}] = \sum_{l=1}^{s} \nabla_l g^{[\sigma]}(\mu_{i_1},\ldots,\mu_{i_s})e^{i_l} = \sum_{l=1}^{s} g^{[\sigma(1)]}(\mu_{i_1},\ldots,\mu_{i_s},\mu_{i_l})e^{i_l}, \tag{39}$$

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

$$(T^l_\sigma(\mu))^{i_1\ldots i_s} := g^{[\sigma(1)]}(\mu_{i_1},\ldots,\mu_{i_s},\mu_{i_l}). \tag{40}$$

Each of these maps is block constant.

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

$$\nabla\mathscr{A}_\sigma(\mu) = \sum_{l=1}^{s} (T^l_\sigma(\mu))^{\tau_l}, \tag{41}$$

where the "lifting" $(T^l_\sigma(\mu))^{\tau_l}$ is defined by (17).

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

$$\nabla[(\mathscr{A}_\sigma(\mu))^{i_1\ldots i_s}] = ((\nabla\mathscr{A}_\sigma(\mu))^{i_1\ldots i_s,1}, (\nabla\mathscr{A}_\sigma(\mu))^{i_1\ldots i_s,2},\ldots,(\nabla\mathscr{A}_\sigma(\mu))^{i_1\ldots i_s,n})^{\mathrm{T}}.$$

We compute the pth entry in the last vector. Using (39), we get

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

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

$$\begin{aligned}\left(\sum_{l=1}^{s}(T^l_\sigma(\mu))^{\tau_l}\right)^{i_1\ldots i_s,p} &= \sum_{l=1}^{s}((T^l_\sigma(\mu))^{\tau_l})^{i_1\ldots i_s,p}\\ &= \sum_{l=1}^{s}(T^l_\sigma(\mu))^{i_1\ldots i_s}\delta_{i_l p}\\ &= \sum_{\substack{l=1\\ i_l=p}}^{s}(T^l_\sigma(\mu))^{i_1\ldots i_s}\\ &= \sum_{\substack{l=1\\ i_l=p}}^{s} g^{[\sigma(1)]}(\mu_{i_1},\ldots,\mu_{i_s},\mu_{i_l}).\end{aligned}$$

The lemma follows. □

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

$$\begin{aligned}
U\left(\sum_{\sigma\in\tilde{P}^s}\operatorname{Diag}^{\sigma}(\nabla\mathscr{A}_\sigma(\mu)[h])\right)U^{\mathrm{T}} &= U\left(\sum_{\sigma\in\tilde{P}^s}\operatorname{Diag}^{\sigma}\left(\left(\sum_{l=1}^{s}(T_\sigma^l(\mu))^{\tau_l}\right)[h]\right)\right)U^{\mathrm{T}}\\
&= U\left(\sum_{\sigma\in\tilde{P}^s}\sum_{l=1}^{s}\operatorname{Diag}^{\sigma}((T_\sigma^l(\mu))^{\tau_l}[h])\right)U^{\mathrm{T}}\\
&= \sum_{\substack{\sigma\in\tilde{P}^s\\ l\in\mathbb{N}_s}} U(\operatorname{Diag}^{\sigma}((T_\sigma^l(\mu))^{\tau_l}[h]))U^{\mathrm{T}}\\
&= \sum_{\substack{\sigma\in\tilde{P}^s\\ l\in\mathbb{N}_s}} (\operatorname{Diag}^{\sigma(1)}(T_\sigma^l(\mu))^{(l)}_{\mathrm{in}})[M]; \qquad (42)
\end{aligned}$$

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

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

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

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

This shows that $\nabla^s F(\operatorname{Diag}\mu)$ is differentiable. We show now that $\nabla^{(s+1)}F(\operatorname{Diag}\mu)$ 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*

$$\mathscr{A}_{\sigma(1)}(\mu) = (T_\sigma^l(\mu))^{(l)}_{\mathrm{in}} + (\mathscr{A}_\sigma(\mu))^{(l)}_{\mathrm{out}}. \qquad (43)$$

Proof. Fix an $l \in \mathbb{N}_s$ and a multi index $(i_1, \dots, i_s, i_{s+1})$. We consider two cases depending on whether $\mu_{i_{s+1}}$ equals μ_{i_l} or not.

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 $(i_1, \dots, i_s, i_{s+1})$ is

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

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

$$\begin{aligned}
((T_\sigma^l(\mu))^{(l)}_{\mathrm{in}} + (\mathscr{A}_\sigma(\mu))^{(l)}_{\mathrm{out}})^{i_1\dots i_s i_{s+1}} &= ((T_\sigma^l(\mu))^{(l)}_{\mathrm{in}})^{i_1\dots i_s i_{s+1}} + ((\mathscr{A}_\sigma(\mu))^{(l)}_{\mathrm{out}})^{i_1\dots i_s i_{s+1}}\\
&= ((T_\sigma^l(\mu))^{(l)}_{\mathrm{in}})^{i_1\dots i_s i_{s+1}} + 0 = (T_\sigma^l(\mu))^{i_1\dots i_s}\\
&= g^{[\sigma(1)]}(\mu_{i_1}, \dots, \mu_{i_s}, \mu_{i_l}) = \nabla_l g^{[\sigma]}(\mu_{i_1}, \dots, \mu_{i_s});
\end{aligned}$$

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

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

$$
\begin{aligned}
(\mathscr{A}_{\sigma_{(1)}}(\mu))^{i_1\ldots i_s i_{s+1}} &= g^{[\sigma_{(1)}]}(\mu_{i_1},\ldots,\mu_{i_s},\mu_{i_{s+1}})\\
&= \frac{g^{[\sigma]}(\mu_{i_1},\ldots,\mu_{i_l},\ldots,\mu_{i_s}) - g^{[\sigma]}(\mu_{i_1},\ldots,\mu_{i_{s+1}},\ldots,\mu_{i_s})}{\mu_{i_l}-\mu_{i_{s+1}}},
\end{aligned}
$$

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

$$
\begin{aligned}
&((T^l_\sigma(\mu))^{(l)}_{\text{in}} + (\mathscr{A}_\sigma(\mu))^{(l)}_{\text{out}})^{i_1\ldots i_s i_{s+1}}\\
&\quad = ((T^l_\sigma(\mu))^{(l)}_{\text{in}})^{i_1\ldots i_s i_{s+1}} + ((\mathscr{A}_\sigma(\mu))^{(l)}_{\text{out}})^{i_1\ldots i_s i_{s+1}}\\
&\quad = 0 + ((\mathscr{A}_\sigma(\mu))^{(l)}_{\text{out}})^{i_1\ldots i_s i_{s+1}}\\
&\quad = \frac{(\mathscr{A}_\sigma(\mu))^{i_1\ldots i_{l-1} i_{s+1} i_{l+1}\ldots i_s} - (\mathscr{A}_\sigma(\mu))^{i_1\ldots i_{l-1} i_l i_{l+1}\ldots i_s}}{\mu_{i_{s+1}}-\mu_{i_l}}\\
&\quad = \frac{g^{[\sigma]}(\mu_{i_1},\ldots,\mu_{i_{s+1}},\ldots,\mu_{i_s}) - g^{[\sigma]}(\mu_{i_1},\ldots,\mu_{i_l},\ldots,\mu_{i_s})}{\mu_{i_{s+1}}-\mu_{i_l}}.
\end{aligned}
$$

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 must be reconciled in their common case. This is done by the following theorem proved in Appendix C.

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

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

where $x = \lambda(X)$.

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

Corollary 6.5. *Let g be twice differentiable in I, let $X \in S^n$ have all eigenvalues in I, and suppose that $X = V(\operatorname{Diag}\lambda(X))V^{\mathrm{T}}$ for some orthogonal matrix V. Then*

$$\nabla^2 F(X) = V(\operatorname{Diag}^{(12)}\mathscr{A}_{(12)}(\lambda(X)))V^{\mathrm{T}}, \tag{44}$$

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

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

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

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

where '$\circ$' denotes the usual Hadamard product. We now show that (44) is the same as (45).

Proposition 6.6. *For any $n \times n$ matrix A, any orthogonal V, and any symmetric H_1 and H_2,*

$$(V(\mathrm{Diag}^{(12)} A)V^{\mathrm{T}})[H_1, H_2] = \langle V(A \circ (V^{\mathrm{T}} H_1 V))V^{\mathrm{T}}, H_2\rangle,$$

where '$\circ$' 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

$$V(\mathrm{Diag}^{(12)} A)V^{\mathrm{T}}[H_1, H_2] = \langle A, \tilde{H}_1 \circ_{(12)} \tilde{H}_2\rangle.$$

On the other hand

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

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

6.3. C^k separable spectral functions

Theorem 6.1 holds for every k-times differentiable function g. In this section, we explain why (37) can be significantly simplified if g is k-times continuously differentiable. In particular, we show three properties of the functions $g^{[\sigma]}(x_1, \ldots, x_s)$. First, we express $g^{[\sigma]}(x_1, \ldots, x_s)$ as a ratio of two determinants whenever $x_1,\ldots,x_s$ are distinct. Second, the determinant formula ensures that $g^{[\sigma]}(x_1, \ldots, x_s)$ is a symmetric function of its arguments. Finally, we show that $g^{[\sigma_1]}(x_1, \ldots, x_s) = g^{[\sigma_2]}(x_1, \ldots, x_s)$ for all σ_1 and σ_2 in $\tilde{P}^s$. Thus, all tensors $\{\mathscr{A}_\sigma(x) | \sigma \in \tilde{P}^k\}$ in (37) are equal to each other, but are lifted onto different k-dimensional "diagonal planes" in the $2k$-dimensional tensor.

The Vandermonde determinant

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

has a variant

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

for any $y \in \mathbb{R}^s$; when $s = 1$, we set $V(x_1) = 1$ and $V\begin{pmatrix} y_1 \\ x_1 \end{pmatrix} = y_1$.

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

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

Proof. When $s=1$ we have

$$\frac{V\begin{pmatrix}y_1\\x_1\end{pmatrix}}{V(x_1)}-\frac{V\begin{pmatrix}y_2\\x_2\end{pmatrix}}{V(x_2)}=(x_1-x_2)\frac{V\begin{pmatrix}y_1,y_2\\x_1,x_2\end{pmatrix}}{V(x_1,x_2)}.$$

For the rest of the proof we assume $s\geqslant 2$. Consider both sides of (46) as a multivariate polynomial (of degree one) in the variables $y_1,\dots,y_s,y_{s+1}$. We show that y_k has equal coefficients on both sides for all $k\in\mathbb{N}_{s+1}$. First observe that

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

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

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

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

Now suppose $k=l$. Then the coefficient of y_k on the left-hand side of (46) is

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

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

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

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

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

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

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

which is again the coefficient of y_{s+1} on the right-hand side. □

Theorem 6.8. *Suppose $g \in C^k(I)$. Then for every permutation $\sigma \in \tilde{P}^s$, where $1 \leqslant s \leqslant k$, and every vector $(x_1, \dots, x_s)$ with distinct entries*

$$g^{[\sigma]}(x_1, \dots, x_s) = \frac{V\begin{pmatrix} g'(x_1), \dots, g'(x_s) \\ x_1, \dots, x_s \end{pmatrix}}{V(x_1, \dots, x_s)}. \tag{47}$$

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

Proof. The proof is by induction on s. When $s = 1$, the definitions ensure that

$$g^{[(1)]}(x_1) = g'(x_1) = \frac{V\begin{pmatrix} g'(x_1) \\ x_1 \end{pmatrix}}{V(x_1)}.$$

Suppose (47) holds for s, where $1 \leqslant s < k$. Let $(x_1, \dots, x_s, x_{s+1})$ be a vector with distinct entries and let $y = (g'(x_1), \dots, g'(x_s), g'(x_{s+1}))$. 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

$$\begin{aligned}
&g^{[\sigma_{(l)}]}(x_1, \dots, x_s, x_{s+1}) \\
&\quad = \frac{g^{[\sigma]}(x_1, \dots, x_s) - g^{[\sigma]}(x_1, \dots, x_{l-1}, x_{s+1}, x_{l+1}, \dots, x_s)}{x_l - x_{s+1}} \\
&\quad = \frac{1}{(x_l - x_{s+1})}\left(\frac{V\begin{pmatrix} y_1, \dots, y_s \\ x_1, \dots, x_s \end{pmatrix}}{V(x_1, \dots, x_s)} - \frac{V\begin{pmatrix} y_1, \dots, y_{l-1}, y_{s+1}, y_{l+1}, \dots, y_s \\ x_1, \dots, x_{l-1}, x_{s+1}, x_{l+1}, \dots, x_s \end{pmatrix}}{V(x_1, \dots, x_{l-1}, x_{s+1}, x_{l+1}, \dots, x_s)} \right) \\
&\quad = \frac{V\begin{pmatrix} y_1, \dots, y_l, y_{s+1}, y_{l+1}, \dots, y_s \\ x_1, \dots, x_l, x_{s+1}, x_{l+1}, \dots, x_s \end{pmatrix}}{V(x_1, \dots, x_l, x_{s+1}, x_{l+1}, \dots, x_s)} \\
&\quad = \frac{V\begin{pmatrix} y_1, \dots, y_{s+1} \\ x_1, \dots, x_{s+1} \end{pmatrix}}{V(x_1, \dots, x_{s+1})}.
\end{aligned}$$

Since $\tilde{P}^{s+1} = \{\sigma_{(l)} | \sigma \in \tilde{P}^s, l \in \mathbb{N}_s\}$ the induction step is completed. Finally, since $g^{[\sigma]}(x_1, \dots, x_s)$ 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 $\mathscr{A} : \mathbb{R}^n \to T^{k,n}$ by

$$(\mathscr{A}(x))^{i_1 \dots i_k} := \frac{V\begin{pmatrix} g'(x_{i_1}), \dots, g'(x_{i_k}) \\ x_{i_1}, \dots, x_{i_k} \end{pmatrix}}{V(x_{i_1}, \dots, x_{i_k})}. \tag{48}$$

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

Theorem 6.9. *Let g be a C^k function defined on an interval I. Let $X \in S^n$ have eigenvalues in the interval I, and let $V \in O^n$ be such that $X = V(\operatorname{Diag} \lambda(X))V^{\mathrm{T}}$. Then the separable spectral function F defined by (31) and (32) is k-times continuously differentiable at X, and its kth derivative is*

$$\nabla^k F(X) = V\left(\sum_{\sigma \in \tilde{P}^k} \operatorname{Diag}^{\sigma} \mathscr{A}(\lambda(X))\right)V^{\mathrm{T}}, \tag{49}$$

where $\mathscr{A}(x)$ 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 the result is 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 representations of the formula for the kth derivative at X of a C^k separable spectral function, applied at k symmetric matrices.

6.3.1. Derivatives as multi linear operators

The next corollary specializes Theorem 6.9 to the case $k = 3$. It should be compared with [2, (V.22)] and [8, §6.6]. One should keep in mind that we are differentiating separable spectral functions, whose gradients are the primary matrix functions considered in [2, Chapter V].

Corollary 6.10. *For $g \in C^3(I)$ and any H_1, H_2, H_3 in S^n we have*

$$\nabla^3 F(X)[H_1, H_2, H_3] = 2 \sum_{p_1,p_2,p_3=1}^{n,n,n} \mathscr{A}(\lambda(X))^{p_1p_2p_3} \tilde{H}_1^{p_1p_2} \tilde{H}_2^{p_2p_3} \tilde{H}_3^{p_3p_1},$$

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

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

$$\begin{aligned}
&\nabla^3 F(\operatorname{Diag}\mu)[H_1, H_2, H_3] \\
&\quad = \left(\sum_{\sigma \in \tilde{P}^3} \operatorname{Diag}^{\sigma} \mathscr{A}(\mu) \right) [H_1, H_2, H_3] \\
&\quad = \sum_{\sigma \in \tilde{P}^3} \langle \mathscr{A}(\mu), H_1 \circ_{\sigma} H_2 \circ_{\sigma} H_3 \rangle \\
&\quad = \langle \mathscr{A}(\mu), H_1 \circ_{(123)} H_2 \circ_{(123)} H_3 \rangle + \langle \mathscr{A}(\mu), H_1 \circ_{(132)} H_2 \circ_{(132)} H_3 \rangle \\
&\quad = \sum_{q_1,q_2,q_3=1}^{n,n,n} \mathscr{A}(\mu)^{q_1q_2q_3} H_1^{q_1q_3} H_2^{q_2q_1} H_3^{q_3q_2} + \sum_{p_1,p_2,p_3=1}^{n,n,n} \mathscr{A}(\mu)^{p_1p_2p_3} H_1^{p_1p_2} H_2^{p_2p_3} H_3^{p_3p_1}.
\end{aligned}$$

After re-parametrization of the first sum ($q_1 = p_2, q_2 = p_3, q_3 = p_1$), using the symmetry of the tensor $\mathscr{A}(\mu)$ and the matrices H_1, H_2, H_3, we continue

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

which is what we wanted to show. □

In the general case when $H_1,\ldots,H_k$ are distinct symmetric matrices, we cannot simplify the formula for $\nabla^k F(X)[H_1, \ldots, H_k]$ 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 $\mathscr{A}$ 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

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

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

$$\begin{aligned}
&\sum_{p_1,\ldots,p_k=1}^{n,\ldots,n} \mathscr{A}^{p_{\sigma(1)}\ldots p_{\sigma(k)}} H_1^{p_{\sigma(1)} p_1} \cdots H_k^{p_{\sigma(k)} p_k} \\
&\quad= \sum_{p_1,\ldots,p_k=1}^{n,\ldots,n} \mathscr{A}^{p_1\ldots p_k} H_1^{p_{\sigma(1)} p_1} \cdots H_k^{p_{\sigma(k)} p_k} \\
&\quad= \sum_{p_1,\ldots,p_k=1}^{n,\ldots,n} \mathscr{A}^{p_1\ldots p_k} H_1^{p_1 p_{\theta^{-1}(1)}} \cdots H_k^{p_k p_{\theta^{-1}(k)}}.
\end{aligned}$$

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

We summarize the preceding paragraph in the following theorem.

Theorem 6.11. *Let $\tilde{P}_0^k$ be a subset of $\tilde{P}^k$, $k \geqslant 3$, 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 $H_1, \ldots, H_k$ in S^n we have

$$\nabla^k F(X)[H_1, \ldots, H_k] = 2 \sum_{\sigma \in \tilde{P}_0^k} \sum_{p_1,\ldots,p_k=1}^{n,\ldots,n} \mathscr{A}(\lambda(X))^{p_1\ldots p_k} \tilde{H}_1^{p_1 p_{\sigma(1)}} \cdots \tilde{H}_k^{p_k p_{\sigma(k)}}, \tag{50}$$

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

If, $H_1 = \cdots = H_k$ then (50) can be simplified even more.

Theorem 6.12. *For $g \in C^k(I)$ and any $H \in S^n$*

$$\nabla^k F(X)[H, \ldots, H] = (k-1)! \sum_{p_1,\ldots,p_k=1}^{n,\ldots,n} \mathscr{A}(\lambda(X))^{p_1\ldots p_k} \tilde{H}^{p_1 p_2} \tilde{H}^{p_2 p_3} \cdots \tilde{H}^{p_k p_1}, \tag{51}$$

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

Proof. Let H be in S^n. Using (49), (10), and (5) we find

$$\begin{aligned}
\nabla^k F(X)[H, \ldots, H] &= V \left(\sum_{\sigma \in \tilde{P}^k} \operatorname{Diag}^{\sigma} \mathscr{A}(\lambda(X)) \right) V^{\mathrm{T}}[H, \ldots, H] \\
&= \sum_{\sigma \in \tilde{P}^k} \langle \mathscr{A}(\lambda(X)), \tilde{H} \circ_\sigma \tilde{H} \circ_\sigma \cdots \circ_\sigma \tilde{H} \rangle \\
&= \sum_{\sigma \in \tilde{P}^k} \sum_{p_1,\ldots,p_k=1}^{n,\ldots,n} \mathscr{A}(\lambda(X))^{p_1\ldots p_k} \tilde{H}^{p_1 p_{\sigma^{-1}(1)}} \cdots \tilde{H}^{p_k p_{\sigma^{-1}(k)}}.
\end{aligned}$$

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

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

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 gives 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

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

Notice that $p_{i+1} = q_{\sigma^{-(i+1)}(1)} = q_{\sigma^{-1}(\sigma^{-i}(1))}$. After the substitution $\tilde{H}^{p_1 p_2} \tilde{H}^{p_2 p_3} \cdots \tilde{H}^{p_k p_1}$ becomes the product

$$\begin{aligned} &\tilde{H}^{q_{\sigma^{-1}(1)} q_{\sigma^{-2}(1)}} \tilde{H}^{q_{\sigma^{-2}(1)} q_{\sigma^{-3}(1)}} \cdots \tilde{H}^{q_{\sigma^{-k}(1)} q_{\sigma^{-1}(1)}} \\ &\quad = \tilde{H}^{q_{\sigma^{-1}(1)} q_{\sigma^{-1}(\sigma^{-1}(1))}} \tilde{H}^{q_{\sigma^{-2}(1)} q_{\sigma^{-1}(\sigma^{-2}(1))}} \cdots \tilde{H}^{q_{\sigma^{-k}(1)} q_{\sigma^{-1}(\sigma^{-k}(1))}} \\ &\quad = \tilde{H}^{q_1 q_{\sigma^{-1}(1)}} \tilde{H}^{q_2 q_{\sigma^{-1}(2)}} \cdots \tilde{H}^{q_k q_{\sigma^{-1}(k)}}. \end{aligned}$$

The final equality follows by a re-ordering of the product since the indices $\{\sigma^{-1}(1), \sigma^{-2}(1), \dots, \sigma^{-k}(1)\}$ are a permutation of the indices $\{1, 2, \dots, k\}$. Finally we have

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

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

7. The Hessian of a general spectral function

In this section we calculate a formula for the Hessian of a general spectral functions at an arbitrary symmetric matrix. That formula was first obtained in [17] but the insight for it came from [18]. Our current approach is streamlined and shows clearly 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 $\mathscr{A}_{(1)(2)} : \mathbb{R}^n \to M^n$ be defined by

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

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

$$\mathscr{A}^{i_1 i_2}_{(12)}(x) = \begin{cases} 0, & \text{if } i_1 = i_2, \\ f''_{i_1 i_1}(x) - f''_{i_1 i_2}(x), & \text{if } i_1 \sim_x i_2 \text{ and } i_1 \neq i_2, \\ \frac{f'_{i_2}(x) - f'_{i_1}(x)}{x_{i_2} - x_{i_1}}, & \text{if } i_1 \nsim_x i_2. \end{cases}$$

Several of the properties of $\mathscr{A}_{(12)}(x)$ 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*

$$\begin{aligned}\mathscr{A}^{i_1 i_2}_{(12)}(x) = \int_0^1 & f''_{i_1 i_1}(\ldots, x_{i_1} + t(x_{i_2} - x_{i_1}), \ldots, x_{i_2} + t(x_{i_1} - x_{i_2}), \ldots) \\ & - f''_{i_1 i_2}(\ldots, x_{i_1} + t(x_{i_2} - x_{i_1}), \ldots, x_{i_2} + t(x_{i_1} - x_{i_2}), \ldots)\, \mathrm{d}t,\end{aligned}$$

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 does not depend on t. In the third case, $i_1 \nsim_x i_2$, the Fundamental Theorem of Calculus, tell us that

$$\begin{aligned} & \frac{1}{x_{i_2} - x_{i_1}} \int_0^1 \frac{\partial}{\partial t} f'_{i_1}(\ldots, x_{i_1} + t(x_{i_2} - x_{i_1}), \ldots, x_{i_2} + t(x_{i_1} - x_{i_2}), \ldots)\, \mathrm{d}t \\ & \quad = \frac{f'_{i_1}(\ldots, x_{i_2}, \ldots, x_{i_1}, \ldots) - f'_{i_1}(\ldots, x_{i_1}, \ldots, x_{i_2}, \ldots)}{x_{i_2} - x_{i_1}} \\ & \quad = \frac{f'_{i_2}(\ldots, x_{i_1}, \ldots, x_{i_2}, \ldots) - f'_{i_1}(\ldots, x_{i_1}, \ldots, x_{i_2}, \ldots)}{x_{i_2} - x_{i_1}} \\ & \quad = \mathscr{A}^{i_1 i_2}_{(12)}(x). \end{aligned}$$

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 maps $x \to \mathscr{A}_{(1)(2)}(x)$ and $x \to \mathscr{A}_{(12)}(x)$ are point symmetric.*

Proof. Lemma 2.5 shows that $x \mapsto \mathscr{A}_{(1)(2)}(x)$ is point symmetric, so if $i_1 \sim_x j_1$, then $f''_{i_1 i_1}(x) = f''_{j_1 j_1}(\mu)$. 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 $f''_{i_1 i_2}(x) = f''_{j_1 j_2}(x)$. Point symmetry of $x \mapsto \nabla f(x)$ implies that if $i_1 \sim_x j_1$, then $f'_{i_1}(x) = f'_{j_1}(x)$. □

Lemma 7.1 ensures that if $f(x)$ is twice continuously differentiable then $\mathscr{A}_{(1)(2)}(x)$ and $\mathscr{A}_{(12)}(x)$ are symmetric matrices and are 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 same at $\lambda(X)$. The 'only if' direction can be seen by restricting $f \circ \lambda$ to the subspace of diagonal matrices. To show the 'if' direction, without loss of generality assume that $X = \operatorname{Diag} \mu$, 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:

$$
\begin{aligned}
&\nabla^2(f\circ\lambda)(\operatorname{Diag}\mu)[M]\\
&\quad=\lim_{m\to\infty}\frac{\nabla(f\circ\lambda)(\operatorname{Diag}\mu+M_m)-\nabla(f\circ\lambda)(\operatorname{Diag}\mu)}{\|M_m\|}\\
&\quad=\lim_{m\to\infty}\frac{U_m(\operatorname{Diag}^{(1)}\nabla f(\lambda(\operatorname{Diag}\mu+M_m)))U_m^{\mathrm{T}}-\operatorname{Diag}^{(1)}\nabla f(\mu)}{\|M_m\|}\\
&\quad=\lim_{m\to\infty}\frac{U_m(\operatorname{Diag}^{(1)}\nabla f(\mu+h_m+\mathrm{o}(\|M_m\|)))U_m^{\mathrm{T}}-\operatorname{Diag}^{(1)}\nabla f(\mu)}{\|M_m\|}\\
&\quad=\lim_{m\to\infty}\frac{U_m(\operatorname{Diag}^{(1)}(\nabla f(\mu)+\nabla^2 f(\mu)[h_m]+\mathrm{o}(\|M_m\|)))U_m^{\mathrm{T}}-\operatorname{Diag}^{(1)}\nabla f(\mu)}{\|M_m\|}\\
&\quad=\lim_{m\to\infty}\frac{U_m(\operatorname{Diag}^{(1)}(\nabla f(\mu)))U_m^{\mathrm{T}}-\operatorname{Diag}^{(1)}\nabla f(\mu)}{\|M_m\|}+U(\operatorname{Diag}^{(1)}(\nabla^2 f(\mu)[h]))U^{\mathrm{T}}.
\end{aligned}
\tag{53}
$$

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

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

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

$$
\begin{aligned}
&U(\operatorname{Diag}^{(1)}(\nabla^2 f(\mu)[h]))U^{\mathrm{T}}\\
&\quad=U(\operatorname{Diag}^{(1)}((\mathscr{A}_{(1)(2)}-\operatorname{Diag}b+\operatorname{Diag}b)[h]))U^{\mathrm{T}}\\
&\quad=U(\operatorname{Diag}^{(1)}((\mathscr{A}_{(1)(2)}-\operatorname{Diag}b)[h]))U^{\mathrm{T}}+U(\operatorname{Diag}^{(1)}((\operatorname{Diag}b)[h]))U^{\mathrm{T}}\\
&\quad=(\operatorname{Diag}^{(1)(2)}(\mathscr{A}_{(1)(2)}-\operatorname{Diag}b))[M]+(\operatorname{Diag}^{(12)}b_{\mathrm{in}}^{(1)})[M].
\end{aligned}
\tag{55}
$$

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

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

$$
\mathscr{A}_{(12)}+\operatorname{Diag}b=T_{\mathrm{out}}^{(1)}+b_{\mathrm{in}}^{(1)}. \tag{56}
$$

Putting (53)–(56) together we obtain:

$$
\begin{aligned}
\nabla^2(f\circ\lambda)(\operatorname{Diag}\mu)&=\operatorname{Diag}^{(12)}T_{\mathrm{out}}^{(1)}+\operatorname{Diag}^{(1)(2)}(\mathscr{A}_{(1)(2)}-\operatorname{Diag}b)+\operatorname{Diag}^{(12)}b_{\mathrm{in}}^{(1)}\\
&=\operatorname{Diag}^{(1)(2)}(\mathscr{A}_{(1)(2)}-\operatorname{Diag}b)+\operatorname{Diag}^{(12)}(\mathscr{A}_{(12)}+\operatorname{Diag}b)\\
&=\operatorname{Diag}^{(1)(2)}\mathscr{A}_{(1)(2)}+\operatorname{Diag}^{(12)}\mathscr{A}_{(12)}.
\end{aligned}
$$

In the third 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 is

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

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

Finally, when f is C^2 both $\mathscr{A}_{(1)(2)}(x)$ and $\mathscr{A}_{(12)}(x)$ are continuous and by [22, Proposition 6.2] $\nabla^2(f \circ \lambda)(X)$ is continuous as well.

Appendix A. A refinement of a perturbation result for eigenvectors

The main tool in the derivation of the formula for the Hessian in [17] was Lemma 2.4. The statement of that lemma was broken down into nine parts, which led to consideration of a 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 [17] into a form more suitable for computations. Section 3 gives the relevant notation.

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 $l = 1, \dots, r$.

Theorem A.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*

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

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:*

$$U^{\mathrm{T}} M_{\mathrm{in}} U = \operatorname{Diag} h, \tag{A.1}$$

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

$$\sum_{p \in I_t} U_m^{ip} U_m^{jp} = \delta_{ij}\delta_{lt} + \frac{\delta_{lt} - \delta_{st}}{\mu_i - \mu_j} M^{ij} \|M_m\| + \mathrm{o}(\|M_m\|), \tag{A.2}$$

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 [17]. Indeed, Part (i) is [17, Lemma 2.4 Part (i)]. The equality in Part (ii) is an aggregate version of Parts (iv) and (vii) from Lemma 2.4 in [17]. To prove Part (iii) we consider several cases.

Case 1. If $i = j \in I_l$ and $t = l$, then (A.2) becomes $\sum_{p \in I_l} (U_m^{ip})^2 = 1 + \mathrm{o}(\|M_m\|)$, which is exactly Part (ii), Lemma 2.4 in [17].

Case 2. If $i = j \in I_l$ and $t \neq l$, then (A.2) becomes $\sum_{p \in I_t} (U_m^{ip})^2 = \mathrm{o}(\|M_m\|)$, which is a consequence of Part (iii), Lemma 2.4 in [17].

Case 3. If $i \neq j \in I_l$ and $t = l$, then (A.2) becomes $\sum_{p \in I_l} U_m^{ip} U_m^{jp} = \mathrm{o}(\|M_m\|)$, which is exactly Part (vi), Lemma 2.4 in [17].

Case 4. If $i \neq j \in I_l$ and $t \neq l$, then (A.2) becomes $\sum_{p \in I_t} U_m^{ip} U_m^{jp} = \mathrm{o}(\|M_m\|)$, which is a consequence of Part (v), Lemma 2.4 in [17].

Case 5. If $i \in I_l$, $j \in I_s$, with $l \neq s \neq t \neq l$, then (A.2) becomes $\sum_{p \in I_t} U_m^{ip} U_m^{jp} = \mathrm{o}(\|M_m\|)$, which is a consequence of Part (viii), Lemma 2.4 in [17].

Case 6. If $i \in I_l$, $j \in I_s$, with $l \neq s$ and $t = l$, then (A.2) becomes

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

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 (A.2) becomes

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

We now show that the expressions in both Case 6 and Case 7 are valid. Part (ix) from Lemma 2.4 in [17] says that if $i \in I_l$, $j \in I_s$ with $l \neq s$, we have

$$\lim_{m \to \infty} g \left(\mu_{\iota_l} \frac{\sum_{p \in I_l} U_m^{ip} U_m^{jp}}{\|M_m\|} + \mu_{\iota_s} \frac{\sum_{p \in I_s} U_m^{ip} U_m^{jp}}{\|M_m\|} g \right) = M^{ij}. \tag{A.3}$$

Introduce the notation

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

and notice that

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

because U_m is orthogonal and the numerator of the last sum is the product of its ith and jth row. Next, by Case 5 we have

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

so

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

For arbitrary reals a and b we compute

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

as $m \to \infty$. Using (A.3), 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. $\square$

Appendix B. Tensor analysis

The aim of this appendix is to prove 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 $l = 1, \dots, r$.

Theorem B.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*

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

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

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

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 (B.1) and compare the results. Both sides of (B.1) are linear in each argument H_s. That is why it is enough to prove the result when H_s, for $s = 1, \dots, k$, is an arbitrary matrix, $H_{i_s j_s}$, from the standard basis on M^n. In that case

$$(U_m(\operatorname{Diag}^{\sigma}T)U_m^{\mathrm{T}} - \operatorname{Diag}^{\sigma}T)[H_{i_1 j_1}, \dots, H_{i_k j_k}]$$
$$= (U_m(\operatorname{Diag}^{\sigma}T)U_m^{\mathrm{T}})^{i_1\dots i_k}_{j_1\dots j_k} - (\operatorname{Diag}^{\sigma}T)^{i_1\dots i_k}_{j_1\dots j_k}. \tag{B.2}$$

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 (B.2):

$$\begin{aligned}(U_m(\operatorname{Diag}^{\sigma}T)U_m^{\mathrm{T}})^{i_1\dots i_k}_{j_1\dots j_k} &= \sum_{\substack{p_\eta,q_\eta=1\\ \eta=1,\dots,k}}^{n,\dots,n}(\operatorname{Diag}^{\sigma}T)^{p_1\dots p_k}_{q_1\dots q_k}\prod_{\nu=1}^{k}U_m^{i_\nu p_\nu}U_m^{j_\nu q_\nu}\\
&= \sum_{\substack{p_\eta=1\\ \eta=1,\dots,k}}^{n,\dots,n}T^{p_1\dots p_k}\prod_{\nu=1}^{k}U_m^{i_\nu p_\nu}U_m^{j_\nu p_{\sigma^{-1}(\nu)}}\\
&= \sum_{\substack{p_\eta=1\\ \eta=1,\dots,k}}^{n,\dots,n}T^{p_1\dots p_k}\prod_{\nu=1}^{k}U_m^{i_\nu p_\nu}U_m^{j_{\sigma(\nu)} p_\nu}\\
&= \sum_{\substack{t_\eta=1\\ \eta=1,\dots,k}}^{r,\dots,r}T^{\iota_{t_1}\dots\iota_{t_k}}\prod_{\nu=1}^{k}\left(\sum_{p_\nu\in I_{t_\nu}}U_m^{i_\nu p_\nu}U_m^{j_{\sigma(\nu)} p_\nu}\right).\end{aligned}$$

Putting everything together, we see that to evaluate the limit on the left-hand side of (B.1) we must compute

$$\lim_{m\to\infty} \frac{\sum_{t_1,\dots,t_k=1}^{r,\dots,r} T^{\iota_{t_1}\dots\iota_{t_k}} \prod_{\nu=1}^{k} \left(\sum_{p_\nu\in I_{t_\nu}} U_m^{i_\nu p_\nu} U_m^{j_{\sigma(\nu)} p_\nu}\right) - (\mathrm{Diag}^\sigma T)_{j_1\cdots j_k}^{i_1\dots i_k}}{\|M_m\|}. \tag{B.3}$$

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

$$(i_1, j_{\sigma(1)}), (i_2, j_{\sigma(2)}), \dots, (i_k, j_{\sigma(k)}) \tag{B.4}$$

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 (A.2), for any t_1, t_2 we observe that:

$$\begin{aligned}
&\lim_{m\to\infty} \frac{1}{\|M_m\|} \left(\sum_{p_1\in I_{t_1}} U_m^{i_1 p_1} U_m^{j_{\sigma(1)} p_1}\right) \left(\sum_{p_2\in I_{t_2}} U_m^{i_2 p_2} U_m^{j_{\sigma(2)} p_2}\right)\\
&= \lim_{m\to\infty} \frac{1}{\|M_m\|} \left(\delta_{i_1 j_{\sigma(1)}} \delta_{v_1 t_1} + \frac{\delta_{v_1 t_1} - \delta_{s_1 t_1}}{\mu_{i_1} - \mu_{j_{\sigma(1)}}} M^{i_1 j_{\sigma(1)}} \|M_m\| + \mathrm{o}(\|M_m\|)\right)\\
&\quad\times \left(\delta_{i_2 j_{\sigma(2)}} \delta_{v_2 t_2} + \frac{\delta_{v_2 t_2} - \delta_{s_2 t_2}}{\mu_{i_2} - \mu_{j_{\sigma(2)}}} M^{i_2 j_{\sigma(2)}} \|M_m\| + \mathrm{o}(\|M_m\|)\right)\\
&= \lim_{m\to\infty} \frac{1}{\|M_m\|} \left(\frac{\delta_{v_1 t_1} - \delta_{s_1 t_1}}{\mu_{i_1} - \mu_{j_{\sigma(1)}}} M^{i_1 j_{\sigma(1)}} \|M_m\| + \mathrm{o}(\|M_m\|)\right)\\
&\quad\times \left(\frac{\delta_{v_2 t_2} - \delta_{s_2 t_2}}{\mu_{i_2} - \mu_{j_{\sigma(2)}}} M^{i_2 j_{\sigma(2)}} \|M_m\| + \mathrm{o}(\|M_m\|)\right)\\
&= 0.
\end{aligned}$$

Since in this case by definition $(\mathrm{Diag}^\sigma T)_{j_1\cdots j_k}^{i_1\dots i_k} = 0$, we see that (B.3) is 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 (A.2), for arbitrary t we obtain:

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

In this subcase we again have $(\mathrm{Diag}^\sigma T)_{j_1\cdots j_k}^{i_1\dots i_k} = 0$; thus (B.3) is equal to zero.

If i_l and $j_{\sigma(l)}$ are in different blocks, $v_l \neq s_l$, then $(\mathrm{Diag}^\sigma T)^{i_1\ldots i_k}_{j_1\ldots j_k} = 0$ and by (A.2) we obtain:

$$\lim_{m\to\infty}\frac{1}{\|M_m\|}\left(\sum_{t_1,\ldots,t_k=1}^{r,\ldots,r} T^{\iota_{t_1}\ldots\iota_{t_k}}\prod_{\nu=1}^{k}\left(\sum_{p_\nu\in I_{t_\nu}} U_m^{i_\nu p_\nu}U_m^{j_{\sigma(\nu)}p_\nu}\right)\right)$$
$$= \lim_{m\to\infty}\left(\sum_{t_1,\ldots,t_k=1}^{r,\ldots,r}\frac{T^{\iota_{t_1}\ldots\iota_{t_k}}}{\|M_m\|}\prod_{\nu=1}^{k}\left(\delta_{i_\nu j_{\sigma(\nu)}}\delta_{v_\nu t_\nu}+\frac{\delta_{v_\nu t_\nu}-\delta_{s_\nu t_\nu}}{\mu_{i_\nu}-\mu_{j_{\sigma(\nu)}}}M^{i_\nu j_{\sigma(\nu)}}\|M_m\|+\mathrm{o}(\|M_m\|)\right)\right). \tag{B.5}$$

We show that the limit of at most two terms of the big sum in (B.5) may be non zero. Indeed, summands corresponding to k-tuples $(t_1,\ldots,t_k)$ 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

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

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

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

since $i_\nu = j_{\sigma(\nu)}$ for $\nu \neq l$. Thus, the limit of the summand in (B.5) corresponding to the k-tuple $(v_1,\ldots,v_{l-1},v_l,v_{l+1},\ldots,v_k)$ is

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

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

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

Putting these two limits together we see that (B.5), and therefore (B.3) is

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

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

Consider now the final case: $i_\nu = j_{\sigma(\nu)}$ for all $\nu = 1, \ldots, k$. Using (A.2) one can see that the only summand that may have a non-zero limit in the sum in the numerator of (B.3) is the one corresponding to the multi-index $(t_1, \ldots, t_k) = (v_1, \ldots, v_k)$. Thus, using the block-constant structure of T ($i_\nu \in I_{v_\nu}$ for all $\nu = 1, \ldots, k$), (B.3) is equal to

$$\lim_{m\to\infty} \frac{1}{\|M_m\|}(T^{i_1\ldots i_k}(1 + \mathrm{o}(\|M_m\|)) - T^{i_1\ldots i_k}) = 0.$$

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

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

$$\begin{aligned}
&\sum_{l=1}^{k}(\mathrm{Diag}^{\sigma_{(1)}} T_{\mathrm{out}}^{(l)})[H_{i_1 j_1}, \ldots, H_{i_k j_k}, M_{\mathrm{out}}]\\
&\quad = \sum_{l=1}^{k}\langle T_{\mathrm{out}}^{(l)}, H_{i_1 j_1} \circ_{\sigma_{(1)}} \cdots \circ_{\sigma_{(1)}} H_{i_k j_k} \circ_{\sigma_{(1)}} M_{\mathrm{out}}\rangle\\
&\quad = \sum_{l=1}^{k}(T_{\mathrm{out}}^{(l)})^{i_1\ldots i_k j_{\sigma_{(1)}(k+1)}}(\delta_{i_1 j_{\sigma_{(1)}(1)}} \cdots \widehat{\delta_{i_l j_{\sigma_{(1)}(l)}}} \cdots \delta_{i_k j_{\sigma_{(1)}(k)}})M_{\mathrm{out}}^{j_{\sigma_{(1)}(k+1)} i_{\sigma_{(l)}^{-1}(k+1)}}\\
&\quad = \sum_{l=1}^{k}(T_{\mathrm{out}}^{(l)})^{i_1\ldots i_k j_{\sigma(l)}}(\delta_{i_1 j_{\sigma_{(1)}(1)}} \cdots \widehat{\delta_{i_l j_{\sigma_{(1)}(l)}}} \cdots \delta_{i_k j_{\sigma_{(1)}(k)}})M_{\mathrm{out}}^{j_{\sigma(l)} i_l}\\
&\quad = \sum_{l=1}^{k}(T_{\mathrm{out}}^{(l)})^{i_1\ldots i_k j_{\sigma(l)}}(\delta_{i_1 j_{\sigma(1)}} \cdots \widehat{\delta_{i_l j_{\sigma(l)}}} \cdots \delta_{i_k j_{\sigma(k)}})M_{\mathrm{out}}^{j_{\sigma(l)} i_l}.
\end{aligned}$$

The final equality results from changing the circumflexed factor (for each fixed l) while keeping the other factors the same. If at least two of the pairs

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

have different entries, then the final sum is zero. Now suppose exactly one of the pairs has unequal entries, say $i_l \neq j_{\sigma(l)}$. Then the sum is equal to

$$(T_{\mathrm{out}}^{(l)})^{i_1\ldots i_k j_{\sigma(l)}}(\delta_{i_1 j_{\sigma(1)}} \cdots \widehat{\delta_{i_l j_{\sigma(l)}}} \cdots \delta_{i_k j_{\sigma(k)}})M_{\mathrm{out}}^{j_{\sigma(l)} i_l}. \tag{B.6}$$

If i_l and $j_{\sigma(l)}$ are in the same block, then $(T_{\mathrm{out}}^{(l)})^{i_1\ldots i_k j_{\sigma(l)}} = 0$ by the definition of $T_{\mathrm{out}}^{(l)}$. If i_l and $j_{\sigma(l)}$ are not in the same block, then (B.6) is equal to

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

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

Proposition B.2. *Let T be any $k+1$-tensor on $\mathbb{R}^n$, let $x \in \mathbb{R}^n$, let V be in O^n, and let σ be in P^k. Then*

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

Proof. Let $H_{i_1 j_1}, \dots, H_{i_k j_k}$ be any k basic matrices. Since $\sigma_{(k+1)}(i) = \sigma(i)$ for all $i \in \mathbb{N}_k$ and $\sigma_{(k+1)}(k+1) = k+1$, we can use Theorem 2.6 twice to compute

$$\begin{aligned}
&(V(\mathrm{Diag}^{\sigma}(T[x]))V^{\mathrm{T}})^{i_1 \dots i_k}_{j_1 \dots j_k} = (V(\mathrm{Diag}^{\sigma}(T[x]))V^{\mathrm{T}})[H_{i_1 j_1}, \dots, H_{i_k j_k}] \\
&\quad = \langle T[x], \tilde{H}_{i_1 j_1} \circ_{\sigma} \dots \circ_{\sigma} \tilde{H}_{i_k j_k} \rangle \\
&\quad = \sum_{p_1,\dots,p_k=1}^{n,\dots,n} (T[x])^{p_1 \dots p_k} \tilde{H}_{i_1 j_1}^{p_1 p_{\sigma^{-1}(1)}} \dots \tilde{H}_{i_k j_k}^{p_k p_{\sigma^{-1}(k)}} \\
&\quad = \sum_{p_1,\dots,p_k,p_{k+1}=1}^{n,\dots,n} T^{p_1 \dots p_{k+1}} x^{p_{k+1}} \tilde{H}_{i_1 j_1}^{p_1 p_{\sigma^{-1}(1)}} \dots \tilde{H}_{i_k j_k}^{p_k p_{\sigma^{-1}(k)}} \\
&\quad = \sum_{p_1,\dots,p_k,p_{k+1}=1}^{n,\dots,n} T^{p_1 \dots p_{k+1}} \tilde{H}_{i_1 j_1}^{p_1 p_{\sigma^{-1}_{(k+1)}(1)}} \dots \tilde{H}_{i_k j_k}^{p_k p_{\sigma^{-1}_{(k+1)}(k)}} (\mathrm{Diag}\, x)^{p_{k+1} p_{\sigma^{-1}_{(k+1)}(k+1)}} \\
&\quad = \langle T, \tilde{H}_{i_1 j_1} \circ_{\sigma_{(k+1)}} \dots \circ_{\sigma_{(k+1)}} \tilde{H}_{i_k j_k} \circ_{\sigma_{(k+1)}} \mathrm{Diag}\, x \rangle \\
&\quad = (V(\mathrm{Diag}^{\sigma_{(k+1)}} T)V^{\mathrm{T}})[H_{i_1 j_1}, \dots, H_{i_k j_k}, V(\mathrm{Diag}\, x)V^{\mathrm{T}}] \\
&\quad = ((V(\mathrm{Diag}^{\sigma_{(k+1)}} T)V^{\mathrm{T}})[V(\mathrm{Diag}\, x)V^{\mathrm{T}}])^{i_1 \dots i_k}_{j_1 \dots j_k}.
\end{aligned}$$

Since the indices $i_1, \dots, i_k$ and $j_1, \dots, j_k$ are arbitrary we are done. □

The next lemma says that for any block-constant tensor T, $\mathrm{Diag}^{\sigma} T$ is invariant under conjugation with a block-diagonal orthogonal matrix.

Lemma B.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 $\mathbb{N}_k$*

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

Proof. Let $\{I_1, \dots, I_r\}$ be the partitioning of the integers $\mathbb{N}_n$ that determines the block structure, so $U^{ip}U^{jp} = 0$ whenever $i \nsim j$ or $i \nsim p$, and $\sum_{p \in I_s} U^{ip}U^{jp} = \delta_{ij}$ whenever $i \in I_s$. Let $(i_1, \dots, i_k)$ be an arbitrary multi index and suppose that $i_l \in I_{v_l}$ for $l = 1, \dots, k$. We expand the left-hand side of the identity:

$$\begin{aligned}
(U(\mathrm{Diag}^{\sigma} T)U^{\mathrm{T}})^{i_1 \dots i_k}_{j_1 \dots j_k} &= \sum_{\substack{p_s, q_s = 1 \\ s=1,\dots,k}}^{n,\dots,n} (\mathrm{Diag}^{\sigma} T)^{p_1 \dots p_k}_{q_1 \dots q_k} U^{i_1 p_1} U^{j_1 q_1} \dots U^{i_k p_k} U^{j_k q_k} \\
&= \sum_{p_1,\dots,p_k=1}^{n,\dots,n} T^{p_1 \dots p_k} U^{i_1 p_1} U^{j_1 p_{\sigma^{-1}(1)}} \dots U^{i_k p_k} U^{j_k p_{\sigma^{-1}(k)}} \\
&= \sum_{p_1,\dots,p_k=1}^{n,\dots,n} T^{p_1 \dots p_k} U^{i_1 p_1} U^{j_{\sigma(1)} p_1} \dots U^{i_k p_k} U^{j_{\sigma(k)} p_k}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{t_1,\dots,t_k=1}^{r,\dots,r} T^{\iota_{t_1}\dots\iota_{t_k}} \sum_{\substack{p_l\in I_{t_l}\\ l=1,\dots,k}} U^{i_1p_1}U^{j_{\sigma(1)}p_1}\dots U^{i_kp_k}U^{j_{\sigma(k)}p_k}\\
&= T^{\iota_{v_1}\dots\iota_{v_k}} \sum_{\substack{p_l\in I_{v_l}\\ l=1,\dots,k}} U^{i_1p_1}U^{j_{\sigma(1)}p_1}\dots U^{i_kp_k}U^{j_{\sigma(k)}p_k}\\
&= T^{\iota_{v_1}\dots\iota_{v_k}}\delta_{i_1j_{\sigma(1)}}\cdots\delta_{i_kj_{\sigma(k)}}\\
&= T^{i_1\dots i_k}\delta_{i_1j_{\sigma(1)}}\cdots\delta_{i_kj_{\sigma(k)}}\\
&= (\mathrm{Diag}^{\sigma}T)^{i_1\dots i_k}_{j_1\dots j_k}.
\end{aligned}
$$

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 in (3).

Theorem B.4. *Let $U\in O^n$ be a block diagonal orthogonal matrix. Let $M\in S^n$ be given and let $h\in\mathbb{R}^n$ be such that*

$$U^{\mathrm{T}}M_{\mathrm{in}}U=\mathrm{Diag}\,h. \tag{B.7}$$

Let $H_1,\dots,H_k$ 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 $\mathbb{R}^n$,*

$$\langle T^{\tau_l}[h],\tilde H_1\circ_\sigma\cdots\circ_\sigma\tilde H_k\rangle=\langle T^{(l)}_{\mathrm{in}},H_1\circ_{\sigma(1)}\cdots\circ_{\sigma(1)}H_k\circ_{\sigma(1)}M_{\mathrm{in}}\rangle \quad \textit{for all } l=1,\dots,k,$$

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

Proof. To see that the first identity holds we use Theorem 2.6, Proposition B.2, (B.7), and Lemma B.3 in that order, as follows:

$$
\begin{aligned}
\langle T[h],\tilde H_1\circ_\sigma\cdots\circ_\sigma\tilde H_k\rangle &= (U(\mathrm{Diag}^{\sigma}T[h])U^{\mathrm{T}})[H_1,\dots,H_k]\\
&= (U(\mathrm{Diag}^{\sigma_{(k+1)}}T)U^{\mathrm{T}})[H_1,\dots,H_k,U(\mathrm{Diag}\,h)U^{\mathrm{T}}]\\
&= (U(\mathrm{Diag}^{\sigma_{(k+1)}}T)U^{\mathrm{T}})[H_1,\dots,H_k,M_{\mathrm{in}}]\\
&= (\mathrm{Diag}^{\sigma_{(k+1)}}T)[H_1,\dots,H_k,M_{\mathrm{in}}]\\
&= \langle T,H_1\circ_{\sigma(k+1)}\cdots\circ_{\sigma_{(k+1)}}H_k\circ_{\sigma_{(k+1)}}M_{\mathrm{in}}\rangle.
\end{aligned}
$$

The final equality follows from Theorem 2.6.

To show the second identity, it suffices to prove it for arbitrary basic matrices $H_{i_sj_s}s=1,\dots,k$. Fix k basic matrices $H_{i_1j_1},\dots,H_{i_kj_k}$ and suppose that $i_l\in I_{v_l}$ for $l=1,\dots,k$. Then

$$
\begin{aligned}
&\langle T^{\tau_l}[h],\tilde H_{i_1j_1}\circ_\sigma\cdots\circ_\sigma\tilde H_{i_kj_k}\rangle\\
&\quad=(U(\mathrm{Diag}^{\sigma}T^{\tau_l}[h])U^{\mathrm{T}})[H_{i_1j_1},\dots,H_{i_kj_k}]
\end{aligned}
$$

$$
\begin{aligned}
&= (U(\mathrm{Diag}^{\sigma} T^{\tau_l}[h])U^{\mathrm{T}})^{i_1\ldots i_k}_{j_1\cdots j_k}\\
&= \sum_{\substack{p_1,\ldots,p_k=1\\ q_1,\ldots,q_k=1}}^{n,\ldots,n} (\mathrm{Diag}^{\sigma} T^{\tau_l}[h])^{p_1\cdots p_k}_{q_1\cdots q_k} U^{i_1p_1}U^{j_1q_1}\ldots U^{i_kp_k}U^{j_kq_k}\\
&= \sum_{p_1,\ldots,p_k=1}^{n,\ldots,n} (T^{\tau_l}[h])^{p_1\cdots p_k} U^{i_1p_1}U^{j_1p_{\sigma^{-1}(1)}}\ldots U^{i_kp_k}U^{j_kp_{\sigma^{-1}(k)}}\\
&= \sum_{p_1,\ldots,p_k=1}^{n,\ldots,n} (T^{\tau_l}[h])^{p_1\cdots p_k} U^{i_1p_1}U^{j_{\sigma(1)}p_1}\ldots U^{i_kp_k}U^{j_{\sigma(k)}p_k}\\
&= \sum_{p_1,\ldots,p_k=1}^{n,\ldots,n}\sum_{p_{k+1}=1}^{n} (T^{\tau_l})^{p_1\cdots p_kp_{k+1}}h^{p_{k+1}} U^{i_1p_1}U^{j_{\sigma(1)}p_1}\ldots U^{i_kp_k}U^{j_{\sigma(k)}p_k}\\
&= \sum_{p_1,\ldots,p_k=1}^{n,\ldots,n} T^{p_1\cdots p_k}h^{p_l} U^{i_1p_1}U^{j_{\sigma(1)}p_1}\ldots U^{i_kp_k}U^{j_{\sigma(k)}p_k}\\
&= \sum_{t_1,\ldots,t_k=1}^{r,\ldots,r} T^{\iota_{t_1}\cdots\iota_{t_k}} \sum_{\substack{p_\eta\in I_{t_\eta}\\ \eta=1,\ldots,k}} h^{p_l} U^{i_1p_1}U^{j_{\sigma(1)}p_1}\ldots U^{i_kp_k}U^{j_{\sigma(k)}p_k}\\
&= T^{\iota_{v_1}\cdots\iota_{v_k}} \sum_{\substack{p_l\in I_{v_l}\\ l=1,\ldots,k}} h^{p_l} U^{i_1p_1}U^{j_{\sigma(1)}p_1}\ldots U^{i_kp_k}U^{j_{\sigma(k)}p_k}\\
&= T^{\iota_{v_1}\cdots\iota_{v_k}}\delta_{i_1j_{\sigma(1)}}\cdots\widehat{\delta_{i_lj_{\sigma(l)}}}\cdots\delta_{i_kj_{\sigma(k)}}\sum_{p_l\in I_{v_l}} h^{p_l}U^{i_lp_l}U^{j_{\sigma(l)}p_l}\\
&= T^{i_1\ldots i_k}\delta_{i_1j_{\sigma(1)}}\cdots\widehat{\delta_{i_lj_{\sigma(l)}}}\cdots\delta_{i_kj_{\sigma(k)}}\sum_{p_l\in I_{v_l}} h^{p_l}U^{i_lp_l}U^{j_{\sigma(l)}p_l}\\
&= T^{i_1\ldots i_k}\delta_{i_1j_{\sigma(1)}}\cdots\widehat{\delta_{i_lj_{\sigma(l)}}}\cdots\delta_{i_kj_{\sigma(k)}}M_{\mathrm{in}}^{i_lj_{\sigma(l)}}.
\end{aligned}
$$

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

$$
\begin{aligned}
&\langle T_{\mathrm{in}}^{(l)}, H_{i_1j_1}\circ_{\sigma_{(l)}}\cdots\circ_{\sigma_{(l)}} H_{i_kj_k}\circ_{\sigma_{(l)}} M_{\mathrm{in}}\rangle\\
&= (T_{\mathrm{in}}^{(l)})^{i_1\ldots i_kj_{\sigma_{(l)}(k+1)}}\delta_{i_1j_{\sigma_{(l)}(1)}}\cdots\widehat{\delta_{i_lj_{\sigma_{(l)}(l)}}}\cdots\delta_{i_kj_{\sigma_{(l)}(k)}}M_{\mathrm{in}}^{j_{\sigma_{(l)}(k+1)}i_{\sigma_{(l)}^{-1}(k+1)}}\\
&= (T_{\mathrm{in}}^{(l)})^{i_1\ldots i_kj_{\sigma(l)}}\delta_{i_1j_{\sigma(1)}}\cdots\widehat{\delta_{i_lj_{k+1}}}\cdots\delta_{i_kj_{\sigma(k)}}M_{\mathrm{in}}^{j_{\sigma(l)}i_l}\\
&= T^{i_1\ldots i_k}\delta_{i_1j_{\sigma(1)}}\cdots\widehat{\delta_{i_lj_{k+1}}}\cdots\delta_{i_kj_{\sigma(k)}}M_{\mathrm{in}}^{j_{\sigma(l)}i_l}\\
&= T^{i_1\ldots i_k}\delta_{i_1j_{\sigma(1)}}\cdots\widehat{\delta_{i_lj_{k+1}}}\cdots\delta_{i_kj_{\sigma(k)}}M_{\mathrm{in}}^{i_lj_{\sigma(l)}}\\
&= T^{i_1\ldots i_k}\delta_{i_1j_{\sigma(1)}}\cdots\widehat{\delta_{i_lj_{\sigma(l)}}}\cdots\delta_{i_kj_{\sigma(k)}}M_{\mathrm{in}}^{i_lj_{\sigma(l)}}.
\end{aligned}
$$

In the third equality we used the fact that T is block constant, as well as the fact that $M_{\text{in}}^{j_{\sigma(l)} i_l} = 0$ if $j_{\sigma(l)} \nsim i_l$. In the fourth equality we used the fact that M is symmetric. The final equality holds because we changed the missing factor, while keeping the other factors the same. □

Proposition B.5. *Let $U \in O^n$ be block diagonal, let H be an $n \times n$ matrix, and let σ be any 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 $T^{p_1 \cdots p_l \cdots p_{k+1}} = 0$ whenever $p_l \sim p_{k+1}$, then*
$(U(\mathrm{Diag}^{\sigma_{(l)}} T)U^{\mathrm{T}})[H_{\text{in}}] = 0.$

(ii) *If T is a $(k+1)$-tensor such that for some fixed $l \in \mathbb{N}_k$ we have $T^{p_1 \cdots p_l \cdots p_{k+1}} = 0$ whenever $p_l \nsim p_{k+1}$, then*
$(U(\mathrm{Diag}^{\sigma_{(l)}} T)U^{\mathrm{T}})[H_{\text{out}}] = 0.$

(iii) *If T is any $(k+1)$-tensor, then*
$(U(\mathrm{Diag}^{\sigma_{(k+1)}} T)U^{\mathrm{T}})[H_{\text{out}}] = 0.$

Proof. Fix an index l in $\mathbb{N}_k$. Let $H_{i_1 j_1}, \ldots, H_{i_k j_k}$ be arbitrary basic matrices, and let H be an arbitrary matrix. Using the definitions we compute

$$
\begin{aligned}
&(U(\mathrm{Diag}^{\sigma_{(l)}} T)U^{\mathrm{T}})[H_{i_1 j_1}, \ldots, H_{i_k j_k}, H] \\
&= \sum_{i_{k+1}, j_{k+1}=1}^{n,n} (U(\mathrm{Diag}^{\sigma_{(l)}} T)U^{\mathrm{T}})^{i_1 \ldots i_{k+1}}_{j_1 \ldots j_{k+1}} H^{i_{k+1} j_{k+1}} \\
&= \sum_{i_{k+1}, j_{k+1}=1}^{n,n} \sum_{\substack{p_s, q_s=1 \\ s=1,\ldots,k+1}}^{n,\ldots,n} (\mathrm{Diag}^{\sigma_{(l)}} T)^{p_1 \cdots p_{k+1}}_{q_1 \cdots q_{k+1}} U^{i_1 p_1} U^{j_1 q_1} \ldots U^{i_{k+1} p_{k+1}} U^{j_{k+1} q_{k+1}} H^{i_{k+1} j_{k+1}} \\
&= \sum_{i_{k+1}, j_{k+1}=1}^{n,n} \sum_{\substack{p_s=1 \\ s=1,\ldots,k+1}}^{n,\ldots,n} T^{p_1 \cdots p_{k+1}} U^{i_1 p_1} U^{j_1 p_{\sigma_{(l)}^{-1}(1)}} \ldots U^{i_{k+1} p_{k+1}} U^{j_{k+1} p_{\sigma_{(l)}^{-1}(k+1)}} H^{i_{k+1} j_{k+1}} \\
&= \sum_{i_{k+1}, j_{k+1}=1}^{n,n} \sum_{\substack{p_s=1 \\ s=1,\ldots,k+1}}^{n,\ldots,n} \Big(T^{p_1 \cdots p_{k+1}} U^{i_1 p_1} U^{j_{\sigma_{(l)}(1)} p_1} \ldots U^{i_l p_l} U^{j_{\sigma_{(l)}(l)} p_l} \\
&\quad \ldots U^{i_{k+1} p_{k+1}} U^{j_{\sigma_{(l)}(k+1)} p_{k+1}} H^{i_{k+1} j_{k+1}} \Big) \\
&= \sum_{i_{k+1}, j_{k+1}=1}^{n,n} \sum_{\substack{p_s=1 \\ s=1,\ldots,k+1}}^{n,\ldots,n} (T^{p_1 \cdots p_{k+1}} U^{i_1 p_1} U^{j_{\sigma(1)} p_1} \ldots U^{i_l p_l} U^{j_{k+1} p_l} \\
&\quad \ldots U^{i_{k+1} p_{k+1}} U^{j_{\sigma(l)} p_{k+1}} H^{i_{k+1} j_{k+1}}).
\end{aligned}
$$

Now suppose that T is a $(k+1)$-tensor with $T^{p_1 \cdots p_l \cdots p_{k+1}} = 0$ 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 $T^{p_1 \cdots p_l \cdots p_{k+1}} = 0$. Thus every summand in the final double sum is zero.

In the second case, suppose T is a $(k+1)$-tensor with $T^{p_1\cdots p_l\cdots p_{k+1}} = 0$ whenever $p_l \nsim p_{k+1}$ and $H = H_{\text{out}}$. Then $H^{i_{k+1}j_{k+1}} \neq 0$ implies that $i_{k+1} \nsim j_{k+1}$. 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 $p_l \nsim p_{k+1}$, which implies that $T^{p_1\cdots p_l\cdots p_{k+1}} = 0$. The sum is zero.

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

$$
\begin{aligned}
&(U(\mathrm{Diag}^{\sigma_{(k+1)}} T)U^{\mathrm{T}})[H_{i_1j_1}, \ldots, H_{i_kj_k}, H] \\
&\quad = \sum_{i_{k+1}, j_{k+1}=1}^{n,n} \sum_{\substack{p_s=1 \\ s=1,\ldots,k+1}}^{n,\ldots,n} T^{p_1\cdots p_{k+1}} U^{i_1p_1} U^{j_{\sigma(1)}p_1} \\
&\qquad \cdots U^{i_kp_k} U^{j_{\sigma(k)}p_k} U^{i_{k+1}p_{k+1}} U^{j_{k+1}p_{k+1}} H^{i_{k+1}j_{k+1}}.
\end{aligned}
$$

Then $H^{i_{k+1}j_{k+1}} \neq 0$ implies that $i_{k+1} \nsim j_{k+1}$. 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 two main analytical tools.

Proof of Theorem 9. A consequence of Theorem B.1 and Proposition B.5. □

Proof of Theorem 10. A consequence of Theorem 2.6, Theorem B.4, Proposition B.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 entries, 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 $T_{\text{in}}^{(l)} = T^{\tau_l}$. Moreover, since Proposition B.5 holds for arbitrary matrices (symmetric or not), Theorem 2.10 implies the next corollary, valid for an arbitrary matrix H.

Corollary B.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 $\mathbb{R}^n$,*

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

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

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

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

Appendix C. Proof of Theorem 6.4

Let $X \in S^n$ have 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 $\tilde{\mathscr{A}}_{(1)}(x) = \nabla f(x) = \mathscr{A}_{(1)}(x)$ for every $x \in \mathbb{R}^n$. Suppose that for some integer s in $[1, k)$ we have

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

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

By definition, the tensor $\mathscr{A}_\sigma(x)$ is equal to zero if the permutation σ has more than one cycle in its cycle decomposition. Then using Lemma 6.3 gives

$$\begin{aligned}\sum_{\sigma \in P^{s+1}} \mathrm{Diag}^\sigma \mathscr{A}_\sigma(x) &= \sum_{\substack{\sigma \in P^s \\ l \in \mathbb{N}_{s+1}}} \mathrm{Diag}^{\sigma(1)} \mathscr{A}_{\sigma(1)}(x) \\ &= \sum_{\substack{\sigma \in P^s \\ l \in \mathbb{N}_s}} \mathrm{Diag}^{\sigma(1)} \mathscr{A}_{\sigma(1)}(x) \\ &= \sum_{\substack{\sigma \in P^s \\ l \in \mathbb{N}_s}} \mathrm{Diag}^{\sigma(1)} ((\mathscr{A}_\sigma(x))^{(l)}_{\mathrm{out}} + (T^l_\sigma(x))^{(l)}_{\mathrm{in}}).\end{aligned}$$

Let $M \in \S^n$ and suppose $\|M\| = 1$. 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

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

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 partition of the integers $\mathbb{N}_n$ into blocks is determined by the repeated eigenvalues of the matrix X, and the latter are all distinct, we have $M_{\mathrm{in}} = \mathrm{Diag}\,(\mathrm{diag}\, M)$. (Moreover, every tensor is block constant.) Thus, defining $h \in \mathbb{R}^n$ as in (23) we see that $h = \mathrm{diag}\, M$ and by (25) we have $U^{\mathrm{T}} M_{\mathrm{in}} U = \mathrm{Diag}\,(\mathrm{diag}\, M)$. on the one hand, the induction hypothesis and the first part of Theorem 2.10 give

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

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

$$\begin{aligned}
&\left(\sum_{\substack{\sigma\in P^s\\ l\in\mathbb{N}_s}} \mathrm{Diag}^{\sigma(1)}(\mathscr{A}_\sigma(x))^{(l)}_{\mathrm{out}}\right)[M]\\
&\quad= \lim_{m\to\infty}\frac{U_m(\sum_{\sigma\in P^s}\mathrm{Diag}^\sigma \mathscr{A}_\sigma(x))U_m^{\mathrm{T}} - \sum_{\sigma\in P^s}\mathrm{Diag}^\sigma \mathscr{A}_\sigma(x)}{\|M_m\|}\\
&\quad= \lim_{m\to\infty}\frac{U_m(\sum_{\sigma\in P^s}\mathrm{Diag}^\sigma \tilde{\mathscr{A}}_\sigma(x))U_m^{\mathrm{T}} - \sum_{\sigma\in P^s}\mathrm{Diag}^\sigma \tilde{\mathscr{A}}_\sigma(x)}{\|M_m\|}\\
&\quad= \left(\sum_{\substack{\sigma\in P^s\\ l\in\mathbb{N}_s}} \mathrm{Diag}^{\sigma(1)}(\tilde{\mathscr{A}}_\sigma(x))^{(l)}_{\mathrm{out}}\right)[M]\\
&\quad= \left(\sum_{\substack{\sigma\in P^s\\ l\in\mathbb{N}_s}} \mathrm{Diag}^{\sigma(1)}\tilde{\mathscr{A}}_{\sigma(1)}(x)\right)[M].
\end{aligned}$$

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

$$\left(\sum_{\sigma\in P^{s+1}} \mathrm{Diag}^\sigma \mathscr{A}_\sigma(x)\right)[M] = \left(\sum_{\sigma\in P^{s+1}} \mathrm{Diag}^\sigma \tilde{\mathscr{A}}_\sigma(x)\right)[M],$$

and since M was arbitrary, we are done.

Acknowledgments

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

References

[1] V.I. Arnold, On matrices depending on parameters, Russian Math. Surveys 26 (1971) 29–43.
[2] R. Bhatia, Matrix Analysis, Springer-Verlag, NY, 1997.
[3] X. Chen, H.-D. Qi, P. Tseng, Analysis of nonsmooth symmetric matrix functions with applications to semidifinite programming, SIAM J. Optim. 13 (2003) 960–985.
[4] Yu.L. Daleckiĭ, Integration and differentiation of functions of Hermitian operators depending on a parameter, Amer. Math. Soc. Transl. Ser. 2 16 (1960) 396–400.
[5] Yu.L. Daleckiĭ, S.G. Kreĭn, Formulas of differentiation according to a parameter of functions of Hermitian operators, Dokl. Akad. Nauk SSSR (N.S.) 76 (1951) 13–16.
[6] W.F. Donoghue, Monotone Matrix Functions and Analytic Continuation, Die Grundlehren der mathematischen Wissenschaften, Band 207, Springer-Verlag, New York–Heidelberg, 1974.
[7] J.-B. Hiriart-Urruty, D. Ye, Sensitivity analysis of all eigenvalues of a symmetric matrix, Numer. Math. 70 (1992) 45–72.
[8] R.A. Horn, C.R. Johnson, Topics in Matrix Analysis, Cambridge University Press, 1994 (first paperback edition with corrections).

[9] T. Kato, Perturbation Theory for Linear Operators, Grundlehren der Mathematischen Wissenschaften Band 132, second ed., Springer-Verlag, New York, Berlin, 1976.
[10] E.C. Kemble, The Fundamental Principles of Quantum Mechanics, Dover, New York, 1958.
[11] F. Kraus, Über Konvexe Matrixfunctionen, Math. Z. 41 (1) (1936) 18–42.
[12] A.S. Lewis, Convex analysis on the Hermitian matrices, SIAM J. Optim. 6 (1996) 164–177.
[13] A.S. Lewis, Derivatives of spectral functions, Math. Oper. Res. 21 (1996) 576–588.
[14] A.S. Lewis, Nonsmooth analysis of eigenvalues, Math. Program. 84 (1999) 1–24.
[15] A.S. Lewis, The mathematics of eigenvalue optimization, Math. Program. Ser. B 97 (1–2) (2003) 155–176.
[16] A.S. Lewis, M.L. Overton, Eigenvalue optimization, Acta Numer. 5 (1996) 149–190.
[17] A.S. Lewis, H.S. Sendov, Twice differentiable spectral functions, SIAM J. Matrix Anal. Appl. 23 (2) (2001) 368–386.
[18] A.S. Lewis, H.S. Sendov, Quadratic expansions of spectral functions, Linear Algebra Appl. 340 (2002) 97–121.
[19] K. Löwner, Über Monotone Matrixfunctionen, Math. Z. 38 (1) (1934) 177–216.
[20] J. Malick, H.S. Sendov, Clarke generalized Jacobian of the projection onto the cone of positive semidefinite matrices, Set-Valued Anal. 14 (3) (2006) 273–293.
[21] R. Mathias, A chain rule for matrix functions and applications, SIAM J. Matrix Anal. Appl. 17 (3) (1996) 610–620.
[22] H.S. Sendov, Generalized Hadamard product and the derivatives of spectral functions, SIAM J. Matrix Anal. Appl. 28 (3) (2006) 667–681.
[23] D. Sun, J. Sun, Semismooth matrix-valued functions, Math. Oper. Res. 27 (1) (2002) 150–169.
[24] N.-K. Tsing, M.K.H. Fan, E.I. Verriest, On analyticity of functions involving eigenvalues, Linear Algebra Appl. 207 (1994) 159–180.