Quadratic Expansions of Spectral Functions

Adrian S. Lewis*and Hristo S. Sendov†

June 22, 2000

Abstract

A function, F, on the space of $n \times n$ real symmetric matrices is called *spectral* if it depends only on the eigenvalues of its argument, that is $F(A) = F(UAU^T)$ for every orthogonal U and symmetric A in its domain. Spectral functions are in one-to-one correspondence with the symmetric functions on $\mathbb{R}^n$: those that are invariant under arbitrary swapping of their arguments. In this paper we show that a spectral function has a *quadratic expansion* around a point A if and only if its corresponding symmetric function has quadratic expansion around $\lambda(A)$ (the vector of eigenvalues). We also give a concise and easy to use formula for the 'Hessian' of the spectral function. In the case of convex functions we show that a positive definite 'Hessian' of f implies positive definiteness of the 'Hessian' of F.

1 Introduction

In this work we investigate a property of functions F on the real vector space of symmetric matrices that are *orthogonally invariant*:

$$F(U^T AU) = F(A), \text{ for all } A \text{ symmetric and } U \text{ orthogonal.}$$

*Department of Combinatorics & Optimization, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. Email: aslewis@math.uwaterloo.ca. Research supported by NSERC.

†Department of Combinatorics & Optimization, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. Email: hssendov@barrow.uwaterloo.ca. Research supported by NSERC.

Every such function can be decomposed as

$$F(A) = (f \circ \lambda)(A),$$

where λ is the map that gives the eigenvalues of the matrix A and f is a permutation invariant function. (See the next section for more details.) We call such functions F *spectral functions* (or just functions of eigenvalues) because de facto they depend only on the spectrum of the operator A.

In the past, such functions have been of interest for example to people working in the field of quantum mechanics [7], [14]. With developments in semidefinite programming, functions on eigenvalues became an inseparable part of mathematical programming. Optimization problems now often involve spectral functions like $\log\det(A)$, the largest eigenvalue of the matrix argument A, or the constraint that A must be positive definite and so on. Remarkably, many properties of the function f are inherited by the spectral function F. For example, this holds for differentiability and convexity [8], various types of generalized differentiability [9], analyticity [19], various second-order properties [18], [17], [16], and so on. Second-order properties of matrix functions are of great interest for optimization because the application of Newton's method and recent interior point methods [11] require that we know the second-order behaviour of the functions involved in the mathematical model.

The standard reference for the behaviour of the eigenvalues of a matrix subject to perturbations in a particular *direction* is [6]. Second-order properties of eigenvalue value functions in a particular direction are derived in [18]. What interests us in this paper is a second-order property of spectral functions subject to perturbation by an arbitrary *matrix.* Analytical properties subject to matrix perturbations are discussed in [19]. In some sense our result about spectral functions having quadratic expansions lies between the results in [8] and the results in [19]. In a parallel paper [10] we show that F is twice differentiable if and only if f is, and also that $F \in C^2$ if and only if $f \in C^2$. Having a quadratic expansion is a property that many functions possess. For example a theorem of Alexandrov [1] states that every finite, convex function on an open subset of $\mathbb{R}^n$ has quadratic expansion at almost every point. Also, it is not necessary for a function to be twice differentiable in order to have quadratic expansion. For example the function

$$f(x) = \begin{cases} x^3 \sin(1/x), & \text{if } x \neq 0 \\ 0, & \text{if } x = 0 \end{cases}$$

has quadratic expansion around $x = 0$ but is not twice differentiable there. While on the other hand being twice differentiable at x implies quadratic expansion at x.

2 Notation and definitions

Let S^n be the Euclidean space of all $n \times n$ symmetric matrices with inner product $\langle A, B\rangle = \mathrm{tr}\,(AB)$, and for $A \in S^n$ denote by $\lambda(A)^T = (\lambda_1(A), ..., \lambda_n(A))$ the vector of its eigenvalues ordered in nonincreasing order. (All vectors in this paper are assumed to be column vectors unless stated otherwise.) For any vector x in $\mathbb{R}^n$, $\mathrm{Diag}\, x$ will denote the diagonal matrix with the vector x on the main diagonal, and $\bar{x}$ will denote the vector with the same entries as x ordered in nonincreasing order, that is $\bar{x}_1 \geq \bar{x}_2 \geq \cdots \geq \bar{x}_n$. Let $\mathbb{R}^n_\downarrow$ denote the convex cone of all vectors x in $\mathbb{R}^n$ such that $x_1 \geq x_2 \geq \cdots \geq x_n$. The following definition explains the property that interests us in this paper.

Definition 2.1 *We say that a function $f : \mathbb{R}^n \to \mathbb{R}$ has a* **weak quadratic expansion** *at the point x if there exists a vector $\nabla f(x)$ and a symmetric matrix $\nabla^2 f(x)$ such that for small $h \in \mathbb{R}^n$*

$$f(x+h) = f(x) + \langle \nabla f(x), h\rangle + \frac{1}{2}\langle h, \nabla^2 f(x) h\rangle + o(\|h\|^2),$$

and a **strong quadratic expansion** *at the point x if*

$$f(x+h) = f(x) + \langle \nabla f(x), h\rangle + \frac{1}{2}\langle h, \nabla^2 f(x) h\rangle + O(\|h\|^3).$$

The vector h is called a **perturbation vector***.*

A few comments on this definition are necessary. Clearly having strong quadratic expansion implies the weak quadratic expansion. We want to alert the reader that a function may not be twice differentiable at the point x but still possesses a strong quadratic expansion at that point. (See the example at the end of the Introduction.) It is clear that if the function has quadratic expansion at the point x then it is differentiable at x and its gradient is the vector $\nabla f(x)$ from the above definition. If the function has weak quadratic expansion, then there is a unique vector $\nabla f(x)$ and a unique symmetric matrix $\nabla^2 f(x)$ (*Hessian*) for which the expansion holds. There is a slight abuse of notation when we call $\nabla^2 f(x)$ the Hessian of f, but no danger of confusion exists because when f is in C^2 around x the symmetric matrix $\nabla^2 f(x)$ is exactly the Hessian. Finally, another way to write the quadratic expansion of a function f, consistent with [11], is

$$(1) \qquad f(x+h) = f(x) + \nabla f(x)[h] + \frac{1}{2}\nabla^2 f(x)[h,h] + O(\|h\|^3).$$

We give some less common notation which will be used throughout the paper. It is taken directly from [18]. We are interested in quadratic expansions of matrix functions $f \circ \lambda$ around a matrix A. (In all of our preliminary results the matrix A will be a diagonal matrix, $\mathrm{Diag}\,\mu$.) Let $H \in S^n$ be the perturbation matrix. Fix a number $m \in \mathbb{N}$, $1 \leq m \leq n$ and let the "block structure" of the vector $\lambda(A)$ be given by

$$\lambda_1(A) = \cdots = \lambda_{k_1}(A) > \cdots > \lambda_{k_{l-1}+1}(A) = \cdots = \lambda_m(A) = \cdots = \lambda_{k_l}(A)$$
$$> \cdots \lambda_{k_r}(A), \quad (k_0 = 0,\ k_r = n).$$

That is, the eigenvalue $\lambda_m(A)$ lies in the l'th block of equal eigenvalues. Let $X = [x^1, ..., x^n]$ be an orthogonal matrix such that $X^T A X = \mathrm{Diag}\,\lambda(A)$ (so x^i is a unit eigenvector corresponding to $\lambda_i(A)$) and let

$$X_l = [x^{k_{l-1}+1}, ..., x^{k_l}].$$

Let $U_l = [v^1, ..., v^{k_l - k_{l-1}}]$ be a $(k_l - k_{l-1}) \times (k_l - k_{l-1})$ orthogonal matrix such that

$$U_l^T (X_l^T H X_l) U_l = \mathrm{Diag}\,\lambda(X_l^T H X_l).$$

Set $H_l := X_l^T H X_l$, $1 \leq l \leq r$, and suppose

$$\lambda_1(H_l) = \cdots = \lambda_{t_{l,1}}(H_l) > \cdots > \lambda_{t_{l,j-1}+1}(H_l) = \cdots \lambda_{m-k_{l-1}}(H_l) \cdots$$
$$= \lambda_{t_{l,j}}(H_l) > \cdots \lambda_{t_{l,s_l}}(H_l), \quad (t_{l,0} = 1,\ t_{l,s_l} = k_l - k_{l-1})$$

Finally let

$$U_{l,j} = [v^{t_{l,j-1}+1}, ..., v^{t_{l,j}}].$$

We should point out that $X_l = X_l(A, m)$, and $U_{l,j} = U_{l,j}(A, H, X, m)$ but from now on we will write only X_l and $U_{l,j}$ to simplify the notation.

By $A^\dagger$ we denote the Moore-Penrose generalized inverse of the matrix A. For more information on the topic see [15, p.102]. But for our needs, because we will be working only with symmetric matrices, the concept can be quickly explained. First, $(\mathrm{Diag}\,x)^\dagger_{i,j}$ is equal to $1/x_i$ if $i = j$ and $x_i \neq 0$, and is 0 otherwise. Second, for any orthogonal matrix U, that diagonalizes A, we have $A^\dagger = (U \mathrm{Diag}\,\lambda(A) U^T)^\dagger := U(\mathrm{Diag}\,\lambda(A))^\dagger U^T$.

3 Supporting results

Let A be in S^n and its eigenvalues have the following block structure

$$\lambda_1(A) = \cdots = \lambda_{k_1}(A) > \lambda_{k_1+1}(A) = \cdots = \lambda_{k_2}(A) > \lambda_{k_2+1}(A) \cdots\cdots \lambda_{k_r}(A),$$

where $k_r = n$. All our results rest on the fact that for every block $l = 1, ..., r$, the following two functions have quadratic expansion at A

$$\sigma_{k_l}(\cdot) = \sum_{i=1}^{k_l} \lambda_i(\cdot)$$
$$S_l(\cdot) = \sum_{i=k_{l-1}+1}^{k_l} \lambda_i^2(\cdot).$$

We are going to give three justifications of this fact and two of them will show that these functions are even analytic at A. For every index $m = 1, ..., n$ and every block $l = 1, ..., r$ define the functions

$$f_m(x) = \sum_{i=1}^{m} \bar{x}_i$$
$$s_l(x) = \sum_{i=k_{l-1}+1}^{k_l} \bar{x}_i^2.$$

The function f_m is the sum of the m largest entries in x. The functions f_m and $s_l(x)$ are symmetric. (A function f is *symmetric* if $f(x) = f(Px)$ for any permutation matrix P. We denote the set of all $n \times n$ permutation matrices with $P(n)$.) It is clear that if the point x is such that $\bar{x}_m > \bar{x}_{m+1}$ then f_m is linear near x. In particular, for points x near $\lambda(A)$ the functions $f_{k_l}(x)$ and $s_l(x)$ are both polynomials in the entries of x. Notice also that

$$\sigma_{k_l}(\cdot) = (f_{k_l} \circ \lambda)(\cdot)$$
$$S_l(\cdot) = (s_l \circ \lambda)(\cdot).$$

The first justification comes from the following result in [10, Theorem 3.3].

Theorem 3.1 *The symmetric function $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable at the point $\lambda(A)$ if and only if $f \circ \lambda$ is twice differentiable at the point A.* ■

The second justification is from [19, Theorem 2.1].

Theorem 3.2 *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a function analytic at the point $\lambda(A)$ for some A in S^n. Let also $f(Px) = f(x)$ for every permutation matrix, P, for which $P\lambda(A) = \lambda(A)$. Then the function $f \circ \lambda$ is analytic at A.* ■

For the third justification we use the standard algebraic fact that every symmetric polynomial in several variables can be written as a polynomial in the elementary symmetric functions. We also use the following result [2]. Until the end of this section only, $\lambda_i(X)$ will denote an arbitrary eigenvalue of a matrix X, not necessarily the i'th largest one.

Theorem 3.3 (Arnold 1971) *Suppose that $A \in \mathbb{C}^{n\times n}$ has q eigenvalues $\lambda_1(A), ..., \lambda_q(A)$ (counting multiplicities) in an open set $\Omega \subset \mathbb{C}$, and the rest $n-q$ eigenvalues not in Ω. Then for all matrices X in a neighbourhood of A there are holomorphic mappings $S : \mathbb{C}^{n\times n} \mapsto \mathbb{C}^{q\times q}$ and $T : \mathbb{C}^{n\times n} \mapsto \mathbb{C}^{(n-q)\times(n-q)}$ such that*

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

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

Using Arnold's theorem we can prove that in fact the functions σ_{k_l} and S_l are holomorphic around A.

Theorem 3.4 *For every symmetric polynomial $p : \mathbb{C}^q \mapsto \mathbb{C}$, the function $(p \circ \lambda)(S(X))$ is holomorphic around A.*

Proof. It suffices to prove the theorem in the case of an elementary symmetric polynomial. By continuity of the eigenvalues, for every $i = 1, ..., n$ we can define functions $\lambda_i : C^{n\times n} \mapsto \mathbb{C}$ such that for matrices X near A, $\{\lambda_i(X)\}_{i=1}^n$ are the eigenvalues of X, $\{\lambda_i(X)\}_{i=1}^q$ are the eigenvalues of $S(X)$. So the elementary symmetric functions of $\lambda_1(X), ..., \lambda_q(X)$ are the coefficients of the characteristic polynomial $\det(\lambda I - S(X))$. Consequently they are holomorphic around A. ■

4 Quadratic expansion of spectral functions

Our goal in this section is to prove the main result of the paper.

Theorem 4.1 (Quadratic Expansion) *The symmetric function $f : \mathbb{R}^n \to \mathbb{R}$ has a strong quadratic expansion at the point $x = \lambda(Y)$ ($Y \in S^n$) if and only if $f \circ \lambda$ has a strong quadratic expansion at Y, with*

$$\begin{aligned}
\nabla(f \circ \lambda)(Y)[H] &= \operatorname{tr}\left(\tilde{H} \operatorname{Diag} \nabla f(\mu)\right) \\
\nabla^2(f \circ \lambda)(Y)[H,H] &= \sum_{p,q=1}^{n} \tilde{h}_{pp} f''_{pq}(\mu) \tilde{h}_{qq} + \\
&\quad \sum_{\substack{p \neq q \\ \mu_p = \mu_q}} b_p \tilde{h}^2_{pq} + \sum_{p,q: \mu_p \neq \mu_q} \frac{f'_p(\mu) - f'_q(\mu)}{\mu_p - \mu_q} \tilde{h}^2_{pq},
\end{aligned}$$

Where

$$\begin{aligned}
&\mu = \lambda(Y) \\
&\tilde{H} = U^T H U \\
&Y = U(\operatorname{Diag} \mu) U^T \\
&b \text{ – as defined in Lemma 4.8.}
\end{aligned}$$

The same statement holds for the weak quadratic expansion. ■

We will only talk about strong quadratic expansions in this paper: the development for the weak version is analogous. We need the following result from [18, Remark 6].

Lemma 4.2 *Every eigenvalue, $\lambda_m(Y)$, of a symmetric matrix, Y, has the following expansion in the direction of the symmetric matrix H:*

$$\begin{aligned}
\lambda_m(Y + tH) &= \lambda_m(Y) + t\lambda_{m-k_{l-1}}(X_l^T H X_l) \\
&\quad + \frac{t^2}{2} \lambda_{m-k_{l-1}-t_{l,j-1}}\left(2U_{l,j}^T X_l^T H (\lambda_m(Y)I - Y)^{\dagger} H X_l U_{l,j}\right) + O(t^3),
\end{aligned} \tag{2}$$

where the meaning of X_l and $U_{l,j}$ is explained in the previous section. ■

Definition 4.3 (Lewis, [9]) *We say that vector $\mu \in \mathbb{R}^n$* **block refines** *the vector $b \in \mathbb{R}^n$ if $\mu_i = \mu_j$ implies $b_i = b_j$ for all $i,j \in \{1, ..., n\}$. Equivalently*

$$P\mu = \mu \ \Rightarrow \ Pb = b \quad \textit{for all } P \in P(n).$$

Next we give a technical lemma that will allow us to cut on the notation and skipped computations.

Lemma 4.4 *Let $\mu \in \mathbb{R}^n$ be such that*

$$\mu_1 = \cdots = \mu_{k_1} > \mu_{k_1+1} = \cdots = \mu_{k_2} > \mu_{k_2+1} \cdots \mu_{k_r}, \qquad (k_0 = 0,\ k_r = n),$$

and let vector $b \in \mathbb{R}^n$ be block refined by μ. Let $H \in S^n$ be an arbitrary matrix and $X_i[e^{k_{i-1}+1}, ..., e^{k_i}]$ for every $i = 1, ..., r$. Then we have the identities:

$$\langle H, \sum_{i=1}^{r} b_{k_i}(\mu_{k_i} I - \operatorname{Diag}\mu)^\dagger H X_i X_i^T \rangle = \sum_{p,q:\mu_p>\mu_q} \frac{b_p - b_q}{\mu_p - \mu_q} h_{qp}^2.$$

$$\langle H, \sum_{i=1}^{l} b_{k_i}(\mu_{k_i} I - \operatorname{Diag}\mu)^\dagger H X_i X_i^T \rangle = \sum_{p=1}^{k_l} \sum_{\substack{q=1 \\ \mu_q \neq \mu_p}}^{n} \frac{b_p}{\mu_p - \mu_q} h_{qp}^2$$

$$\langle H, b_{k_l}(\mu_{k_l} I - \operatorname{Diag}\mu)^\dagger H X_l X_l^T \rangle = \sum_{p=k_{l-1}+1}^{k_l} \sum_{\substack{q=1 \\ \mu_q \neq \mu_p}}^{n} \frac{b_p}{\mu_p - \mu_q} h_{qp}^2.$$

Proof. The product $X_i X_i^T$ is an $n \times n$ matrix with zero entries, except $(X_l X_l^T)^{p,p} = 1$ for $p = k_{i-1} + 1, ..., k_i$. That is why the columns of $H X_i X_i^T$ are zero vectors, except the columns with indexes $p = k_{i-1} + 1, ..., k_i$ which are equal to the corresponding columns of H. The matrix $b_{k_i}(\mu_{k_i} I - \operatorname{Diag}\mu)^\dagger$ is equal to

$$\operatorname{Diag}\left(\frac{b_{k_i}}{\mu_{k_i} - \mu_1}, ..., \frac{b_{k_i}}{\mu_{k_i} - \mu_{k_{i-1}}}, 0, ..., 0, \frac{b_{k_i}}{\mu_{k_i} - \mu_{k_i+1}}, ..., \frac{b_{k_i}}{\mu_{k_i} - \mu_{k_r}} \right).$$

Consequently we have

$$\langle H, \sum_{i=1}^{r} b_{k_i}(\mu_{k_i} I - \operatorname{Diag}\mu)^\dagger H X_i X_i^T \rangle = \sum_{i=1}^{r} \sum_{p=k_{i-1}+1}^{k_i} \sum_{\substack{q=1 \\ \mu_q \neq \mu_p}}^{n} \frac{b_{k_i}}{\mu_{k_i} - \mu_q} h_{qp}^2$$

$$= \sum_{i=1}^{r} \sum_{p=k_{i-1}+1}^{k_i} \left(\sum_{q=1}^{k_{i-1}} \frac{-b_{k_i}}{\mu_q - \mu_{k_i}} h_{qp}^2 + \sum_{q=k_i+1}^{n} \frac{b_{k_i}}{\mu_{k_i} - \mu_q} h_{qp}^2 \right)$$

$$= \sum_{i=1}^{r} \sum_{p=k_{i-1}+1}^{k_i} \left(\sum_{q=1}^{k_{i-1}} \frac{-b_p}{\mu_q - \mu_p} h_{qp}^2 + \sum_{q=k_i+1}^{n} \frac{b_p}{\mu_p - \mu_q} h_{qp}^2 \right)$$

$$= \sum_{p,q:\mu_p > \mu_q} \frac{b_p - b_q}{\mu_p - \mu_q} h_{pq}^2.$$

The other two identities can now be easily obtained as well. ■

Our first goal is to find a formula for the Hessian of σ_{k_l}, $1 \leq l \leq r$. We denote the standard basis in $\mathbb{R}^n$ by e^1,e^2,...,e^n. As a byproduct in the following lemma we derive a formula for the derivative of the function σ_{k_l} at the point $\operatorname{Diag}\mu$. This formula appeared many times in the literature: see for example Corollary 3.10 in [5], or the proof of Corollary 3.3 in [8], or formula (3.28) in [12].

Lemma 4.5 *For a real vector $\mu \in \mathbb{R}^n$, such that*

$$\mu_1 = \cdots = \mu_{k_1} > \mu_{k_1+1} = \cdots = \mu_{k_2} > \mu_{k_2+1} \cdots \mu_{k_r}, \qquad (k_0 = 0,\ k_r = n),$$

the function

$$\sigma_{k_l}(\cdot) = \sum_{i=1}^{k_l} \lambda_i(\cdot)$$

is analytic at the matrix $\operatorname{Diag}\mu$ *with first and second derivatives satisfying*

$$\begin{aligned} \nabla \sigma_{k_l}(\operatorname{Diag}\mu)[H] &= \operatorname{tr}\left(H \operatorname{Diag} \sum_{i=1}^{k_l} e^i \right) \\ \nabla^2 \sigma_{k_l}(\operatorname{Diag}\mu)[H,H] &= 2 \sum_{p=1}^{k_l} \sum_{q=k_l+1}^{n} \frac{h_{qp}^2}{\mu_p - \mu_q} \\ &= \operatorname{tr}\left(2H \sum_{i=1}^{l} (\mu_{k_i} I - \operatorname{Diag}\mu)^{\dagger} H X_i X_i^T \right), \end{aligned}$$

where $X_i = [e^{k_{i-1}+1}, ..., e^{k_i}]$.

Proof. The fact that σ_{k_l} is analytic at the point $\operatorname{Diag}\mu$ follows from Section 3. Next, summing equations (2) with $A = \operatorname{Diag}\mu$, for $m = 1, ..., k_l$ and using the fact that $X = I$ (so $X_i = [e^{k_{i-1}+1}, ..., e^{k_i}]$), we get

$$\sigma_{k_l}(\operatorname{Diag}\mu + tH) = \sum_{i=1}^{k_l} \lambda_i(\operatorname{Diag}\mu + tH) = \sigma_{k_l}(\operatorname{Diag}\mu) + t\sum_{i=1}^{l} \operatorname{tr}(X_i^T H X_i)$$
$$+\frac{t^2}{2}\sum_{i=1}^{l}\sum_{j=1}^{s_i}\sum_{v=1}^{t_{i,j}-t_{i,j-1}} \lambda_v\big(2U_{i,j}^T X_i^T H\big(\mu_{k_i} I - \operatorname{Diag}\mu\big)^{\dagger} H X_i U_{i,j}\big) + O(t^3)$$
$$= \sigma_{k_l}(\operatorname{Diag}\mu) + t\langle \operatorname{Diag}\sum_{i=1}^{k_l} e^i, H\rangle$$
$$+\frac{t^2}{2}\sum_{i=1}^{l}\sum_{j=1}^{s_i} \operatorname{tr}\big(2U_{i,j}^T X_i^T H(\mu_{k_i} I - \operatorname{Diag}\mu)^{\dagger} H X_i U_{i,j}\big) + O(t^3).$$

We concentrate on the double sum above.

$$\sum_{i=1}^{l}\sum_{j=1}^{s_i} \operatorname{tr}\big(2U_{i,j}^T X_i^T H(\mu_{k_i} I - \operatorname{Diag}\mu)^{\dagger} H X_i U_{i,j}\big) =$$
$$= \sum_{i=1}^{l}\sum_{j=1}^{s_i} \operatorname{tr}\big(2X_i^T H(\mu_{k_i} I - \operatorname{Diag}\mu)^{\dagger} H X_i U_{i,j} U_{i,j}^T\big)$$
$$= \sum_{i=1}^{l} \operatorname{tr}\left(2X_i^T H(\mu_{k_i} I - \operatorname{Diag}\mu)^{\dagger} H X_i \sum_{j=1}^{s_i} U_{i,j} U_{i,j}^T\right)$$
$$= \sum_{i=1}^{l} \operatorname{tr}\big(2X_i^T H(\mu_{k_i} I - \operatorname{Diag}\mu)^{\dagger} H X_i\big)$$
$$= \operatorname{tr}\left(2H\sum_{i=1}^{l}(\mu_{k_i} I - \operatorname{Diag}\mu)^{\dagger} H X_i X_i^T\right)$$
$$= \sum_{p=1}^{k_l}\sum_{\substack{q=1\\ \mu_q\neq\mu_p}}^{n} \frac{2}{\mu_p - \mu_q} h_{qp}^2$$
$$= 2\sum_{p=1}^{k_l}\sum_{q=k_l+1}^{n} \frac{h_{qp}^2}{\mu_p - \mu_q}.$$

The next to the last equality follows from Lemma 4.4, with $b = (2, ..., 2)$, while the last equality after canceling all terms with opposite signs. By the uniqueness of the Hessian in the quadratic expansion of a function, we conclude that the last expression above must be indeed the Hessian. ■

Note 4.6 *Notice that the Hessian above is positive definite quadratic form. This is not a surprise since a well known result of Fan [4] says that σ_m is a convex function for all $m = 1, ..., n$.*

Lemma 4.7 *For a real vector $\mu \in \mathbb{R}^n$, such that*

$$\mu_1 = \cdots = \mu_{k_1} > \mu_{k_1+1} = \cdots = \mu_{k_2} > \mu_{k_2+1} \cdots \mu_{k_r} \qquad (k_0 = 1,\ k_r = n),$$

the function

$$S_l(\cdot) = \sum_{m=k_{l-1}+1}^{k_l} \lambda_m^2(\cdot)$$

is analytic with first and second derivatives at the matrix $\mathrm{Diag}\,\mu$*, satisfying*

$$\nabla S_l(\mathrm{Diag}\,\mu)[H] = 2\mu_{k_l} \mathrm{tr}\left(H\,\mathrm{Diag} \sum_{i=k_{l-1}+1}^{k_l} e^i\right)$$

$$\begin{aligned}\nabla^2 S_l(\mathrm{Diag}\,\mu)[H,H] &= 2 \sum_{p,q=k_{l-1}+1}^{k_l} h_{qp}^2 + 4 \sum_{p=k_{l-1}+1}^{k_l} \sum_{\substack{q=1 \\ \mu_p \neq \mu_q}}^{n} \frac{\mu_p}{\mu_p - \mu_q} h_{qp}^2 \\ &= \langle H, 2X_l X_l^T H X_l X_l^T + 4\mu_{k_l}(\mu_{k_l} I - \mathrm{Diag}\,\mu)^\dagger H X_l X_l^T \rangle,\end{aligned}$$

where $X_l = [e^{k_{l-1}+1}, ..., e^{k_l}]$.

Proof. The analyticity of $S(\cdot)$ at the point $\mathrm{Diag}\,\mu$ follows from Section 3. Next, summing the squares of equations (2) with $A = \mathrm{Diag}\,\mu$, for $m = 1, ..., k_l$ and using the fact that $X = I$ (so $X_i = [e^{k_{i-1}+1}, ..., e^{k_i}]$), we get

$$\begin{aligned}\sum_{m=k_{l-1}+1}^{k_l} \lambda_m^2(\mathrm{Diag}\,\mu + tH) &= \sum_{m=k_{l-1}+1}^{k_l} \Bigg(\mu_{k_l} + t\lambda_{m-k_{l-1}}(X_l^T H X_l) \\ &\quad + \frac{t^2}{2}\lambda_{m-k_{l-1}-t_{l,j-1}}\left(2U_{l,j}^T X_l^T H(\mu_{k_l} I - \mathrm{Diag}\,\mu)^\dagger H X_l U_{l,j}\right) + O(t^3)\Bigg)^2 \\ &= (k_l - k_{l-1})\mu_{k_l}^2 + t^2 \sum_{m=k_{l-1}+1}^{k_l} \lambda_{m-k_{l-1}}^2(X_l^T H X_l)\end{aligned}$$

$$+ 2t\mu_{k_l} \sum_{m=k_{l-1}+1}^{k_l} \lambda_{m-k_{l-1}}(X_l^T H X_l)$$

$$+ t^2\mu_{k_l} \sum_{j=1}^{s_l} \sum_{v=1}^{t_{l,j}-t_{l,j-1}} \lambda_v\big(2U_{l,j}^T X_l^T H(\mu_{k_l} I - \operatorname{Diag}\mu)^\dagger H X_l U_{l,j}\big) + O(t^3).$$

We recall the fact that for every symmetric $n \times n$ matrix Q we have

$$\sum_{i=1}^{n} \lambda_i^2(Q) = \langle Q, Q \rangle.$$

We use this fact to evaluate the second summand in the formula above.

$$\sum_{m=k_{l-1}+1}^{k_l} \lambda_{m-k_{l-1}}^2(X_l^T H X_l) = \langle X_l^T H X_l, X_l^T H X_l \rangle = \langle H, X_l X_l^T H X_l X_l^T \rangle.$$

Observe as in Lemma 4.5 that for the fourth summand in the formula above we have

$$\sum_{j=1}^{s_l} \sum_{v=1}^{t_{l,j}-t_{l,j-1}} \lambda_v\big(2U_{l,j}^T X_l^T H(\mu_{k_l} I - \operatorname{Diag}\mu)^\dagger H X_l U_{l,j}\big)$$

$$= \sum_{j=1}^{s_l} \operatorname{tr}\big(2U_{l,j}^T X_l^T H(\mu_{k_l} I - \operatorname{Diag}\mu)^\dagger H X_l U_{l,j}\big)$$

$$= \operatorname{tr}\big(2X_l^T H(\mu_{k_l} I - \operatorname{Diag}\mu)^\dagger H X_l\big).$$

Substituting everything in the original formula we get

$$\sum_{m=k_{l-1}+1}^{k_l} \lambda_m^2(\operatorname{Diag}\mu + tH) = (k_l - k_{l-1})\mu_{k_l}^2 + t^2\langle H, X_l X_l^T H X_l X_l^T \rangle +$$

$$2t\mu_{k_l}\langle \operatorname{Diag} \sum_{i=k_{l-1}+1}^{k_l} e^i, H \rangle + t^2\mu_{k_l}\langle H, 2(\mu_{k_l} I - \operatorname{Diag}\mu)^\dagger H X_l X_l^T \rangle + O(t^3)$$

$$= (k_l - k_{l-1})\mu_{k_l}^2 + 2t\mu_{k_l}\langle \operatorname{Diag} \sum_{i=k_{l-1}+1}^{k_l} e^i, H \rangle +$$

$$\frac{t^2}{2}\langle H, 2X_l X_l^T H X_l X_l^T + 4\mu_{k_l}(\mu_{k_l} I - \operatorname{Diag}\mu)^\dagger H X_l X_l^T \rangle + O(t^3).$$

Using the third identity in Lemma 4.4, with $b = 4\mu$, we conclude that

$$\nabla^2 S_l(\mathrm{Diag}\,\mu)[H,H] = 2\sum_{p,q=k_{l-1}+1}^{k_l} h_{qp}^2 + 4\sum_{p=k_{l-1}+1}^{k_l}\sum_{\substack{q=1\\ \mu_p\neq\mu_q}}^{n} \frac{\mu_p}{\mu_p-\mu_q} h_{qp}^2.$$

By the uniqueness of the Hessian in the quadratic expansion of a function, we conclude that the last expression above must be indeed the Hessian. ■

Lemma 4.8 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a symmetric function having quadratic expansion at the point $\mu \in \mathbb{R}^n_{\downarrow}$, where*

$$\mu_1 = \cdots = \mu_{k_1} > \mu_{k_1+1} = \cdots = \mu_{k_2} > \mu_{k_2+1} \cdots \mu_{k_r}, \qquad (k_0 = 1,\ k_r = n).$$

Then we can write

$$\nabla^2 f(\mu) = \begin{pmatrix} a_{11}E_{11} + b_{k_1}I_1 & a_{12}E_{12} & \cdots & a_{1r}E_{1r} \\ a_{21}E_{21} & a_{22}E_{22} + b_{k_2}I_2 & \cdots & a_{2r}E_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{r1}E_{r1} & a_{r2}E_{r2} & \cdots & a_{rr}R_{rr} + b_{k_r}I_r \end{pmatrix},$$

where each E_{uv} is a matrix of all ones with dimensions $(k_u - k_{u-1}) \times (k_v - k_{v-1})$, $(a_{ij})_{i,j=1}^{r,r}$ is a real symmetric matrix, $b := (b_1, ..., b_n)$ is a real vector which is block refined by μ, and I_u is a square identity matrix of the same dimensions as E_{uu}. We also define the following matrix

$$A := \nabla^2 f(\mu) - \mathrm{Diag}\,(b_{k_1}I_1, ..., b_{k_r}I_r) = (a_{ij}E_{ij})_{i,j=1}^r.$$

Before we give the proof, some comments about the above representation are necessary.

(i) We make the convention that if the i-th diagonal block in the above representation has dimensions 1×1 then we set $a_{ii} = 0$ and $b_{k_i} = f''_{k_ik_i}(\mu)$. Otherwise the value of b_{k_i} is uniquely determined as the difference between a diagonal and an off-diagonal element of this block.

(ii) Note that the matrix A as well as the vector b depend on the point around which we form the quadratic expansion (in this case μ) and on the function f.

Proof. We have

$$f(\mu + h) = f(\mu) + \langle \nabla f(\mu), h\rangle + \frac{1}{2}\langle h, \nabla^2 f(\mu) h\rangle + O(\|h\|^3).$$

Let P be a permutation matrix such that $P\mu = \mu$. Then

$$\begin{aligned} f(P(\mu + h)) &= f(\mu) + \langle \nabla f(\mu), Ph\rangle + \frac{1}{2}\langle Ph, \nabla^2 f(\mu) Ph\rangle + O(\|Ph\|^3) \\ &= f(\mu) + \langle P^T \nabla f(\mu), h\rangle + \frac{1}{2}\langle h, (P^T \nabla^2 f(\mu) P) h\rangle + O(\|h\|^3). \end{aligned}$$

Using the fact that f is symmetric gives us that $f(P(\mu + h)) = f(\mu + h)$ so $\nabla f(\mu) = P^T \nabla f(\mu)$. Subtracting the above two equalities we obtain

$$\nabla^2 f(\mu) = P^T \nabla^2 f(\mu) P, \quad \forall P \in P(n) \text{ s.t. } P\mu = \mu. \tag{3}$$

The claimed block structure of $\nabla^2 f(\mu)$ is now easy to check. ■

Note 4.9 *Observe that equation (3) holds for arbitrary $\mu \in \mathbb{R}^n$.*

Lemma 4.10 *Vector μ block refines $\nabla^2 f(\mu)\mu$.*

Proof. Suppose $P\mu = \mu$. Then using twice Equation (3) and the above note, we get

$$P\nabla^2 f(\mu)\mu \;= \nabla^2 f(\mu) P\mu = P(P^T \nabla^2 f(\mu) P)\mu = \nabla^2 f(\mu)\mu.$$ ■

Lemma 4.11 *Let $\mu \in \mathbb{R}^n_\downarrow$ be such that*

$$\mu_1 = \cdots = \mu_{k_1} > \mu_{k_1+1} = \cdots = \mu_{k_2} > \mu_{k_2+1} \cdots \mu_{k_r} \quad (k_0 = 0,\ k_r = n).$$

Suppose μ block-refines a vector $b \in \mathbb{R}^n$. Then $b^T\lambda$ has the quadratic expansion:

$$b^T \lambda(\operatorname{Diag}\mu + H) = b^T\mu + \langle \operatorname{Diag} b, H\rangle + \sum_{p,q:\mu_p > \mu_q} \frac{b_p - b_q}{\mu_p - \mu_q} h_{qp}^2.$$

Proof. Because the vector μ block-refines the vector b there exist reals $b'_1, b'_2, \ldots, b'_r$ with

$$b_j = b'_i \text{ whenever } k_{i-1} + 1 \le j \le k_i,\ i = 1, 2, \ldots, r.$$

We obtain

$$b^T\lambda(\cdot) = \sum_{i=1}^{r} b'_i \sum_{j=k_{i-1}+1}^{k_i} \lambda_j(\cdot) = \sum_{i=1}^{r} b'_i\big(\sigma_{k_i}(\cdot) - \sigma_{k_{i-1}}(\cdot)\big).$$

Now applying Lemma 4.5 gives

$$\begin{aligned}
b^T\lambda(\operatorname{Diag}\mu + H) &= \sum_{i=1}^{r} b'_i\big(\sigma_{k_i}(\operatorname{Diag}\mu + H) - \sigma_{k_{i-1}}(\operatorname{Diag}\mu + H)\big) \\
&= \sum_{i=1}^{r} b'_i\Big(\sum_{j=1}^{k_i}\mu_j + \langle \operatorname{Diag}\sum_{j=1}^{k_i} e^j, H\rangle + \langle H, \sum_{i=1}^{l}(\mu_{k_i}I - \operatorname{Diag}\mu)^\dagger H X_i X_i^T\rangle \\
&\quad - \sum_{j=1}^{k_{i-1}}\mu_j - \langle \operatorname{Diag}\sum_{j=1}^{k_{i-1}} e^j, H\rangle - \langle H, \sum_{i=1}^{l-1}(\mu_{k_i}I - \operatorname{Diag}\mu)^\dagger H X_i X_i^T\rangle\Big) + O(\|H\|^3) \\
&= \sum_{i=1}^{r} b'_i\Big(\sum_{j=k_{i-1}+1}^{k_i}\mu_j + \langle \operatorname{Diag}\sum_{j=k_{i-1}+1}^{k_i} e^j, H\rangle \\
&\qquad + \langle H, (\mu_{k_l}I - \operatorname{Diag}\mu)^\dagger H X_l X_l^T\rangle\Big) + O(\|H\|^3) \\
&= b^T\mu + \langle \operatorname{Diag} b, H\rangle + \langle H, \sum_{i=1}^{r} b_{k_i}(\mu_{k_i}I - \operatorname{Diag}\mu)^\dagger H X_i X_i^T\rangle + O(\|H\|^3) \\
&= b^T\mu + \langle \operatorname{Diag} b, H\rangle + \sum_{p,q:\mu_p>\mu_q}\frac{b_p - b_q}{\mu_p - \mu_q}h_{qp}^2.
\end{aligned}$$

The last equality above follows from Lemma 4.4. ■

Lemma 4.12 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a symmetric function having quadratic expansion at the point $\mu \in \mathbb{R}^n_{\downarrow}$, where*

$$\mu_1 = \cdots = \mu_{k_1} > \mu_{k_1+1} = \cdots = \mu_{k_2} > \mu_{k_2+1}\cdots\mu_{k_r} \quad (k_0 = 0,\ k_r = n).$$

Then the following matrix functions on S^n,

(i) $F(\cdot) := \nabla f(\mu)^T\lambda(\cdot)$,

(ii) $H(\cdot) := \mu^T\nabla^2 f(\mu)\lambda(\cdot)$,

(iii) $G(\cdot) := \lambda(\cdot)^T\nabla^2 f(\mu)\lambda(\cdot)$,

have quadratic expansions at the matrix $\operatorname{Diag}\mu$.

Proof. Later we will need the formulae giving the quadratic expansions of these functions derived in the following proof. Notice that the first two parts follow immediately from the previous two lemmas. So we can write

$$F(\operatorname{Diag}\mu + H) \approx \nabla f(\mu)^T \mu + \langle \operatorname{Diag}\nabla f(\mu), H\rangle + \sum_{p,q:\mu_p>\mu_q} \frac{f'_p(\mu) - f'_q(\mu)}{\mu_p - \mu_q} h_{qp}^2,$$

$$H(\operatorname{Diag}\mu + H) \approx \mu^T \nabla^2 f(\mu)\mu + \langle \operatorname{Diag}\nabla^2 f(\mu)\mu, H\rangle + \sum_{p,q:\mu_p>\mu_q} \frac{(\mu^T\nabla^2 f(\mu))_p - (\mu^T\nabla^2 f(\mu))_q}{\mu_p - \mu_q} h_{qp}^2.$$

(iii) Because of the block structure of $\nabla^2 f(\mu)$ described in Lemma 4.8, we have

$$\lambda(\cdot)^T \nabla^2 f(\mu)\lambda(\cdot) = \sum_{i,j=1}^{r} a_{ij}(\sigma_{k_i}(\cdot) - \sigma_{k_{i-1}}(\cdot))(\sigma_{k_j}(\cdot) - \sigma_{k_{j-1}}(\cdot)) + \sum_{l=1}^{r} b_{k_l} S_l(\cdot),$$

where the matrix $(a_{ij})_{i,j=1}^r$, vector b, and $S_l(\cdot)$ are defined in Lemma 4.8 and Lemma 4.7. Now by Lemma 4.5

$$\sigma_{k_l}(\operatorname{Diag}\mu + H) - \sigma_{k_{l-1}}(\operatorname{Diag}\mu + H) = \sum_{i=k_{l-1}+1}^{k_l} \mu_i + \langle \operatorname{Diag} \sum_{i=k_{l-1}+1}^{k_l} e^i, H\rangle + \frac{1}{2}\langle H, 2(\mu_{k_l} I - \operatorname{Diag}\mu)^\dagger H X_l X_l^T\rangle + O(\|H\|^3)$$

$$\sum_{i=k_{l-1}+1}^{k_l} \mu_i + \sum_{i=k_{l-1}+1}^{k_l} h_{ii} + \sum_{i=k_{l-1}+1}^{k_l} \langle H, (\mu_{k_l} I - \operatorname{Diag}\mu)^\dagger H e^i (e^i)^T\rangle + O(\|H\|^3).$$

We can evaluate the first summand in the above representation of the function $G(\cdot)$.

$$\sum_{i,j=1}^{r} a_{ij}(\sigma_{k_i}(\operatorname{Diag}\mu + H) - \sigma_{k_{i-1}}(\operatorname{Diag}\mu + H))$$

$$\times\,(\sigma_{k_j}(\mathrm{Diag}\,\mu + H) - \sigma_{k_{j-1}}(\mathrm{Diag}\,\mu + H))$$

$$= \mu^T A\mu + (\mathrm{diag}\,H)^T A(\mathrm{diag}\,H) + 2\mu^T A(\mathrm{diag}\,H)$$

$$+ 2\langle H, \sum_{i,j=1}^{n} \mu_i A^{ij}(\mu_j I - \mathrm{Diag}\,\mu)^{\dagger} H e^j (e^j)^T\rangle + O(\|H\|^3)$$

$$= \mu^T A\mu + 2\langle \mathrm{Diag}\,A\mu, H\rangle + \langle H, \mathrm{Diag}\,A(\mathrm{diag}\,H)\rangle$$

$$+ 2\langle H, \sum_{i,j=1}^{n} \mu_i A^{ij}(\mu_j I - \mathrm{Diag}\,\mu)^{\dagger} H e^j (e^j)^T\rangle + O(\|H\|^3),$$

where diag: $S^n \to \mathbb{R}^n$ defined by diag $(H) = (h_{11}, ..., h_{nn})$ is the conjugate operator of Diag: $\mathbb{R}^n \to S^n$. On the other hand Lemma 4.7 gives us:

$$\sum_{l=1}^{r} b_{k_l} S_l(\mathrm{Diag}\,\mu + H) = \sum_{l=1}^{r} b_{k_l}\Big((k_l - k_{l-1})\mu_{k_l}^2 + 2\mu_{k_l}\langle \mathrm{Diag} \sum_{i=k_{l-1}+1}^{k_l} e^i, H\rangle$$

$$+ \langle H, X_l X_l^T H X_l X_l^T + 2\mu_{k_l}(\mu_{k_l} I - \mathrm{Diag}\,\mu)^{\dagger} H X_l X_l^T\rangle\Big) + O(\|H\|^3)$$

$$= \mu^T(\mathrm{Diag}\,b)\mu + 2\langle \mathrm{Diag}\,(Ib)\mu, H\rangle + \langle H, \sum_{l=1}^{r} b_{k_l} X_l X_l^T H X_l X_l^T\rangle$$

$$+ \langle H, 2\sum_{i,j=1}^{n} \mu_i (Ib)^{ij}(\mu_i I - \mathrm{Diag}\,\mu)^{\dagger} H e^i (e^i)^T\rangle + O(\|H\|^3).$$

Adding these two formulae together we finally get:

$$\lambda(\mathrm{Diag}\,\mu + H)^T \nabla^2 f(\mu)\lambda(\mathrm{Diag}\,\mu + H) = \mu^T \nabla^2 f(\mu)\mu + 2\langle \mathrm{Diag}\,\nabla^2 f(\mu)\mu, H\rangle$$

$$+ \langle H, \mathrm{Diag}\,A(\mathrm{diag}\,H)\rangle + \langle H, \sum_{l=1}^{r} b_{k_l} X_l X_l^T H X_l X_l^T\rangle$$

$$+ \langle H, 2\sum_{j=1}^{n} \big(\mu^T \nabla^2 f(\mu)\big)_j (\mu_j I - \mathrm{Diag}\,\mu)^{\dagger} H e^j (e^j)^T\rangle + O(\|H\|^3)$$

$$= \mu^T \nabla^2 f(\mu)\mu + 2\langle \mathrm{Diag}\,\nabla^2 f(\mu)\mu, H\rangle + \langle H, \mathrm{Diag}\,A(\mathrm{diag}\,H)\rangle$$

$$+ \langle H, \sum_{l=1}^{r} b_{k_l} X_l X_l^T H X_l X_l^T\rangle + 2\sum_{p,q:\mu_p > \mu_q} \frac{(\mu^T \nabla^2 f(\mu))_p - (\mu^T \nabla^2 f(\mu))_q}{\mu_p - \mu_q} h_{qp}^2.$$

In the last equality we used Lemma 4.10 and Lemma 4.4. ■

Now we are ready to prove a preliminary case of Theorem 4.1, namely, that it holds at $X = \mathrm{Diag}\,\mu$, ($\mu \in \mathbb{R}_\downarrow$) and to give a formula for the Hessian of $f \circ \lambda$ at that point. The results for the gradient of $f \circ \lambda$ that we will obtain along the way were first obtained in [9].

Theorem 4.13 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a symmetric function having quadratic expansion at the point $\mu \in \mathbb{R}^n_\downarrow$, where*

$$\mu_1 = \cdots = \mu_{k_1} > \mu_{k_1+1} = \cdots = \mu_{k_2} > \mu_{k_2+1} \cdots \mu_{k_r} \quad (k_0 = 0,\ k_r = n).$$

Then $f \circ \lambda$ has quadratic expansion at the point $\mathrm{Diag}\,\mu$*, with*

$$\nabla(f \circ \lambda)(\mathrm{Diag}\,\mu)[H] = \mathrm{tr}\,(H \mathrm{Diag}\,\nabla f(\mu))$$

$$\nabla^2(f \circ \lambda)(\mathrm{Diag}\,\mu)[H,H] = \sum_{p,q=1}^{n} h_{pp} f''_{pq}(\mu) h_{qq} + \sum_{\substack{p \neq q \\ \mu_p = \mu_q}} b_p h_{pq}^2 + \sum_{p,q:\mu_p \neq \mu_q} \frac{f'_p(\mu) - f'_q(\mu)}{\mu_p - \mu_q} h_{pq}^2.$$

(With vector b defined by Lemma 4.8.)

Note 4.14 *Corollary 4.15 will show that the requirement that $\mu \in \mathbb{R}^n_\downarrow$ can be omitted. For a matrix representation of the above formula combine equation (4) below, and the first identity in Lemma 4.4.*

Proof. We are given that

$$f(x) = f(\mu) + \nabla f(\mu)^T (x - \mu) + \frac{1}{2}(x - \mu)^T \nabla^2 f(\mu)(x - \mu) + O(\|x - \mu\|^3),$$

so after letting $x = \lambda(\mathrm{Diag}\,\mu + H)$ and using the fact that

$$\lambda(\mathrm{Diag}\,\mu + H) = \lambda(\mathrm{Diag}\,\mu) + O(\|H\|)$$

we get

$$\begin{aligned}(f \circ \lambda)(\mathrm{Diag}\,\mu + H) &= f(\mu) + \nabla f(\mu)^T \lambda(\mathrm{Diag}\,\mu + H) - \nabla f(\mu)^T \mu \\ &+ \frac{1}{2}\lambda(\mathrm{Diag}\,\mu + H)^T \nabla^2 f(\mu) \lambda(\mathrm{Diag}\,\mu + H) - \mu^T \nabla^2 f(\mu) \lambda(\mathrm{Diag}\,\mu + H) \\ &+ \frac{1}{2}\mu^T \nabla^2 f(\mu)\mu + O(\|H\|^3).\end{aligned}$$

Substituting the three expressions in the proof of Lemma 4.12 we obtain

$$
\begin{aligned}
(f \circ \lambda)(\operatorname{Diag}\mu + H) &= (f \circ \lambda)(\operatorname{Diag}\mu) + \langle \operatorname{Diag}\nabla f(\mu), H\rangle \\
&\quad + \frac{1}{2}\langle H, \operatorname{Diag} A(\operatorname{diag} H) + \sum_{l=1}^{r} b_{k_l} X_l X_l^T H X_l X_l^T\rangle \qquad (4)\\
&\quad + \sum_{p,q:\mu_p>\mu_q} \frac{f_p'(\mu) - f_q'(\mu)}{\mu_p - \mu_q} h_{qp}^2 + O(\|H\|^3).
\end{aligned}
$$

Recall that $X_l = [e^{k_{l-1}+1}, ..., e^{k_l}]$. In order to obtain the representation given in the theorem one has to use the definition of A and $b = (b_1, ..., b_n)$ given in Lemma 4.8 and the note that follows it. ■

Proof of Theorem 4.1. Suppose f has quadratic expansion at the point $\lambda(Y)$, and choose any orthogonal matrix $U = [u^1 \dots u^n]$ that gives the ordered spectral decomposition of Y, $Y = U\big(\operatorname{Diag}\lambda(Y)\big)U^T$. Here we actually have $A = A(\lambda(Y))$ and $b_i = b_i(\lambda(Y))$. While in formula (4) we had $A = A(\mu)$ and $b_i = b_i(\mu)$, to make the formulae here easier to read we will write again simply A and b_i. Then we have, using Formula (4) and some easy manipulations,

$$
\begin{aligned}
(f \circ \lambda)(Y+H) &= (f \circ \lambda)(\operatorname{Diag}\lambda(Y) + U^T H U) \\
&= (f \circ \lambda)(Y) + \langle \operatorname{Diag}\nabla f(\lambda(Y)), U^T H U\rangle \\
&\quad + \frac{1}{2}\langle U^T H U, \operatorname{Diag} A(\operatorname{diag} U^T H U) + \sum_{l=1}^{r} b_{k_l} X_l X_l^T U^T H U X_l X_l^T\rangle \\
&\quad + \sum_{\substack{p,q \\ \lambda_p(Y)>\lambda_q(Y)}} \frac{f_p'(\lambda(Y)) - f_q'(\lambda(Y))}{\lambda_p(Y) - \lambda_q(Y)} ((U^T H U)^{qp})^2 + O(\|H\|^3),
\end{aligned}
$$

where $X_l = [e^{k_{l-1}+1}, ..., e^{k_l}]$. ■

Corollary 4.15 *Theorem 4.13 holds for arbitrary* $\mu \in \mathbb{R}^n$, *where*

$$
b(\mu) := P b(\bar{\mu}), \qquad (5)
$$

and P *is a permutation matrix, such that* $P^T\mu = \bar{\mu}$.

Proof. Pick a permutation matrix P such that $P^T\mu = \bar{\mu}$ and let π be the associated with it permutation, that is $\bar{\mu} = (\mu_{\pi(1)}, ..., \mu_{\pi(n)})$, or in other words $Pe^i = e^{\pi(i)}$. We have that f has quadratic expansion at the point μ, that is

$$f(\mu + h) = f(\mu) + \langle \nabla f(\mu), h\rangle + \frac{1}{2}\langle h, \nabla^2 f(\mu)h\rangle + O(\|h\|^3).$$

Using the fact that f is symmetric we obtain

$$\begin{aligned} f(\bar{\mu} + P^T h) &= f(P^T(\mu + h)) = f(\mu + h) \\ &= f(\mu) + \langle \nabla f(\mu), h\rangle + \frac{1}{2}\langle h, \nabla^2 f(\mu)h\rangle + O(\|h\|^3) \\ &= f(\bar{\mu}) + \langle P^T\nabla f(\mu), P^T h\rangle + \frac{1}{2}\langle P^T h, P^T\nabla^2 f(\mu)PP^T h\rangle + O(\|P^T h\|^3). \end{aligned}$$

So f has quadratic expansion at the point $\bar{\mu}$ as well, and we have the relationships:

$$\begin{aligned} \nabla f(\bar{\mu}) &= P^T\nabla f(\mu) \\ \nabla^2 f(\bar{\mu}) &= P^T\nabla^2 f(\mu)P. \end{aligned} \tag{6}$$

We have $\operatorname{Diag}\mu = P(\operatorname{Diag}\bar{\mu})P^T$. Applying Theorem 4.1 with $Y = \operatorname{Diag}\mu$ and $U = P$, and using Equations (6) and (5) we get

$$\begin{aligned} \nabla^2(f\circ\lambda)(\operatorname{Diag}\mu)[H,H] &= \sum_{p,q=1}^{n} h_{\pi(p)\pi(p)} f''_{pq}(\bar{\mu}) h_{\pi(q)\pi(q)} \\ &\quad + \sum_{\substack{p\neq q \\ \bar{\mu}_p=\bar{\mu}_q}} b_p(\bar{\mu}) h^2_{\pi(p)\pi(q)} + \sum_{\bar{\mu}_p\neq\bar{\mu}_q} \frac{f'_p(\bar{\mu}) - f'_q(\bar{\mu})}{\bar{\mu}_p - \bar{\mu}_q} h^2_{\pi(p)\pi(q)} \\ &= \sum_{p,q=1}^{n} h_{pp} f''_{pq}(\mu) h_{qq} + \sum_{\substack{p\neq q \\ \mu_p=\mu_q}} b_p(\mu) h^2_{pq} + \sum_{\mu_p\neq\mu_q} \frac{f'_p(\mu) - f'_q(\mu)}{\mu_p - \mu_q} h^2_{pq}. \end{aligned}$$

The invariance of the formula for the gradient is shown in a similar manner. See also [9]. ■

5 Strictly convex functions

As we mentioned in the introduction, a symmetric function f is convex if and only if $f \circ \lambda$ is convex. The analogous result also holds for for essential strict convexity []. Here we study yet a stronger property.

In this section we show that if a symmetric, convex function f has a quadratic expansion at the point $x = \lambda(Y)$ then the symmetric matrix $\nabla^2 f(x)$ is positive *definite*, if and only if the same is true for the bilinear operator $\nabla^2(f \circ \lambda)(Y)$.

Lemma 5.1 *If a function $f : \mathbb{R}^n \to \mathbb{R}$ is symmetric, strictly convex, and differentiable at the point μ*

$$\mu_1 = \cdots = \mu_{k_1} > \mu_{k_1+1} = \cdots = \mu_{k_2} > \mu_{k_2+1} \cdots \mu_{k_r}, \qquad (k_r = n).$$

then its gradient satisfies

$$\frac{f'_p(\mu) - f'_q(\mu)}{\mu_p - \mu_q} \ > \ 0 \qquad \text{for all } p,\ q \text{ such that } \mu_p \neq \mu_q.$$

Proof. Because f is strictly convex and differentiable at μ, for every $x \in \mathbb{R}^n$ $(\mu \neq x)$ we have that (see for example [13, Theorem 2.3.5])

$$\langle \nabla f(\mu), x - \mu \rangle \ < \ f(x) - f(\mu).$$

Suppose $\mu_p \neq \mu_q$. Let P be the permutation matrix that transposes p and q only. Then we have

$$f'_q(\mu) - f'_p(\mu))(\mu_p - \mu_q) = \langle \nabla f(\mu), P\mu - \mu \rangle \ \ < \ f(P\mu) - f(\mu) = 0. \quad \blacksquare$$

Lemma 5.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a symmetric function having quadratic expansion at μ, where*

$$\mu_1 = \cdots = \mu_{k_1} > \mu_{k_1+1} = \cdots = \mu_{k_2} > \mu_{k_2+1} \cdots \mu_{k_r}, \qquad (k_r = n).$$

If the Hessian $\nabla^2 f(\mu)$ is positive definite then vector $b = (b_1, ..., b_n)$, defined in Lemma 4.8, has strictly positive entries.

Proof. It is well known that every principal minor in a positive definite matrix is positive definite. Fix an index $1 \leq i \leq n$. If $\mu_{i-1} > \mu_i > \mu_{i+1}$ then from the representation of the matrix $\nabla^2 f(\mu)$ in Lemma 4.8 and the note after it, it is clear that $b_i > 0$. Suppose now that i is in a block of length at least 2. Then some principal minor of $\nabla^2 f(\mu)$ of the form

$$\begin{pmatrix} a + b_i & a \\ a & a + b_i \end{pmatrix}$$

is positive definite, and the result follows. $\blacksquare$

Theorem 5.3 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a symmetric, strictly convex function having quadratic expansion at μ*

$$\mu_1 = \cdots = \mu_{k_1} > \mu_{k_1+1} = \cdots = \mu_{k_2} > \mu_{k_2+1} \cdots \mu_{k_r}, \qquad (k_r = n).$$

Then the symmetric matrix $\nabla^2 f(\mu)$ is positive definite if and only if the bilinear operator $\nabla^2(f \circ \lambda)(Diag\, \mu)$ is positive definite.

Note 5.4 *In fact by Alexandrov's Theorem, if a function is convex it has quadratic expansion at almost every point of its domain [1].*

Proof. Suppose first that the symmetric matrix $\nabla^2 f(\mu)$ is positive definite. Take a symmetric matrix $H \neq 0$. Then we have

$$\sum_{p,q=1}^{n} h_{pp} f''_{pq}(\mu) h_{qq} \geq 0,$$

because $\nabla^2 f(\mu)$ is positive definite,

$$2 \sum_{l=1}^{r} b_{k_l} \sum_{k_{i-1} < p < q \leq k_i} h_{pq}^2 \geq 0,$$

follows from Lemma 5.2, and

$$2 \sum_{p,q:\mu_p > \mu_q} \frac{f'_p(\mu) - f'_q(\mu)}{\mu_p - \mu_q} h_{pq}^2 \geq 0,$$

which follows from Lemma 5.1. Now because $H \neq 0$ at least one of the above inequalities will be strict.

In the other direction the argument is easy: take $H = \operatorname{Diag} x$, for $0 \neq x \in \mathbb{R}^n$ in the formula for $\nabla^2(f \circ \lambda)(\operatorname{Diag} \mu)$ given in Theorem 4.13 to get immediately $x^T \nabla^2 f(\mu) x > 0$. ■

Theorem 5.5 *If $f : \mathbb{R}^n \to \mathbb{R}$ is a symmetric, strictly convex function having quadratic expansion at the point Y, then $\nabla^2 f(\lambda(Y))$ is positive definite if and only if $\nabla^2(f \circ \lambda)(Y)$ is.*

Proof. The proof of this theorem is now clear since $\nabla^2(f \circ \lambda)(Y)$ is positive definite if and only if $\nabla^2(f \circ \lambda)(\operatorname{Diag} \lambda(Y))$ is. ■

6 Examples

Example 6.1 Let g be a function on a scalar argument. Consider the following *separable* symmetric function with its corresponding spectral function:

$$
\begin{aligned}
f(x_1,...,x_n) &= \sum_{i=1}^{n} g(x_i) \\
(f \circ \lambda)(Y) &= \sum_{i=1}^{n} g(\lambda_i(Y)).
\end{aligned}
$$

Then if g has quadratic expansion at the points $x_1,...,x_n$ so does f at $x = (x_1,...,x_n)$ and we have

$$
\begin{aligned}
\nabla f(x) &= (g'(x_1),...,g'(x_n))^T, \\
\nabla^2 f(x) &= \mathrm{Diag}\,(g''(x_1),...,g''(x_n)), \\
b(x) &= (g''(x_1),...,g''(x_n))^T.
\end{aligned}
$$

Suppose g has quadratic expansion at each entry of the vector $\mu \in \mathbb{R}^n_{\downarrow}$ that satisfies

$$
\mu_1 = \cdots = \mu_{k_1} > \mu_{k_1+1} = \cdots = \mu_{k_2} > \mu_{k_2+1} \cdots \mu_{k_r}, \qquad (k_r = n).
$$

Then Theorem 4.13 says that

$$
\begin{aligned}
\nabla^2(f \circ \lambda)(\mathrm{Diag}\,\mu)[H,H] &= \sum_{p=1}^{n} g''(\mu_p)h_{pp}^2 + \sum_{\substack{p \neq q \\ \mu_p = \mu_q}} g''(\mu_p)h_{pq}^2 \\
&\quad + \sum_{p,q:\mu_p \neq \mu_q} \frac{g'(\mu_p) - g'(\mu_q)}{\mu_p - \mu_q} h_{pq}^2 \\
&= \sum_{p,q:\mu_p = \mu_q} g''(\mu_p)h_{pq}^2 + \sum_{p,q:\mu_p \neq \mu_q} \frac{g'(\mu_p) - g'(\mu_q)}{\mu_p - \mu_q} h_{pq}^2.
\end{aligned}
$$

Let us define the following notation consistent with [3, Section V.3]. For any function h defined on a subset of $\mathbb{R}$ define

$$
h^{[1]}(\alpha,\beta) = \begin{cases} \frac{h(\alpha)-h(\beta)}{\alpha-\beta}, & \text{if } \alpha \neq \beta \\ h'(\alpha), & \text{if } \alpha = \beta. \end{cases}
$$

If Λ is a diagonal matrix with diagonal entries α_1,...,α_n, we denote by $h^{[1]}(\Lambda)$ the $n \times n$ symmetric matrix matrix whose (i,j)-entry is $h^{[1]}(\alpha_i, \alpha_j)$. If Y is Hermitian and $Y = U(\mathrm{Diag}\,\lambda(Y))U^T$, let $h^{[1]}(Y) = Uh^{[1]}(\mathrm{Diag}\,\lambda(Y))U^T$.

Using this notation, for the function $h = g'$, we clearly have

$$\begin{aligned} \nabla^2(f \circ \lambda)(\mathrm{Diag}\,\mu)[H,H] &= \langle H, h^{[1]}(\mathrm{Diag}\,\mu) \circ H \rangle, \\ \nabla^2(f \circ \lambda)(Y)[H,H] &= \langle U^T H U, h^{[1]}(\mathrm{Diag}\,\lambda(Y)) \circ (U^T H U) \rangle, \end{aligned} \tag{7}$$

where $Y = U(\mathrm{Diag}\,\lambda(Y))U^T$, and $X \circ Y = (x_{ij}y_{ij})_{i,j=1}^n$ is the Hermitian product of matrices X and Y.

This result is not surprising. Let us extend the domain of the function h to include a subset of the Hermitian matrices in the following way. If $\Lambda = \mathrm{Diag}\,(\alpha_1, ..., \alpha_n)$ is a diagonal matrix whose entries are in the domain of h, we define $h(\Lambda) = \mathrm{Diag}\,(h(\alpha_1), ..., h(\alpha_n))$. If Y is a Hermitian matrix with eigenvalues $\alpha_1, ..., \alpha_n$ in the domain of h, we choose e unitary matrix U such that $Y = U\Lambda U^T$ and define $h(Y) = Uh(\Lambda)U^T$. In this way we can define $h(Y)$ for all Hermitian matrices with eigenvalues in the domain of h. Then the formulae for the gradient in Theorem 4.1 says that for $h = g'$ we have

$$\nabla(f \circ \lambda)(Y) = h(Y).$$

That is why Equations (7) are just the formulas for the derivative ∇h given in Theorem V.3.3 in [3].

Example 6.2 Now we specialize the above example even more. The following spectral function finds a lot of applications in semidefinite programming. Consider the symmetric and strictly convex function and its corresponding spectral function:

$$\begin{aligned} f : x \in \mathbb{R}^n_{++} &\to -\sum_{i=1}^n \log x_i, \\ f \circ \lambda : A \in S^n_+ &\to -\ln \mathrm{Det}\,(A). \end{aligned}$$

(Where S^n_+ denotes the set of all positive definite symmetric matrices.) Then Theorem 4.13 says that for $\mu \in \mathbb{R}^n_\downarrow$ such that

$$\mu_1 = \cdots = \mu_{k_1} > \mu_{k_1+1} = \cdots = \mu_{k_2} > \mu_{k_2+1} \cdots \mu_{k_r}, \qquad (k_r = n),$$

we have

$$\begin{aligned}\nabla^2(f\circ\lambda)(\operatorname{Diag}\mu)[H,H] &= \sum_{p=1}^{n}\frac{h_{pp}^2}{\mu_p^2}+\sum_{\substack{p\neq q\\ \mu_p=\mu_q}}\frac{h_{pq}^2}{\mu_p^2}+\sum_{p,q:\mu_p\neq\mu_q}\frac{h_{pq}^2}{\mu_p\mu_q}\\ &= \sum_{p,q=1,1}^{n,n}\frac{h_{pq}^2}{\mu_p\mu_q}\\ &= \operatorname{tr}\left((\operatorname{Diag}\mu)^{-1}H(\operatorname{Diag}\mu)^{-1}H\right).\end{aligned}$$

The last equality may easily be verified. In general, for arbitrary symmetric matrix A and an orthogonal matrix U such that $U^TAU=\operatorname{Diag}\lambda(A)$, we get

$$\nabla^2(f\circ\lambda)(A)[H,H]=\operatorname{tr}(A^{-1}HA^{-1}H).$$

This agrees with the previously known formula for the second derivative of the function $-\ln\operatorname{Det}(A)$. (See for example [11, Proposition 5.4.5].) Moreover the result in Section 5 tells us that

$$A\succ 0 \ \text{ implies } \ \operatorname{tr}(A^{-1}HA^{-1}H)>0 \ \text{ for all } \ 0\neq H\in S^n,$$

this result can also be verified directly.

Example 6.3 Consider the following symmetric function and its corresponding spectral function:

$$\begin{aligned}\phi_k(x) = &\ k^{\text{th}} \text{ largest element of } \{x_1,x_2\ldots,x_n\},\\ \lambda_k(A) = &\ k^{\text{th}} \text{ largest eigenvalue of } A.\end{aligned}$$

The function $\phi_k(x)$ is linear near every point x such that

$$\bar{x}_{k-1}>\bar{x}_k>\bar{x}_{k+1},$$

since locally we have $\phi_k(x)=x_m$ for the index m such that x_m is the k^{th} largest coordinate of x. In particular if $x\in\mathbb{R}^n_\downarrow$ then $k=m$. So

$$\nabla\phi_k(x)=e^m,\ \nabla^2\phi_k(x)=0,\ b_{\phi_k}(x)=0.$$

Then Theorem 4.13 says that for $\mu\in\mathbb{R}^n_\downarrow$ such that

$$\mu_{k-1}>\mu_k>\mu_{k+1},$$

we have

$$\begin{aligned}\nabla^2\lambda_k(\operatorname{Diag}\mu)[H,H] &= \nabla^2(f\circ\lambda)(\operatorname{Diag}\mu)[H,H]\\ &= 2\sum_{\substack{q\,=\,1\\ q\,\neq\,k}}^{n}\frac{1}{\mu_k-\mu_q}h_{kq}^2\\ &= 2\operatorname{tr}\left((e^k)^T H(\mu_k-\operatorname{Diag}\mu)^\dagger He^k\right).\end{aligned}$$

This agrees with the result in [6, p. 92].

7 The Eigenvalues of $\nabla^2(f\circ\lambda)$

A natural question one may ask is: Is there any relationship between the eigenvalues of $\nabla^2 f(\lambda(Y))$ and those of $\nabla^2(f\circ\lambda)(Y)$? This section shows that in general such a relationship will be quite weak. Let Y be a symmetric matrix such that

$$\begin{aligned}\lambda_1(Y)=\cdots=\lambda_{k_1}(Y)>\cdots>\lambda_{k_{l-1}+1}(Y)=\cdots=\lambda_m(Y)=\cdots=\lambda_{k_l}(Y)\\ >\cdots\lambda_{k_r}(Y),\quad (k_0=0,\ k_r=n).\end{aligned}$$

Using the representation given in Theorem 4.13 and Corollary 4.15 one can see that the eigenvalues of $\nabla^2(f\circ\lambda)(Y)$ are

- $\{\lambda_i\big(\nabla^2 f(\lambda(Y))\big)|i=1,...,n\}$.
- $2b_{k_l}$ is an eigenvalue for every $l=1,...,r$ with multiplicity $(k_l-k_{l-1})(k_l-k_{l-1}-1)/2$.
- $2\frac{f'_{k_l}(\lambda(Y))-f'_{k_s}(\lambda(Y))}{\lambda_{k_l}(Y)-\lambda_{k_s}(Y)}$ is an eigenvalue with multiplicity $(k_l-k_{l-1})(k_s-k_{s-1})$ for every ordered pair $(\lambda_{k_l}(Y),\lambda_{k_s}(Y))$ such that $\lambda_{k_l}(Y)>\lambda_{k_s}(Y)$.

So we can immediately conclude that

$$\begin{aligned}\lambda_{\max}\big(\nabla^2(f\circ\lambda)(Y)\big)&\geq\lambda_{\max}\big(\nabla^2 f(\lambda(Y))\big)\\ \lambda_{\min}\big(\nabla^2 f(\lambda(Y))\big)&\geq\lambda_{\min}\big(\nabla^2(f\circ\lambda)(Y)\big).\end{aligned}\tag{8}$$

We are going to show now that the above inequalities may be strict.

Example 7.1 Consider the convex function

$$f(x,y) := \frac{x^2+y^2}{4} + \frac{\cos 2x + \cos 2y}{8},$$

and the point

$$\mu = (2\pi, \pi) \in \mathbb{R}^2_{\downarrow}.$$

Then

$$\nabla f(x,y) = \begin{pmatrix} \frac{x}{2} - \frac{\sin 2x}{4} \\ \frac{y}{2} - \frac{\sin 2y}{4} \end{pmatrix}, \quad \nabla^2 f(x,y) = \begin{pmatrix} \sin^2 x & 0 \\ 0 & \sin^2 y \end{pmatrix}.$$

Using the representation in Theorem 4.13 we get

$$\nabla^2 f(\mu) = 0, \quad \nabla^2 (f \circ \lambda)(\operatorname{Diag} \mu)[H,H] = h_{12}^2,$$

where

$$H = \begin{pmatrix} h_{11} & h_{12} \\ h_{12} & h_{22} \end{pmatrix}.$$

Then clearly

$$\lambda_{\max}\big(\nabla^2 (f \circ \lambda)(\operatorname{Diag} \mu)\big) = 1 > \lambda_{\max}\big(\nabla^2 f(\mu)\big) = 0.$$

In order to demonstrate a strict inequality between the smallest eigenvalues one needs to consider the function $-f(x,y)$ at the same point μ.

Even though we may not have equalities in (8) at a particular matrix Y, if we consider the eigenvalues of $\nabla^2 f(\lambda(Y))$ and $\nabla^2(f \circ \lambda)(Y)$ over a symmetric, convex set we can see that they vary within the same boundaries. More precisely we have the following theorem. To make its proof precise, we need the main result from [10] saying that: A symmetric function f is C^2 if and only if $f \circ \lambda$ is.

Theorem 7.2 *Let C be a convex and symmetric subset of $\mathbb{R}^n$, and let $f : C \mapsto \mathbb{R}$ be symmetric, C^2 function. Then*

$$\min_{y \in C} \lambda_{\min}(\nabla^2 f(y)) = \min_{Y \in \lambda^{-1}(C)} \lambda_{\min}(\nabla^2 (f \circ \lambda)(Y)). \tag{9}$$

Proof.

$$\begin{aligned}
\lambda_{\min}(\nabla^2 f(y)) > \alpha,\ \ \forall y \in C & \\
&\Leftrightarrow \nabla^2\left(f - \frac{\alpha}{2}\|\cdot\|^2\right)(y) \succ 0,\ \ \forall y \in C \\
&\Leftrightarrow \nabla^2\left(\left(f - \frac{\alpha}{2}\|\cdot\|^2\right) \circ \lambda\right)(Y) \succ 0,\ \ \forall Y \in C \\
&\Leftrightarrow \nabla^2\left(f \circ \lambda - \frac{\alpha}{2}\|\cdot\|_2^2\right)(Y) \succ 0,\ \ \forall Y \in C \\
&\Leftrightarrow \lambda_{\min}(\nabla^2(f \circ \lambda)(Y)) > \alpha,\ \ \forall Y \in C.
\end{aligned}$$

■

Remark 7.3 *(i) The above proof stays the same if '* $>$ *' and* $\succ$ *are changed to '* $\geq$ *' and* $\succeq$ *respectively.*

(ii) If we multiply both sides of Equation (9) by -1 *we will get*

$$\max_{y\in C} \lambda_{\max}(\nabla^2 f(y)) = \max_{Y\in\lambda^{-1}(C)} \lambda_{\max}(\nabla^2(f \circ \lambda)(Y)).$$

Acknowledgments. We thank John Borwein and Heinz Bauschke for pointing out few errors in the manuscript, suggesting Example 6.1, and the connection to [3].

References

[1] A .D. ALEXANDROV. Almost everywhere existence of the second differential of a convex function and some properties of convex surfaces connected with it. *Leningrad State University Annals [Uchenye Zapiski], Mathematical Series*, 6:3–35, 1939. Russian.

[2] V.I. Arnold. On matrices depending on parameters. *Russian Mathematical Surveys*, 26:29–43, 1971.

[3] RAJENDRA BHATIA. *Matrix Analysis.* Springer-Verlag, first edition, 1997.

[4] K. FAN. On a theorem of weyl concerning eigenvalues of linear transformations ii. *Proceedings of the National Academy of Sciences of the United States of America*, 36:31–35, 1950.

[5] J.-B. HIRIART-URRUTY and D. YE. Sensitivity analysis of all eigenvalues of a symmetric matrix. *Numerische Mathematik*, 70:45–72, 1992.

[6] T. KATO. *A Short Introduction to Perturbation Theory for Linear Operators*. Springer-Verlag, Berlin, 1976.

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

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

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

[10] A.S. LEWIS and H.S. SENDOV. Twice differentiable spectral functions. *SIAM Journal of Matrix Analysis*, 2000. submitted.

[11] Y.E. NESTEROV and A.S. NEMIROVSKII. *Interior-Point Polynomial Algorithms in Convex Programming*, volume 13 of *SIAM Studies in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, 1994.

[12] M.L. OVERTON and R.S. WOMERSLEY. Second derivatives for eigenvalue optimization. *SIAM Journal on Matrix Analysis and its Applications*, 16(3):697–718, 1995.

[13] A. L. PERESSINI, F.E. SULLIVAN, and Jr. J.J. UHL. *The Mathematics of Nonlinear Programing*. Undergraduate Texts in Mathematics. Springer-Verlag, 1988.

[14] L.I. SCHIFF. *Quantum Mechanics*. McGraw-Hill, New York, 1955.

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

[16] M. TORKI. First- and second-order epi-differentiability in eigenvalue optimization. *Journal of Mathematical Analysis and Applications*, 234:391–416, 1999.

[17] M. TORKI. *Valeurs propres de matrices symétriques: analyses de la sensibilité d'ordre supérieur et formulations variationelles pour leur estimation.* PhD thesis, Paul Sabatier University, Toulouse, France, 1999.

[18] M. TORKI. Second-order directional derivatives of all eigenvalues of a symmetric matrix. *Nonlinear Analysis, Theory, Methods and Applications*, 2000. To appear.

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