Journal of Convex Analysis
Volume 15 (2008), No. 3, 547–560

Prox-Regularity of Spectral Functions and Spectral Sets

Aris Daniilidis[*]
Departament de Matemàtiques, C1/308,
Universitat Autònoma de Barcelona, 08193 Bellaterra (Cerdanyola del Vallès), Spain
arisd@mat.uab.es, http://mat.uab.es/∼arisd

Adrian Lewis
School of Operations Research and Industrial Engineering,
Cornell University, Ithaca, NY 14853, USA
aslewis@orie.cornell.edu, http://www.orie.cornell.edu/∼aslewis

Jérôme Malick
CNRS, Laboratoire J. Kunztmann, Grenoble, France
jerome.malick@inria.fr, http://bipop.inrialpes.fr/people/malick/

Hristo Sendov[†]
Department of Statistical & Actuarial Sciences,
The University of Western Ontario, London, Ontario, Canada
hssendov@stats.uwo.ca, http://www.stats.uwo.ca/faculty/hssendov/Main.html

Dedicated to the memory of Thomas Lachand-Robert.
Cher ami, tu nous as quittés si tôt...

Received: December 11, 2007
Revised manuscript received: December 17, 2007

Important properties such as differentiability and convexity of symmetric functions in $\mathbb{R}^n$ can be transferred to the corresponding spectral functions and vice-versa. Continuing to built on this line of research, we hereby prove that a spectral function $F\colon \mathbf{S}^n \to \mathbb{R} \cup \{+\infty\}$ is prox-regular if and only if the underlying symmetric function $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is prox-regular. Relevant properties of symmetric sets are also discussed.

Keywords: Spectral function, prox-regular function, eigenvalue optimization, invariant function, permutation theory

1991 Mathematics Subject Classification: Primary 15A18, 49J52; Secondary 47A75, 90C22

1. Introduction

By $\mathbf{S}^n$, $\mathbf{O}^n$, and $\mathbf{\Sigma}^n$ we denote, respectively, the space of $n \times n$ symmetric matrices, the orthogonal group on $\mathbb{R}^n$, and the group of $n \times n$ permutation matrices. For $X \in \mathbf{S}^n$, by $\lambda(X) \in \mathbb{R}^n$ we denote the vector of eigenvalues of X in nonincreasing order:

$$\lambda_1(X) \geq \lambda_2(X) \geq \cdots \geq \lambda_n(X).$$

[*]Research supported by the MEC Grant No. MTM2005-08572-C03-03 (Spain).
[†]Research supported by the NSERC of Canada.

ISSN 0944-6532 / \$ 2.50 © Heldermann Verlag

For any $x \in \mathbb{R}^n$, by $[x] \in \mathbb{R}^n$ we denote the vector with the same coordinates as x ordered nonincreasingly.

A function $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is called *symmetric* if $f(x) = f(\sigma x)$ for all $x \in \operatorname{dom} f := \{x \in \mathbb{R}^n : f(x) < +\infty\}$ and all $\sigma \in \mathbf{\Sigma}^n$. Necessarily, the domain of a symmetric function is a *symmetric set* in $\mathbb{R}^n$: $x \in \operatorname{dom} f$ if and only if $\sigma x \in \operatorname{dom} f$ for all $\sigma \in \mathbf{\Sigma}^n$.

A function $F\colon \mathbf{S}^n \to \mathbb{R} \cup \{+\infty\}$ is called *spectral* if $F(U^\top XU) = F(X)$ for all $X \in \operatorname{dom} F$ and all $U \in \mathbf{O}^n$. Necessarily, the domain of a spectral function is a *spectral set*: $X \in \operatorname{dom} F$ implies that the orbit $\{U^\top XU : U \in \mathbf{O}^n\}$ is also in $\operatorname{dom} F$. Note that if K is a symmetric set, then

$$\lambda^{-1}(K) := \{X \in \mathbf{S}^n : \lambda(X) \in K\}$$

is a spectral set. It is not difficult to see that the spectral functions are in one-to-one correspondence with the symmetric functions. The relationship is given by the formulae

$$F(X) = (f \circ \lambda)(X) \qquad \text{and} \qquad f(x) = F(\operatorname{Diag} x),$$

where $\operatorname{Diag} x$ denotes the $n \times n$ diagonal matrix whose diagonal elements are the coordinates of x. Properties of a symmetric function f on $\mathbb{R}^n$ and the associated spectral function $F := f \circ \lambda$ on $\mathbf{S}^n$ are closely related. Even though the map $X \mapsto \lambda(X)$ can be very badly behaved, for example it is not everywhere differentiable, many problems are corrected by the invariance property of f. An illustration of this is given with the following theorem [7] (see also [11]).

Theorem 1.1 (Convexity preserved). *Let the set K in $\mathbb{R}^n$ be convex and symmetric and suppose that the function $f\colon K \to \mathbb{R}$ is symmetric. Then the set $\lambda^{-1}(K)$ is convex and the spectral function $F = f \circ \lambda$ is convex if and only if f is convex.*

Differentiability is another property that is preserved, [12, Theorem 1.1], as recalled by the next theorem. An analogous result also holds for twice (continuously) differentiable spectral functions, see [14]; for C^∞ spectral functions, see [5]; and for analytic spectral functions, see [15].

Theorem 1.2 (Differentiability preserved). *Let the set K in $\mathbb{R}^n$ be open and symmetric and suppose that the function $f\colon K \to \mathbb{R}$ is symmetric. Then the spectral function $F = f \circ \lambda$ is (continuously) differentiable at (around) the matrix X if and only if f is (continuously) differentiable at (around) the vector $\lambda(X)$.*

In this paper we continue to built on this line of research. We establish that the important variational property of prox-regularity can be added to the list of properties for which the transfer principle is valid. The prox-regularity, studied in [16] and [17] in particular, has proved to be a robust notion of nonsmoothness enjoying nice "geometrical" properties, generalizing both convex functions and smooth functions. The prox-regularity has also been used in algorithms, in particular in the identification of active constraints or the conceptual construction of predictor-corrector algorithms, see for instance [8], [6].

In order to give the precise definition of prox-regularity, we need to recall some basic definitions, following [19]. A set $C \subset \mathbb{R}^n$ is said to be *locally closed* at a point $\bar{x}$ if $C \cap V$ is closed for some closed neighborhood V of $\bar{x}$. A function $\varphi\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is said to be *locally lower semicontinuous* at $\bar{x}$ if $\varphi(\bar{x})$ is finite and $\operatorname{epi} \varphi := \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : \varphi(x) \le t\}$

is locally closed at $(\bar{x}, \varphi(\bar{x}))$. Given a function $\varphi\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, we say that $v \in \mathbb{R}^n$ is a *regular subgradient* of φ at $\bar{x}$, denoted by $v \in \hat{\partial}\varphi(\bar{x})$, if φ is finite at $\bar{x}$ and

$$\varphi(x) \geq \varphi(\bar{x}) + v^\top(x - \bar{x}) + o(\|x - \bar{x}\|).$$

As usual, $t \mapsto o(t)$ denotes a real-valued function defined in a neighborhood of the origin 0 of $\mathbb{R}$ and satisfying $o(t)/t \to 0$ as $t \to 0$. If the function φ is infinite at $\bar{x}$ then we set $\hat{\partial}\varphi(\bar{x}) = \emptyset$. We say that v is a *subgradient* of φ at $\bar{x}$, written $v \in \partial\varphi(\bar{x})$, if φ is finite at $\bar{x}$ and there is a sequence $x_k \to \bar{x}$ with values $\varphi(x_k) \to \varphi(\bar{x})$ and a sequence $v_k \in \hat{\partial}\varphi(x_k)$ such that $v_k \to v$. Analogously, if φ is infinite at $\bar{x}$ then we set $\partial\varphi(\bar{x}) = \emptyset$.

Throughout the text, $B(x, r)$ will denote the open ball with center $x \in \mathbb{R}^n$ and radius $r > 0$. The definition of prox-regular function is then as follows, see [19, Definition 13.27].

Definition 1.3 (Prox-regularity). A function $\varphi\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is called *prox-regular at $\bar{x}$ for $\bar{v}$* if φ is finite and locally lower semicontinuous at $\bar{x}$, $\bar{v} \in \partial\varphi(\bar{x})$ and there exist $\delta > 0$ and $\rho \geq 0$ such that for all $x, y \in B(\bar{x}, \delta)$ and $v \in \partial\varphi(x)$ with $\varphi(x) \leq \varphi(\bar{x}) + \delta$ and $\|v - \bar{v}\| \leq \delta$, we have

$$\varphi(y) \geq \varphi(x) + v^\top(y - x) - \frac{\rho}{2}\|y - x\|^2.$$

The function φ is called prox-regular at $\bar{x}$, if it is prox-regular at $\bar{x}$ for all $\bar{v} \in \partial\varphi(\bar{x})$.

The main result of this paper, stating that the prox-regularity is transferred from a symmetric function to the corresponding spectral function and vice-versa, is the content of the following theorem.

Theorem 1.4 (Prox-regularity preserved). *Let f be a symmetric lower semicontinuous function. Then $F = f \circ \lambda$ is prox-regular at $\bar{X}$ if and only if f is prox-regular at $\lambda(\bar{X})$.*

The proof of the above theorem will be given at the end of the paper (Theorem 4.2 in Section 4). Before, in Section 2, we shall first consider two particular cases of prox-regular spectral functions, for which a direct proof of the transfer principle can be given. In Section 3, we shall take a close look at the subdifferentials of spectral and symmetric functions, building tools for our development. We finish this first section by fixing terminology and notation.

Notation – Terminology. The canonical Euclidean norm on the space $\mathbf{S}^n$ of $n \times n$ symmetric matrices, often called the Frobenius norm, is defined by the formula:

$$\|X\|^2 = \sum_{i,j=1}^{n} X_{ij}^2 = \operatorname{tr}(X^2).$$

The associated inner product is denoted by $\langle X, Y\rangle = \operatorname{tr}(XY)$. The above formula for the norm in $\mathbf{S}^n$, when restricted to diagonal matrices, corresponds to the Euclidean norm in $\mathbb{R}^n$ (still denoted by $\|\cdot\|$), since $\|x\| = \|\operatorname{Diag} x\|$. It is also well-known that

$$\|X\|^2 = \sum_{i=1}^{n} \lambda_i^2(X) = \|\lambda(X)\|^2$$

that is

$$\|X\| = (\|\cdot\| \circ \lambda)(X). \tag{1}$$

The above relation shows that the Frobenius norm is a spectral function on $\mathbf{S}^n$ associated with the canonical Euclidean norm on $\mathbb{R}^n$.

In the sequel we shall say that two matrices X, Y in $\mathbf{S}^n$ admit a *simultaneous spectral decomposition* if they are simultaneously diagonalizable in the same orthonormal basis, that is, if for some orthogonal matrix $U \in \mathbf{O}^n$ the matrices $U^\top X U$ and $U^\top Y U$ are diagonal. It is known that X and Y admit simultaneous spectral decomposition if and only if $XY = YX$ (see [9]). A more restrictive condition is to assume that the matrices X and Y admit a *simultaneous ordered spectral decomposition*, which guarantees that the obtained diagonal matrices are precisely $\operatorname{Diag}\lambda(X)$ and $\operatorname{Diag}\lambda(Y)$, that is, the entries in both diagonals are ordered in a nonincreasing way. The next theorem due to Fan shows precisely when two matrices X and Y admit simultaneous ordered spectral decomposition (see [3, Theorem 1.2.1]).

Theorem 1.5 (Fan). *Any two matrices X and Y in $\mathbf{S}^n$ satisfy the inequality*

$$\langle X, Y\rangle = \operatorname{tr}(XY) \leq \lambda(X)^\top \lambda(Y).$$

Equality holds if and only if X and Y admit a simultaneous ordered spectral decomposition.

2. Examples of prox-regular spectral functions

In this section, we consider two particular cases of Theorem 1.4 for which the transfer principle can be established by direct arguments. Namely, we discuss the case of uniform prox-regular spectral functions and of indicator functions of prox-regular spectral sets. A common point of both cases is a uniform character of prox-regularity.

2.1. Uniform prox-regularity

The notion of uniform prox-regularity corresponds to the standard prox-regularity with parameters independent of the subgradients $v \in \partial\varphi(\bar{x})$ (see [2]).

Definition 2.1 (Uniform prox-regularity). A function $\varphi\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is called *uniformly prox-regular at* $\bar{x}$ if there exist $\delta > 0$ and $\rho \geq 0$ such that for all $x, y \in B(\bar{x}, \delta)$ and $v \in \partial\varphi(x)$ with $\varphi(x) \leq \varphi(\bar{x}) + \delta$, we have

$$\varphi(y) \geq \varphi(x) + v^\top (y - x) - \frac{\rho}{2}\|y - x\|^2.$$

A uniformly prox-regular locally Lipschitz function is also called proximally smooth or lower-C^2 ([18], [4]). Let us recall that a lower semicontinuous (respectively, a locally Lipschitz) function f is uniformly prox-regular (at a point x) if and only if f admits a local representation (around x) of the form

$$f = g - \beta\|\cdot\|^2,$$

where g is a lower semicontinuous (respectively, continuous) convex function (see [2, Corollary 3.12] and [1, Theorem 4.1] for example). Using this representation a straightforward proof of Theorem 1.4 for the case of uniformly prox-regular functions can be given: indeed, we can write

$$f \circ \lambda = (g - \beta\|\cdot\|^2) \circ \lambda = g \circ \lambda - \beta\|\cdot\|^2,$$

the second equality stemming from (1). Thus the result follows from the convex transfer principle (*cf.* Theorem 1.1).

2.2. Indicator functions and spectral sets

Let C be a subset of $\mathbb{R}^n$ and $\bar{x} \in C$. A vector v is called a *regular normal vector* to C at $\bar{x}$, denoted by $v \in \hat{N}_C(\bar{x})$, if

$$v^\top (x - \bar{x}) \le o(\|x - \bar{x}\|) \ \text{ for } x \in C.$$

A vector v is called a *normal vector*, denoted by $v \in N_C(\bar{x})$, if there exist sequences $x_k \to \bar{x}$ and $v_k \to v$ with $v_k \in \hat{N}_C(x_k)$. A closed subset C of $\mathbb{R}^n$ is called *prox-regular at* $\bar{x} \in C$ *for* $\bar{v} \in N_C(\bar{x})$ if there exist $\delta > 0$ and $\rho > 0$ such that whenever $x \in C$ and $v \in N_C(x)$ with $\|x - \bar{x}\| < \delta$ and $\|v - \bar{v}\| < \delta$, then x is the unique nearest point of $\{x' \in C : \|x' - \bar{x}\| < \delta\}$ to $x + v/\rho$. The set C is *prox-regular at* $\bar{x}$ if this property holds for every vector $\bar{v} \in N_C(\bar{x})$.

As expected, C is prox-regular if and only if its indicator function is prox-regular at $\bar{x}$ and, according to [17, Proposition 1.2], C is prox-regular at $\bar{x}$ is and only if it is prox-regular at $\bar{x}$ for $\bar{v} = 0$. Let us now recall from [17, Theorem 1.3] another important characterization of prox-regularity for sets.

Theorem 2.2 (Prox-regular sets vs distance functions). *Let $C \subset \mathbb{R}^n$ be a closed set and $\bar{x} \in C$. Then C is prox-regular at $\bar{x}$ if and only if the distance function d_C is continuously differentiable on $O \setminus C$ for some open neighborhood O of $\bar{x}$.*

In the sequel we use the above characterization to get a direct proof of the transfer principle of prox-regularity for spectral sets, or equivalently, for indicator functions. To this end we need to establish that the distance function $\mathrm{d}_K(x) := \inf_{y \in K} \|x - y\|$ to a symmetric subset K of $\mathbb{R}^n$ is a symmetric function. This is one of the conclusions of the following statement.

Proposition 2.3 (Symmetric distance functions). *Let K be a symmetric subset of $\mathbb{R}^n$. Then the distance function d_K to K is symmetric: $\mathrm{d}_K(x) = \mathrm{d}_K(\sigma x)$ for all $\sigma \in \mathbf{\Sigma}^n$ and $x \in \mathbb{R}^n$. Moreover, the distance function $\mathrm{D}_{\lambda^{-1}(K)}$ to the spectral set $\lambda^{-1}(K)$ satisfies:*

$$\mathrm{D}_{\lambda^{-1}(K)} = \mathrm{d}_K \circ \lambda.$$

Proof. Let $x \in \mathbb{R}^n$ and $\sigma \in \mathbf{\Sigma}^n$. Since we have $\sigma K = K$, making the change of variables $z = \sigma y$ we deduce that

$$\mathrm{d}_K(\sigma x) = \inf_{z \in K} \|\sigma x - z\| = \inf_{y \in K} \|\sigma x - \sigma y\| = \inf_{y \in K} \|x - y\| = \mathrm{d}_K(x),$$

which shows that d_K is permutation invariant. To see that $\mathrm{D}_{\lambda^{-1}(K)}$ is a spectral function we fix $X \in \mathbf{S}^n$ and $U \in \mathbf{O}^n$ such that $X = U^\top \mathrm{Diag}\lambda(X)U$, and we obtain

$$\begin{aligned}
\mathrm{D}_{\lambda^{-1}(K)}(X) &= \inf_{Y \in \lambda^{-1}(K)} \|X - Y\| \\
&= \inf_{Y \in \lambda^{-1}(K)} \|U^\top(\mathrm{Diag}\,\lambda(X))U - Y\| \\
&= \inf_{Y \in \lambda^{-1}(K)} \|\mathrm{Diag}\,\lambda(X) - Y\| \\
&\leq \inf_{y \in K} \|\lambda(X) - y\| \\
&= \mathrm{d}_K(\lambda(X)).
\end{aligned}$$

For the opposite inequality, let us observe that a direct consequence of Theorem 1.5 is the fact that $\|\lambda(X) - \lambda(Y)\| \leq \|X - Y\|$, for any two symmetric matrices X and Y. Using this we deduce

$$\begin{aligned}
\mathrm{D}_{\lambda^{-1}(K)}(X) &= \inf_{Y \in \lambda^{-1}(K)} \|X - Y\| \\
&\geq \inf_{Y \in \lambda^{-1}(K)} \|\lambda(X) - \lambda(Y)\| \\
&\geq \inf_{y \in K} \|\lambda(X) - y\| \\
&= \mathrm{d}_K(\lambda(X)).
\end{aligned}$$

The proof is complete. □

The following result relates the prox-regularity of symmetric sets with the prox-regularity of the corresponding spectral sets; in other words, it proves Theorem 1.4 in the particular case of indicator functions of spectral sets.

Theorem 2.4 (Prox-regular spectral sets). *Let K be a symmetric subset of $\mathbb{R}^n$ and let X be an element of $\lambda^{-1}(K)$. Then the set K is prox-regular at $\lambda(X)$ if and only if $\lambda^{-1}(K)$ is prox-regular at X.*

Proof. Observe first that K is closed if and only if $\lambda^{-1}(K)$ is. We deduce successively

$$\begin{aligned}
& K \text{ is prox-regular at } \lambda(X) & \\
\iff\ & \mathrm{d}_K \text{ is } C^1 \text{ around } \lambda(X) & \text{[Theorem 2.2]} \\
\iff\ & \mathrm{d}_K \circ \lambda \text{ is } C^1 \text{ around } X & \text{[Theorem 1.2]} \\
\iff\ & \mathrm{D}_{\lambda^{-1}(K)} \text{ is } C^1 \text{ around } X & \text{[Proposition 2.3]} \\
\iff\ & \lambda^{-1}(K) \text{ is prox-regular at } X & \text{[Theorem 2.2]}
\end{aligned}$$

which completes the proof. □

We end this subsection about spectral sets by stressing an interesting property of the spectral prox-regular set $\lambda^{-1}(K)$. Being prox-regular, the projection mapping is locally unique, that is, there exists a unique nearest point locally around $\lambda^{-1}(K)$; on the other hand, being a spectral set, we can get an explicit expression of its projection: if the point $x \in K$ is the nearest point of K to $y \in \mathbb{R}^n$, then for any orthogonal matrix $U \in \mathbf{O}^n$, the matrix $U^\top(\mathrm{Diag}\,x)U \in \lambda^{-1}(K)$ is a nearest matrix of the spectral set $\lambda^{-1}(K)$ to the matrix $U^\top(\operatorname{Diag} y)U$. This result has been recently established in [10, Theorem A.1] and generalizes several projection results that are used in projection algorithms in a nonconvex setting (see the introduction of [10] for an overview of this question). Using the material of this paper we can hereby give an alternative quick proof of the aforementioned result along the following lines: Since d_K is a symmetric function, we have $d_K(y) = \|x-y\| = d_K([y])$, thus by Proposition 2.3 we obtain

$$\mathrm{D}_{\lambda^{-1}(K)}(U^\top(\operatorname{Diag} y)U) = d_K([y]) = \|x-y\| = \|U^\top(\operatorname{Diag} x)U - U^\top(\operatorname{Diag} y)U\|,$$

which proves the desired assertion.

3. Properties of subdifferentials

In order to tackle the general (non-uniform) case, we have to grind our tools: in this section we study properties of the subdifferentials of spectral and symmetric functions.

Theorem 3.1 below gives a full description of the subdifferential of a spectral function $F = f \circ \lambda$ in terms of the subdifferential of the underlying symmetric function f. This result is a cornerstone for the variational theory of spectral mappings and will play a fundamental role in our analysis. Results of this kind were initially established for subdifferentials of convex spectral functions (see [11], [3] for example). A much more general result holds for the class of lower semicontinuous spectral functions and for the notions of regular, limiting or Clarke subdifferential (see [13] for details).

Theorem 3.1 (Subdifferential of spectral functions). *If f is a lower semicontinuous function, then*

$$\partial F(X) = \{U^\top(\operatorname{Diag} v)U : v \in \partial f(\lambda(X)) \text{ and } U \in \mathbf{O}^n_X\}, \tag{2}$$

where

$$\mathbf{O}^n_X = \{U \in \mathbf{O}^n : X = U^\top(\operatorname{Diag} \lambda(X))U\}. \tag{3}$$

We point out that given two matrices $X, V \in \mathbf{S}^n$ the relation $V \in \partial F(X)$ implies that X and V admit a simultaneous spectral decomposition. Interestingly, when F is a convex function then the relation $V \in \partial F(X)$ implies that X and V admit a simultaneous *ordered* spectral decomposition. Indeed, by (2) we have $V = U^\top(\operatorname{Diag} v)U$ for some $v \in \partial f(\lambda(X))$ and $U \in \mathbf{O}^n_X$. Then, by the convexity of F we obtain

$$F(Y) \geq F(X) + \langle V, Y - X\rangle \quad \text{for all } Y \in \mathbf{S}^n.$$

Let $\sigma \in \mathbf{\Sigma}^n$ be such that $\sigma\lambda(V) = v$ and take $Y = U^\top(\operatorname{Diag} \sigma\lambda(X))U$. Then the above inequality yields

$$F(X) - \lambda(V)^\top\lambda(X) = F(Y) - \langle V, Y\rangle \geq F(X) - \langle V, X\rangle,$$

whence $\langle V, X\rangle \geq \lambda(V)^\top\lambda(X)$, which in view of Theorem 1.5 shows that X and V admit a simultaneous ordered spectral decomposition.

The fact that convexity of F is crucial for the conclusion that X and V admit a simultaneous ordered spectral decomposition is illustrated by the following example.

Example 3.2 (Unordered decomposition). Consider the symmetric function $f(x_1,x_2) = x_1x_2$. It follows easily that the spectral function $f \circ \lambda$ is differentiable at the point $X = \operatorname{Diag}(1,2)$ with gradient $V = \operatorname{Diag}(2,1)$. Obviously the matrices X and V are simultaneously diagonalizable (and then admit simultaneous spectral decomposition), but they do not have a simultaneous ordered spectral decomposition. □

In the convex case the property that the matrices X and V admit a simultaneous ordered spectral decomposition simplifies significantly the variational analysis. We can indeed relate the size of the subgradients of the functions f and F, with the estimation

$$\|\lambda(V) - \lambda(V')\| \leq \|V - V'\|, \tag{4}$$

holding with equality if and only if V and V' admit a simultaneous ordered spectral decomposition (as a direct consequence from Theorem 1.5). On the other hand, if f is a general lower semicontinuous spectral function, the group of permutations over the coordinates should be taken into account: forthcoming Theorem 3.6 will thus be very useful for our purposes.

Given $x \in \mathbb{R}^n$ and $v \in \partial f(x)$ the following set of permutations appears naturally in our study:

$$\mathscr{S}_{x,v} = \{\sigma \in \mathbf{\Sigma}^n : \sigma v \in \partial f(x)\}, \tag{5}$$

that is, permutations that applied to v remain in the subdifferential.

Remark 3.3 (Permutations leaving x invariant). It is straightforward to see that for every permutation $\sigma \in \mathbf{\Sigma}^n$ and any $x \in \mathbb{R}^n$ we have

$$\partial f(\sigma x) = \sigma \partial f(x).$$

Thus, any permutation $\sigma \in \mathbf{\Sigma}^n$ leaving x invariant (that is, $\sigma x = x$) belongs in particular to $\mathscr{S}_{x,v}$ for any $v \in \partial f(x)$. On the other hand, the example of the constant function $f(x_1, x_2) = 0$, for all $(x_1, x_2) \in \mathbb{R}^2$ or of the (symmetric) function

$$g(x_1, x_2) = \min\{|x_1 - x_2 - 1|, |x_1 - x_2 + 1|\}$$

show that, in general, the set $\mathscr{S}_{x,v}$ may contain more elements. Indeed, take (in both cases) $x = (1, 0)$, and let $u = (0, 0) \in \partial f(x)$, $v = (1, -1) \in \partial g(x)$ and σ the non–trivial permutation of Σ^2. □

The following lemma is taken from [13, Proposition 3].

Lemma 3.4 (Simultaneous conjugacy). *Given vectors x, y, u and v in $\mathbb{R}^n$, there is a matrix $U \in \mathbf{O}^n$ with*

$$\operatorname{Diag} x = U^\top (\operatorname{Diag} u) U \qquad \textit{and} \qquad \operatorname{Diag} y = U^\top (\operatorname{Diag} v) U$$

if and only if there is a permutation $\sigma \in \mathbf{\Sigma}^n$ with $x = \sigma u$ and $y = \sigma v$.

Let us continue our analysis with the following technical lemma stating that if two subgradients of the spectral function F are close to each other, then the underlying subgradients of the corresponding symmetric function f are also nearby up to a permutation.

Lemma 3.5 (Proximity of subgradients). *Consider a subgradient $\bar{V}$ of the function F at the matrix $\bar{X}$, and the corresponding decomposition $\bar{V} = \bar{U}^\top(\operatorname{Diag}\bar{v})\bar{U}$, where $\bar{U} \in \mathbf{O}^n_{\bar{X}}$, $\bar{v} \in \partial f(\bar{x})$ and $\bar{x} = \lambda(\bar{X})$. Then for every $\varepsilon > 0$, there exists $\delta \in (0, \varepsilon)$ such that for any $V \in \partial F(X)$ with corresponding decomposition $V = U^\top(\operatorname{Diag} v)U$ for some $v \in \partial f(\lambda(X))$ and $U \in \mathbf{O}^n_X$ satisfying*

$$\|X - \bar{X}\| \le \delta \qquad \text{and} \qquad \|V - \bar{V}\| \le \delta$$

there exists a permutation $\sigma \in \mathscr{S}_{\bar{x},\bar{v}}$ such that $\|v - \sigma\bar{v}\| \le \varepsilon$.

Proof. Let us assume, towards a contradiction, that there exist $\varepsilon > 0$ and sequences $X_k \to \bar{X}$, $V_k \to \bar{V}$, $U_k \in \mathbf{O}^n_{X_k}$, and $v_k \in \partial f(\lambda(X_k))$ satisfying

$$X_k = {U_k}^\top(\operatorname{Diag}\lambda(X_k))U_k \qquad \text{and} \qquad V_k = {U_k}^\top(\operatorname{Diag} v_k)U_k \tag{6}$$

such that

$$\forall \sigma \in \mathscr{S}_{\bar{x},\bar{v}} \quad \|v_k - \sigma\bar{v}\| > \varepsilon. \tag{7}$$

Let $\{\sigma_k\}_{k\ge1} \subset \mathbf{\Sigma}^n$ be such that $v_k = \sigma_k\lambda(V_k)$, for all $k \ge 1$. Since $\mathbf{O}^n$ is compact, there is no loss of generality to assume that $U_k \to U$. Since $\mathbf{\Sigma}^n$ is finite, it follows by the continuity of $\lambda(\cdot)$ that v_k approaches $\tilde{v} := \tilde{\sigma}\,\lambda(\bar{V})$ for some $\tilde{\sigma} \in \mathbf{\Sigma}^n$. Let us now observe that

$$\bar{V} = U^\top(\operatorname{Diag}\tilde{v})U = \bar{U}^\top(\operatorname{Diag}\bar{v})\bar{U},$$

yielding

$$\operatorname{Diag}\tilde{v} = (U\bar{U}^\top)(\operatorname{Diag}\bar{v})(U\bar{U}^\top)^\top.$$

Since $U, \bar{U} \in \mathbf{O}^n_{\bar{X}}$, we also have

$$\operatorname{Diag}\bar{x} = (U\bar{U}^\top)(\operatorname{Diag}\bar{x})(U\bar{U}^\top)^\top.$$

Applying Lemma 3.4 together with Remark 3.3, we conclude that there is a permutation $\bar{\sigma} \in \mathscr{S}_{\bar{x},\bar{v}}$ such that $\tilde{v} = \bar{\sigma}\bar{v}$. This contradicts (7) and the proof is complete. □

Theorem 3.6 (Proximity up to a permutation). *Let $\bar{V}$ be a subgradient of F at $\bar{X}$ and let $\bar{V} = \bar{U}^\top(\operatorname{Diag}\bar{v})\bar{U}$ be its corresponding decomposition, where $\bar{U} \in \mathbf{O}^n_{\bar{X}}$, $\bar{v} \in \partial f(\bar{x})$ and $\bar{x} = \lambda(\bar{X})$. Then there exists $\tilde{\delta} > 0$ such that for all $\delta \in [0, \tilde{\delta})$ and all $V \in \partial F(X)$ with decomposition $V = U^\top(\operatorname{Diag} v)U$ for $v \in \partial f(\lambda(X))$ and $U \in \mathbf{O}^n_X$ satisfying*

$$\|X - \bar{X}\| \le \delta \qquad \text{and} \qquad \|V - \bar{V}\| \le \delta$$

there exists

$$\sigma \in \mathscr{S}_{\bar{x},\bar{v}} \quad \text{such that } \|v - \sigma\bar{v}\| \le \delta.$$

Thus, by (5), we have $\operatorname{dist}(v, \partial f(\bar{x})) \le \delta$.

Proof. Let us set

$$\Delta = \min\{\|\sigma\bar{v} - \tau\bar{v}\| : \tau \in \mathscr{S}_{\bar{x},\bar{v}},\ \sigma \notin \mathscr{S}_{\bar{x},\bar{v}}\} > 0. \tag{8}$$

Applying Lemma 3.5 with $\varepsilon = \Delta/3$ we get a constant $\tilde{\delta} > 0$ (with $\Delta/3 > \tilde{\delta}$) and a permutation $\tau \in \mathscr{S}_{\bar{x},\bar{v}}$ such that $\|v - \tau\bar{v}\| \le \Delta/3$. Let us now consider the permutations σ, $\bar{\sigma}$ such that $v = \sigma\lambda(V)$ and $\bar{v} = \bar{\sigma}\lambda(\bar{V})$, so that

$$\|v - \sigma\bar{\sigma}^{-1}\bar{v}\| = \|\sigma^{-1}v - \bar{\sigma}^{-1}\bar{v}\| = \|\lambda(V) - \lambda(\bar{V})\| \le \|V - \bar{V}\|.$$

Thus setting $\omega = \sigma\bar{\sigma}^{-1}$ we have $\|v - \omega\bar{v}\| \le \|V - \bar{V}\| \le \tilde{\delta} \le \Delta/3$. To conclude, it is sufficient to show that $\omega \in \mathscr{S}_{\bar{x},\bar{v}}$. Indeed,

$$\|\omega\bar{v} - \tau\bar{v}\| \le \|\omega\bar{v} - v\| + \|v - \tau\bar{v}\| \le \Delta/3 + \Delta/3 < \Delta,$$

which in view of (8) yields that $\omega \in \mathscr{S}_{\bar{x},\bar{v}}$. The proof is complete. □

4. Prox-regularity of spectral functions

To prove our main result, we need the following characterization of prox-regularity for the case of symmetric functions.

Lemma 4.1 (Prox-regularity of symmetric functions). *Let f be a lower semicontinuous symmetric function. Then f is prox-regular at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$ if and only if there exist $\rho > 0$ and $\delta > 0$ such that for all $x, y \in B(\bar{x}, \delta)$ and $v \in \partial f(x)$ with $f(x) \le f(\bar{x}) + \delta$ and $\|v - \bar{v}\| \le \delta$ we have*

$$f(\sigma y) \ge f(x) + v^\top(\sigma y - x) - \frac{\rho}{2}\|\sigma y - x\|^2 \quad \text{for all } \sigma \in \mathbf{\Sigma}^n. \tag{9}$$

Proof. The sufficiency of the above condition is obvious (just take $\sigma = id$). Let us prove the necessity part. The prox-regularity of f at $\bar{x}$ for $\bar{v}$ gives $\tilde{\delta} > 0$ and $\tilde{\rho} > 0$ such that for all $x, z \in B(\bar{x}, \tilde{\delta})$ and $v \in \partial f(x)$ satisfying $\|v - \bar{v}\| \le \tilde{\delta}$ and $f(x) \le f(\bar{x}) + \tilde{\delta}$ we have

$$f(z) \ge f(x) + v^\top(z - x) - \frac{\tilde{\rho}}{2}\|z - x\|^2.$$

Let us pick any

$$L > \|\bar{v}\| + \tilde{\delta} \tag{10}$$

and let us use the lower-semicontinuity of f to obtain a positive constant

$$\delta \le \min\{1, \tilde{\delta}\} \tag{11}$$

such that for all $y \in B(\bar{x}, \delta)$

$$f(y) \ge f(\bar{x}) - L + \tilde{\delta}. \tag{12}$$

Let us finally set

$$\rho \ge \max\left\{\tilde{\rho},\, 4L/\delta^2\right\}.$$

Having defined the constants $\delta, \rho > 0$ let us take $\sigma \in \mathbf{\Sigma}^n$, $x, y \in B(\bar{x}, \delta)$ and $v \in \partial f(x)$ such that $\|v - \bar{v}\| \le \delta$ and $f(x) \le f(\bar{x}) + \delta$. We aim to prove that (9) holds. Observe that this is indeed the case whenever $\|\sigma y - x\| \le \tilde{\delta}$, so we may assume $\|\sigma y - x\| > \tilde{\delta}$. Let us further set

$$\Delta_1 = f(\bar{x}) - L + \tilde{\delta} \qquad \text{and} \qquad \Delta_2(\mu) = f(\bar{x}) + \tilde{\delta} + L\mu - \frac{\rho}{2}\mu^2 \quad (\mu \in \mathbb{R}).$$

Claim. $\Delta_1 \geq \Delta_2(\mu)$ for all $\mu \in (\delta, +\infty)$.

Proof of the Claim. We need to show that the inequality

$$\frac{\rho}{2}\mu^2 - L\mu - L > 0,$$

holds for all $\mu \in (\delta, +\infty)$. The discriminant of the left-hand side, as a polynomial in μ, is strictly positive. It is easy to see that since $\rho \geq 4L/\delta^2$ its larger root satisfies

$$\frac{L + \sqrt{L(L+2\rho)}}{\rho} \leq \delta.$$

This proves the claim.

We further infer from (12) and the invariance of f that

$$f(\sigma y) = f(y) \geq \Delta_1, \tag{13}$$

while using (10) we deduce that

$$\|v\| \leq \|\bar{v}\| + \delta \leq \|\bar{v}\| + \tilde{\delta} < L\,.$$

Let us set $\mu := \|\sigma y - x\| > \tilde{\delta}$ and note that in view of (11) we have $\mu > \delta$. We thus deduce

$$f(x) + v^\top(\sigma y - x) - \frac{\rho}{2}\|\sigma y - x\|^2 \leq (f(\bar{x}) + \tilde{\delta}) \; + \; \|v\|\,\mu - \frac{\rho}{2}\,\mu^2 \; \leq \; \Delta_2(\mu)\,.$$

Since $\Delta_1 \geq \Delta_2(\mu)$ we obtain from (13) that

$$f(\sigma y) \geq f(x) + v^\top(\sigma y - x) - \frac{\rho}{2}\|\sigma y - x\|^2,$$

which completes the proof. □

We are now in position to prove the main result of this work.

Theorem 4.2 (Main result). *Let f be a symmetric lower semicontinuous function. Then f is prox-regular at $\lambda(\bar{X})$ if and only if $F = f \circ \lambda$ is prox-regular at $\bar{X}$.*

Proof. **($\Leftarrow$).** Suppose that F is prox-regular at $\bar{X}$ for any $\bar{V} \in \partial F(\bar{X})$. Then, it is easy to see that f is prox-regular at $\lambda(\bar{X})$ for any $\bar{v} \in \partial f(\lambda(\bar{X}))$ by using (4) and the formula $f(x) = F(U^\top(\operatorname{Diag} x)U)$.

($\Rightarrow$). Assume that f is prox-regular at $\lambda(\bar{X})$. We need to prove that F is prox-regular at $\bar{X}$ for any $\bar{V} \in \partial F(\bar{X})$. Set $\bar{x} = \lambda(\bar{X})$, let, by Theorem 3.1, $\bar{v} \in \partial f(\bar{x})$ and $\bar{U} \in \mathbf{O}^n_{\bar{X}}$ be such that $\bar{V} = \bar{U}^\top(\operatorname{Diag}\bar{v})\bar{U}$. To prove the prox-regularity of F at $\bar{X}$ for $\bar{V}$, we proceed in three steps:

- the first step consists in fixing the values of the parameters δ, $\rho > 0$;
- in the second step we introduce the working variables in $\mathbf{S}^n$ and $\mathbb{R}^n$;
- in the final step, we deduce the inequality of prox-regularity of F at $\bar{X}$ from the one of f at $\bar{x}$.

Step 1: Choice of parameters. We first apply Theorem 3.6 with respect to $\bar V \in \partial F(\bar X)$ (and its given decomposition $\bar V = \bar U^\top (\mathrm{Diag}\,\bar v)\bar U$, with $\bar U \in \mathbf{O}^n_{\bar X}$, $\bar v \in \partial f(\bar x)$ and $\bar x = \lambda(\bar X)$) to obtain $\tilde\delta > 0$. Then, for each $\tau \in \mathscr{S}_{\bar x, \bar v}$, we use the prox-regularity of f at $\bar x$ for $\tau\bar v \in \partial f(\bar x)$ to get $\delta_\tau > 0$ and $\rho_\tau > 0$ by Lemma 4.1. We then set

$$\delta = \min\Big\{\{\delta_\tau,\ \tau \in \mathscr{S}_{\bar x,\bar v}\} \cup \{\tilde\delta\}\Big\} \qquad \text{and} \qquad \rho = \max\Big\{\rho_\tau,\ \tau \in \mathscr{S}_{\bar x,\bar v}\Big\}.$$

Step 2: Definition of variables. Consider X, Y, $V \in \mathbf{S}^n$ such that

$$\|Y - \bar X\| \le \delta,\ \|X - \bar X\| \le \delta,\ V \in \partial F(X),\ F(X) \le F(\bar X) + \delta \text{ and } \|V - \bar V\| \le \delta$$

and set

$$\mathscr{F}(X, V, Y) = F(X) + \langle V, Y - X\rangle - \frac{\rho}{2}\|Y - X\|^2 .$$

Our aim is to prove that $\mathscr{F}(X, V, Y) \le F(Y)$. To this end, we set $x = \lambda(X)$ and $y = \lambda(Y)$, and we introduce $v \in \partial f(x)$ and $U \in \mathbf{O}^n_X$ such that $V = U^\top(\mathrm{Diag}\,v)U$. Observe that, by (4), we have $x, y \in B(\bar x, \delta)$; by the property of $F = f \circ \lambda$, we have $f(x) \le f(\bar x) + \delta$; and by Theorem 3.6, we have

$$\|v - \tau\bar v\| \le \delta \quad (\tau \in \mathscr{S}_{\bar x,\bar v}) . \tag{14}$$

Moreover, since X and V admit a simultaneous spectral decomposition, there exists $\sigma \in \mathbf{\Sigma}^n$ such that

$$\lambda_{\sigma(i)}(V + \rho X) = v_i + \rho x_i , \tag{15}$$

where $x = (x_1, \dots, x_n)$ and $v = (v_1, \dots, v_n)$.

Step 3: Final argument. Let us note that

$$\mathscr{F}(X, V, Y) = F(X) - \frac{\rho}{2}(\|X\|^2 + \|Y\|^2) - \langle V, X\rangle + \langle V + \rho X, Y\rangle , \tag{16}$$

and let us observe that

$$F(X) - \frac{\rho}{2}(\|X\|^2 + \|Y\|^2) = f(x) - \frac{\rho}{2}(\|x\|^2 + \|y\|^2). \tag{17}$$

On the other hand, the term $\langle V, X\rangle$ in (16) can be rewritten as follows:

$$\langle V, X\rangle = \langle UVU^\top, UXU^\top\rangle = \langle \mathrm{Diag}\,v, \mathrm{Diag}\,x\rangle = v^\top x. \tag{18}$$

Let us now focus on the term $\langle V + \rho X, Y\rangle$. Using Theorem 1.5, we deduce that

$$\langle V + \rho X, Y\rangle \ \le\ \lambda(V + \rho X)^\top \lambda(Y),$$

and after rearranging the sum by means of the permutation σ given in (15) we obtain

$$\langle V + \rho X, Y\rangle \ \le\ \sum_{i=1}^n (v_i + \rho x_i)^\top y_{\sigma(i)} = v^\top(\sigma y) + \rho x^\top(\sigma y) . \tag{19}$$

Then combining (16) with (17), (18) and (19) we deduce

$$\mathscr{F}(X, V, Y) \leq f(x) + v^\top(\sigma y - x) - \frac{\rho}{2}(\|x\|^2 + \|y\|^2 - 2x^\top(\sigma y)),$$

which, in view of $\|y\| = \|\sigma y\|$ yields

$$\mathscr{F}(X, V, Y) \leq f(x) + v^\top(\sigma y - x) - \frac{\rho}{2}\|\sigma y - x\|^2 .$$

We conclude applying (9): from (14), the prox-regularity of f at $\bar{x}$ for $\tau\bar{v}$ yields

$$\mathscr{F}(X, V, Y) \leq f(\sigma y) = f(y) = F(Y) .$$

The proof is complete. □

References

[1] D. Aussel, A. Daniilidis, L. Thibault: Subsmooth sets: functional characterizations and related concepts, Trans. Amer. Math. Soc. 357(4) (2005) 1275–1301.

[2] F. Bernard, L. Thibault: Uniform prox-regularity of functions and epigraphs in Hilbert spaces, Nonlinear Anal., Theory Methods Appl. 60A(2) (2005) 187–207.

[3] J. M. Borwein, A. S. Lewis: Convex Analysis and Nonlinear Optimization, 2nd Ed., Springer, New York (2006).

[4] F. Clarke, R. Stern, P. Wolenski: Proximal smoothness and the lower-C^2 property, J. Convex Analysis 2(1-2) (1995) 117–144.

[5] J. Dadok: On the C^∞ Chevalley's theorem, Adv. Math. 44 (1982) 121–131.

[6] A. Daniilidis, W. Hare, J. Malick: Geometrical interpretation of proximal-type algorithms in structured nonsmooth optimization, Optimization 55(5-6) (2006) 481–503.

[7] C. Davis: All convex invariant functions of hermitian matrices, Arch. Math. 8 (1957) 276–278.

[8] W. Hare, A. S. Lewis: Identifying active constraints via partial smoothness and prox-regularity, J. Convex Analysis 11(2) (2004) 251–266.

[9] R. A. Horn, Ch. R. Johnson: Matrix Analysis, Cambridge University Press, Cambridge (1989); new Ed. (1999).

[10] A. S. Lewis, J. Malick: Alternating projections on manifolds, Math. Oper. Res. 33(1) (2008) 216–234.

[11] A. S. Lewis: Convex analysis on the Hermitian matrices, SIAM J. Optim. 6 (1996) 164–177.

[12] A. S. Lewis: Derivatives of spectral functions, Math. Oper. Res. 21 (1996) 576–588.

[13] A. S. Lewis: Nonsmooth analysis of eigenvalues, Math. Program. 84(1) (1999) 1–24.

[14] A. S. Lewis, H. S. Sendov: Twice differentiable spectral functions, SIAM J. Matrix Anal. Appl. 23(2) (2001) 368–386.

[15] M. K. H. Fan, N.-K. Tsing, E. I. Verriest: On analyticity of functions involving eigenvalues, Linear Algebra Appl. 207 (1994) 159–180.

[16] R. A. Poliquin, R. T. Rockafellar: Prox-regular functions in variational analysis, Trans. Amer. Math. Soc. 348 (1996) 1805–1838.

[17] R. A. Poliquin, R. T. Rockafellar, L. Thibault: Local differentiability of distance functions, Trans. Amer. Math. Soc. 352 (2000) 5231–5249.

[18] R. T. Rockafellar: Favorable classes of Lipschitz continuous functions in subgradient optimization, in: Progress in Non-Differentiable Optimization, IIASA Collaborative Proceedings Series, International Institute of Applied Systems Analysis, Laxenburg, Austria, E. Nurminski (ed.), (1982) 125–143.

[19] R. T. Rockafellar, R. J.-B. Wets: Variational Analysis, Springer, Berlin (1998).