SIAM J. MATRIX ANAL. APPL.
Vol. 35, No. 2, pp. 580–598

ORTHOGONAL INVARIANCE AND IDENTIFIABILITY*

A. DANIILIDIS†, D. DRUSVYATSKIY‡, AND A. S. LEWIS§

Abstract. Matrix variables are ubiquitous in modern optimization, in part because variational properties of useful matrix functions often expedite standard optimization algorithms. Convexity is one important such property: permutation-invariant convex functions of the eigenvalues of a symmetric matrix are convex, leading to the wide applicability of semidefinite programming algorithms. We prove the analogous result for the property of "identifiability," a notion central to many active-set-type optimization algorithms.

Key words. eigenvalues, symmetric matrix, partial smoothness, identifiable set, polyhedra, duality

AMS subject classifications. 15A18, 53B25, 15A23, 05A05

DOI. 10.1137/130916710

1. Introduction.

The explosion of interest over recent years in semidefinite programming—optimization involving semidefinite matrix variables—derives strongly from the convexity of many important functions of the eigenvalues of a symmetric matrix. This convexity underpins the success of semidefinite programming algorithms.

To be more specific, consider *spectral functions*. These are functions F, defined on the space of symmetric matrices $\mathbf{S}^n$, that depend on matrices only through their eigenvalues, that is, functions that are invariant under the action of the orthogonal group by conjugation. Spectral functions can always be written as the composition $F = f \circ \lambda$, where f is a permutation-invariant function on $\mathbf{R}^n$, and λ is the mapping assigning to each matrix $X \in \mathbf{S}^n$ the vector of its eigenvalues $(\lambda_1(X), \ldots, \lambda_n(X))$ in nonincreasing order; see [3, section 5.2]. Notable examples of functions fitting in this category are $X \mapsto \lambda_1(X)$ and $X \mapsto \sum_{i=1}^n |\lambda_i(X)|$. Though the spectral mapping λ is not straightforward in behavior (being, in particular, nonsmooth), the spectral function F inherits convexity from the underlying function f [22, 14], a fact closely related to von Neumann's theorem on unitarily invariant matrix norms [35].

Convexity opens to matrix optimization a wide arena of algorithms. We study here another variational concept underpinning a broad range of optimization algorithms of "active set" type: the idea of "identifiability." To introduce this notion, for simplicity consider a minimizer $\bar{x} \in \mathbf{R}^n$ for a continuous but nonsmooth convex function f. We say that a sequence $x_i \to \bar{x}$ "approaches criticality" if each point x_i minimizes a perturbed function $f + \langle y_i, \cdot \rangle$ for vectors $y_i \to 0$ in $\mathbf{R}^n$. We call a set $M \subset \mathbf{R}^n$ "identifiable" if every sequence approaching criticality eventually lies in M.

*Received by the editors April 12, 2013; accepted for publication (in revised form) by P.-A. Absil February 7, 2014; published electronically May 13, 2014.
http://www.siam.org/journals/simax/35-2/91671.html

†Departament de Matemàtiques, C1/364, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Spain (on leave), and DIM-CMM, Universidad de Chile, Santiago, Chile (arisd@dim.uchile.cl, http://www.dim.uchile.cl/∼arisd/). This author was supported by grant MTM2011-29064-C01 (Spain) and FONDECYT Regular 1130176 (Chile).

‡Department of Combinatorics and Optimization, University of Waterloo, Ontario N2L 3G1, Canada (dd379@cornell.edu, http://people.orie.cornell.edu/dd379/). This author was partially supported by the NDSEG grant from the Department of Defense.

§School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853 (aslewis@orie.cornell.edu, http://people.orie.cornell.edu/aslewis/). This author was supported in part by National Science Foundation grant DMS-1208338.

The active-set philosophy in optimization amounts to "identifying" such a set M that is also a manifold on which the restriction $f|_M$ is smooth, and then applying smooth techniques to minimize the restriction.

Our aim here is to show how this identifiability approach "lifts" through the spectral mapping λ, in much the same way as we observed above for convexity. Specifically, given a matrix $\bar{X} \in \mathbf{S}^n$, we show that, if, around the vector $\lambda(\bar{X})$, the function f is "locally symmetric" (meaning locally invariant under permutations fixing that vector), and if an identifiable set M for f is also locally symmetric, then the inverse image $\lambda^{-1}(M)$ is identifiable at $\bar{X}$ for the spectral function $f \circ \lambda$. Furthermore, we show that smoothness of the restriction $f|_M$ lifts to smoothness of $(f \circ \lambda)|_{\lambda^{-1}(M)}$.

The existence of an identifiable manifold is equivalent to the notion of *partial smoothness* discussed at length in [23, 20, 36, 16]. The property holds very commonly: in particular, it is generic in convex semialgebraic optimization [2]. More concretely, when minimizing a polyhedral function f, an identifiable manifold always exists. Since many common spectral functions (like the two examples above) can be written in the composite form $f \circ \lambda$, where f is a permutation-invariant *polyhedral* function, they are amenable to optimization using active-set techniques, via identifiability. One immediate application is to nonlinear eigenvalue optimization problems of the form $\min_x \lambda_1\big(C(x)\big)$, where $C(\cdot)$ is a smoothly parametrized symmetric matrix. The problem of estimating the corresponding identifiable manifold is discussed in [26, Example 4.5].

We also develop parallel results for orthogonally invariant functions of nonsymmetric matrices, the role of eigenvalues being taken by singular values. An interesting application is the approach to low-rank solutions X to a linear matrix equation $\mathcal{A}(X) = b$ via nuclear norm regularization:

$$\min_X \|\mathcal{A}(X) - b\|^2 + \rho\|X\|_*.$$

Since the nuclear norm $\|\cdot\|_*$ is a polyhedral function (namely the l_1 norm) of the vector of singular values, this objective function always has an identifiable manifold M. A simple proximal minimization algorithm (requiring only a singular value decomposition at each iteration) generates iterates that must eventually lie on M, opening the possibility of acceleration techniques: for further discussion, see [25].

Our result sits in a broader context. In the language above, the *transfer principle* asserts that the spectral function $F = f \circ \lambda$ inherits many geometric (more generally variational analytic) properties of f, or, equivalently, F inherits many properties of its restriction to diagonal matrices. For example, when f is a permutation-invariant norm, then F is an orthogonally invariant norm on the space of symmetric matrices—a special case of von Neumann's theorem on unitarily invariant matrix norms [35]. The collection of properties known to satisfy this principle is striking: prox-regularity [9], Clarke-regularity [24, 22], smoothness [22, 21, 32, 10, 34, 33], algebraicity [10], and stratifiability [15, Theorem 4.8]. In this work, as we have explained, we add identifiability (and partial smoothness) to the list (Proposition 3.15 and Theorem 3.19).

One of our intermediary theorems is of particular interest. We observe that the main result of [30] immediately implies that a permutation-invariant set M is a C^p manifold if and only if the spectral set $\lambda^{-1}(M)$ is a C^p manifold (for any $p = 2, \ldots, \infty, \omega$); see Theorem 2.7. This nicely complements and simplifies the recent results of [10, 11, 12].

The outline of the manuscript is as follows. In section 2 we establish some basic notation and discuss the spectral lifting property for smooth functions and manifolds. In section 3 we prove the lifting property for identifiable sets and partly smooth manifolds, while in section 4 we explore duality theory of partly smooth manifolds. Section 5 illustrates how our results have natural analogues for nonsymmetric matrices.

2. Spectral functions and lifts of manifolds.

2.1. Notation. Throughout, the symbol $\mathbf{E}$ will denote a Euclidean space (by which we mean a finite-dimensional real inner-product space). The functions that we will be considering will take their values in the extended real line $\overline{\mathbf{R}} := \mathbf{R} \cup \{-\infty, \infty\}$. We will always assume that the functions under consideration are *proper*, meaning they are never $\{-\infty\}$ and are not always $\{+\infty\}$. For a set $Q \subset \mathbf{E}$, the *indicator function* $\delta_Q \colon \mathbf{E} \to \overline{\mathbf{R}}$ is a function that takes the value 0 on Q and $+\infty$ outside of Q. An open ball of radius ϵ around a point $\bar{x}$ will be denoted by $B_\epsilon(\bar{x})$, while the open unit ball will be denoted by $\mathbf{B}$.

To simplify notation, real-analytic functions on $\mathbf{E}$ will be called C^ω-smooth. Given any set $Q \subset \mathbf{E}$ and a mapping $F \colon Q \to \widetilde{Q}$, where $\widetilde{Q}$ is a subset of some other Euclidean space $\mathbf{H}$, we say that F is C^p*-smooth* (for $p = 2, \ldots, \infty, \omega$) if for each point $\bar{x} \in Q$, there is a neighborhood U of $\bar{x}$ and a C^p-smooth mapping $\widehat{F} \colon \mathbf{E} \to \mathbf{H}$ that agrees with F on $Q \cap U$. A subset $M \subset \mathbf{E}$, where $\mathbf{E}$ is d-dimensional, is a C^p *manifold of dimension* r if for each point $\bar{x} \in M$, there is an open neighborhood U around $\bar{x}$ and a mapping F from $\mathbf{E}$ to a $(d-r)$-dimensional Euclidean space so that F is C^p-smooth with the derivative $\nabla F(\bar{x})$ having full rank and we have $M \cap U = \{x \in U : F(x) = 0\}$. In this case, the range of the adjoint of $\nabla F(\bar{x})$ is the *normal space* to M at $\bar{x}$ and will be denoted by $N_M(\bar{x})$.

Two particular realizations of $\mathbf{E}$ will be important for us, namely $\mathbf{R}^n$ and the space $\mathbf{S}^n$ of $n \times n$-symmetric matrices. Throughout, we will fix an orthogonal basis of $\mathbf{R}^n$, along with an inner product $\langle \cdot, \cdot \rangle$. The corresponding norm will be written as $\|\cdot\|$. The group of permutations of coordinates of $\mathbf{R}^n$ will be denoted by Σ^n, while an application of a permutation $\sigma \in \Sigma^n$ to a point $x \in \mathbf{R}^n$ will simply be written as σx. We denote by $\mathbf{R}^n_\geq$ the set of all points $x \in \mathbf{R}^n$ with $x_1 \geq x_2 \geq \cdots \geq x_n$. A function $f \colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is said to be *symmetric* if we have $f(x) = f(\sigma x)$ for every $x \in \mathbf{R}^n$ and every $\sigma \in \Sigma^n$.

The vector space of real $n \times n$-symmetric matrices $\mathbf{S}^n$ will always be endowed with the trace inner product $\langle X, Y \rangle = \mathrm{tr}\,(XY)$, while the associated norm (Frobenius norm) will be denoted by $\|\cdot\|_F$. The group of orthogonal $n \times n$ matrices will be denoted by $\mathbf{O}^n$. Note that the group of permutations Σ^n naturally embeds in $\mathbf{O}^n$. The action of $\mathbf{O}^n$ by conjugation on $\mathbf{S}^n$ will be written as $U.X := U^T X U$, for matrices $U \in \mathbf{O}^n$ and $X \in \mathbf{S}^n$. A function $F \colon \mathbf{S}^n \to \overline{\mathbf{R}}$ is said to be *spectral* if we have $F(X) = F(U.X)$ for every $X \in \mathbf{S}^n$ and every $U \in \mathbf{O}^n$.

2.2. Smooth transfer principle. We can now consider the spectral mapping $\lambda \colon \mathbf{S}^n \to \mathbf{R}^n$, which simply maps symmetric matrices to the vector of their eigenvalues in nonincreasing order. Then a function on $\mathbf{S}^n$ is spectral if and only if it can be written as a composition $f \circ \lambda$, for some symmetric function $f \colon \mathbf{R}^n \to \overline{\mathbf{R}}$. (See, for example [22, Proposition 4].) As was mentioned in the introduction, the *transfer principle* asserts that a number of variational-analytic properties hold for the spectral function $f \circ \lambda$ if and only if they hold for f. Evidently, analogous results are valid even when f is only *locally symmetric* (to be defined below). The proofs follow by a reduction to the symmetric case by simple symmetrization arguments, and hence we will omit details in the current paper. The goal of the current section is to note that the transfer principle, in particular, holds for smooth functions and for smooth manifolds.

For each point $x \in \mathbf{R}^n$, we consider the *stabilizer*

$$\mathrm{Fix}(x) := \{\sigma \in \Sigma^n : \sigma x = x\}.$$

The following notion, borrowed from [10], will be key for us.

DEFINITION 2.1 (local symmetry). *A function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is* locally symmetric *at a point $\bar{x} \in \mathbf{R}^n$ if we have $f(x) = f(\sigma x)$ for all points x near $\bar{x}$ and all permutations $\sigma \in \mathrm{Fix}(\bar{x})$.*

A set $Q \subset \mathbf{R}^n$ is symmetric (respectively, locally symmetric) if the indicator function δ_Q is symmetric (respectively, locally symmetric). The following shows that smooth functions satisfy the transfer principle [33, 34].

THEOREM 2.2 (transfer principle for smooth functions). *Consider a function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and a matrix $\overline{X} \in \mathbf{S}^n$. Suppose that f is locally symmetric around $\bar{x} := \lambda(\overline{X})$. Then f is C^p-smooth around $\bar{x}$ (for $p = 2, \ldots, \infty, \omega$) if and only if the spectral function $f \circ \lambda$ is C^p-smooth around $\overline{X}$.*

Shortly, we will need a slightly strengthened version of Theorem 2.2, where smoothness is considered only relative to a certain locally symmetric subset. We record it now.

COROLLARY 2.3 (transfer principle for restricted smooth functions). *Consider a function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, a matrix $\overline{X} \in \mathbf{S}^n$, and a set $M \subset \mathbf{R}^n$ containing $\bar{x} := \lambda(\overline{X})$. Suppose that f and M are locally symmetric around $\bar{x}$. Then the restriction of f to M is C^p-smooth ($p = 2, \ldots, \infty, \omega$) around $\bar{x}$ if and only if the restriction of $f \circ \lambda$ to $\lambda^{-1}(M)$ is C^p-smooth around $\overline{X}$.*

Proof. Suppose that the restriction of f to M is C^p-smooth around $\bar{x}$. Then there exists a C^p-smooth function $\tilde{f}$, defined on $\mathbf{R}^n$, and agreeing with f on M near $\bar{x}$. Consider then the symmetrized function

$$\tilde{f}_{\mathrm{sym}}(x) := \frac{1}{|\mathrm{Fix}(\bar{x})|} \sum_{\sigma \in \mathrm{Fix}(\bar{x})} \tilde{f}(\sigma x),$$

where $|\mathrm{Fix}(\bar{x})|$ denotes the cardinality of the set $\mathrm{Fix}(\bar{x})$. Clearly $\tilde{f}_{\mathrm{sym}}$ is C^p-smooth, locally symmetric around $\bar{x}$, and moreover it agrees with f on M near $\bar{x}$. Finally, using Theorem 2.2, we deduce that the spectral function $\tilde{f}_{\mathrm{sym}} \circ \lambda$ is C^p-smooth around $\overline{X}$ and it agrees with $f \circ \lambda$ on $\lambda^{-1}(M)$ near $\overline{X}$. This proves the forward implication of the corollary.

To see the converse, define $F := f \circ \lambda$, and suppose that the restriction of F to $\lambda^{-1}(M)$ is C^p-smooth around $\overline{X}$. Then there exists a C^p-smooth function $\widetilde{F}$, defined on $\mathbf{S}^n$, and agreeing with F on $\lambda^{-1}(M)$ near $\overline{X}$. Consider then the function

$$\widetilde{F}_{\mathrm{sym}}(X) := \frac{1}{|\mathbf{O}^n|} \sum_{U \in \mathbf{O}^n} \widetilde{F}(U.X),$$

where $|\mathbf{O}^n|$ denotes the cardinality of the set $\mathbf{O}^n$. Clearly $\widetilde{F}_{\mathrm{sym}}$ is C^p-smooth, spectral, and it agrees with F on $\lambda^{-1}(M)$ near $\overline{X}$. Since $\widetilde{F}_{\mathrm{sym}}$ is spectral, we deduce that there is a symmetric function $\tilde{f}$ on $\mathbf{R}^n$ satisfying $\widetilde{F}_{\mathrm{sym}} = \tilde{f} \circ \lambda$. Theorem 2.2 then implies that $\tilde{f}$ is C^p-smooth. Hence to complete the proof, all we have to do is verify that

$\tilde{f}$ agrees with f on M near $\bar{x}$. To this end, consider a point $x \in M$ near $\bar{x}$ and choose a permutation $\sigma \in \mathrm{Fix}(\bar{x})$ satisfying $\sigma x \in \mathbf{R}^n_{\geq}$. Let $U \in \mathbf{O}^n$ be such that $\overline{X} = U^T(\mathrm{Diag}\,\bar{x})U$. Then we have

$$\tilde{f}(x) = \tilde{f}(\sigma x) = \widetilde{F}_{\mathrm{sym}}\big(U^T(\mathrm{Diag}\,x)U\big) = F\big(U^T(\mathrm{Diag}\,x)U\big) = f(\sigma x) = f(x),$$

as claimed. ∎

Our next goal is to observe that smooth manifolds also satisfy the transfer principle. One path to such a result proceeds by "lifting the defining equations"; see [10, 11, 12]. This strategy, however, encounters serious difficulties. An alternate idea, which we pursue here, is to instead consider the distance function. To this end, recall that the *distance* of a point x to a set $Q \subset \mathbf{E}$ is simply

$$d_Q(x) := \inf\{\|x - y\| : y \in Q\},$$

whereas the *metric projection* of x onto Q is defined by

$$P_Q(x) := \{y \in Q : \|x - y\| = d_Q(x)\}.$$

The following theorem provides a key link between the order of smoothness of a set and that of the squared distance function [30].[1]

Theorem 2.4 (smoothness of sets and the distance function). *Consider a set $Q \subset \mathbf{E}$ that is locally closed around a point $\bar{x} \in Q$. Then Q is a C^p manifold near $\bar{x}$ (for $p = 2, \ldots, \infty, \omega$) if and only if the function d_Q^2 is C^p-smooth near $\bar{x}$.*

Remark 2.5. It is quite illuminating to outline the proof of Theorem 2.4. If Q is a C^p manifold near $\bar{x}$, then it is easy to see, using the implicit function theorem, that the projection P_Q is C^{p-1} smooth near $\bar{x}$, and consequently that d_Q^2 is C^p-smooth near $\bar{x}$. The converse is slightly more delicate. If d_Q^2 is C^p-smooth near $\bar{x}$, then the projection P_Q is C^{p-1} smooth near $\bar{x}$. Moreover, one can then show that P_Q has constant rank on a neighborhood of $\bar{x}$ and therefore that Q is a C^{p-1} manifold near $\bar{x}$. Now, carefully writing out the definition of P_Q in terms of the C^{p-1}-smooth defining equations of M, one easily sees that the defining equations are in fact C^p-smooth.

Finally, the following theorem shows that distance functions interact well with the eigenvalue mapping [9, Proposition 2.3, Theorem 2.4].

Theorem 2.6 (distance functions and the spectral mapping). *Consider a matrix $\overline{X} \in \mathbf{S}^n$ and a set $Q \subset \mathbf{R}^n$ that is locally symmetric around the point $\bar{x} := \lambda(\overline{X})$. Then the function d_Q is locally symmetric near $\bar{x}$ and the distance to the spectral set $\lambda^{-1}(Q)$ satisfies*

$$d_{\lambda^{-1}(Q)} = d_Q \circ \lambda \quad \textit{locally around } \overline{X}.$$

Combining the previous two theorems, the following is immediate.

Theorem 2.7 (transfer principle for smooth manifolds). *Consider a matrix $\overline{X} \in \mathbf{S}^n$ and a set $M \subset \mathbf{R}^n$ that is locally closed and locally symmetric around $\bar{x} := \lambda(\overline{X})$. Then M is a C^p manifold around $\bar{x}$ (for $p = 2, \ldots, \infty, \omega$) if and only if the spectral set $\lambda^{-1}(M)$ is a C^p manifold around $\overline{X}$.*

Proof. By Theorem 2.4, the set M is a C^p manifold around $\bar{x}$ if and only if the function d_M^2 is C^p-smooth near $\bar{x}$. On the other hand, by Theorems 2.2 and 2.6, the latter occurs if and only if the function $d_M^2 \circ \lambda = d_{\lambda^{-1}(M)}^2$ is C^p-smooth near $\overline{X}$. Applying Theorem 2.4 once again, the result follows. ∎

[1]The original version of the current manuscript included a proof of this theorem. During the reviewing process, however, we became aware of [30] where the same result had appeared earlier.

2.3. Dimension of the lifted manifold. The proof of Theorem 2.7 is very short (in light of Theorem 2.4), unlike the involved proof of [10]. One shortcoming, however, is that it does not a priori yield information about the dimension of the lifted manifold $\lambda^{-1}(M)$. In this section, we outline how we can use the fact that $\lambda^{-1}(M)$ is a manifold to establish a formula between the dimensions of M and $\lambda^{-1}(M)$. This section can safely be skipped upon first reading.

We adhere closely to the notation and some of the combinatorial arguments of [10]. With any point $x \in \mathbf{R}^n$ we associate a partition $\mathcal{P}_x = \{I_1, \ldots, I_\rho\}$ of the set $\{1, \ldots, n\}$, whose elements are defined as follows:

$$i, j \in I_\ell \Longleftrightarrow x_i = x_j.$$

It follows readily that for $x \in \mathbf{R}^n_{\geq}$ there exists a sequence

$$1 = i_0 \leq i_1 < \cdots < i_\rho = n$$

such that

$$I_\ell = \{i_{\ell-1}, \ldots, i_\ell - 1\} \quad \text{for each } \ell \in \{1, \ldots, \rho\}.$$

For any such partition $\mathcal{P}$ we set

$$\Delta_{\mathcal{P}} := \{x \in \mathbf{R}^n_{\geq} : \mathcal{P}_x = \mathcal{P}\}.$$

As explained in [10, section 2.2], the set of all such $\Delta_{\mathcal{P}}$'s defines an affine stratification of $\mathbf{R}^n_{\geq}$. Observe further that for every point $x \in \mathbf{R}^n_{\geq}$ we have

$$\lambda^{-1}(x) = \{U^T (\operatorname{Diag} x) U : U \in \mathbf{O}^n\}.$$

Let $\mathbf{O}^n_X := \{U \in \mathbf{O}^n : U^T X U = X\}$ denote the stabilizer of X, which is a C^∞ manifold of dimension

$$\dim \mathbf{O}^n_X \;=\; \dim \left(\prod_{1 \leq \ell \leq \rho} \mathbf{O}^{|I_\ell|} \right) = \sum_{1 \leq \ell \leq \rho} \frac{|I_\ell|\,(|I_\ell| - 1)}{2},$$

as one can easily check. Since the orbit $\lambda^{-1}(x)$ is isomorphic to $\mathbf{O}^n / \mathbf{O}^n_X$, it follows that it is a submanifold of $\mathbf{S}^n$. A computation, which can be found in [10], then yields the equation

$$\dim\, \lambda^{-1}(x) = \dim \mathbf{O}^n - \dim \mathbf{O}^n_X = \sum_{1 \leq i < j \leq \rho} |I_i|\,|I_j|.$$

Consider now any locally symmetric manifold M of dimension d. There is no loss of generality to assume that M is connected and has nonempty intersection with $\mathbf{R}^n_{\geq}$. Let us further denote by Δ_* an affine stratum of the aforementioned stratification of $\mathbf{R}^n_{\geq}$ with the property that its dimension is maximal among all of the strata Δ enjoying a nonempty intersection with M. It follows that there exists a point $\bar{x} \in M \cap \Delta_*$ and $\delta > 0$ satisfying $M \cap B(\bar{x}, \delta) \subset \Delta_*$ (see [10, section 3] for details). Since $\dim\, \lambda^{-1}(M) = \dim\, \lambda^{-1}(M \cap B(\bar{x}, \delta))$ and since $\lambda^{-1}(M \cap B(\bar{x}, \delta))$ is a fibration, we obtain

$$\dim\, \lambda^{-1}(M) \;=\; \dim\, M \;+ \sum_{1 \leq i < j \leq \rho_*} |I^*_i|\,|I^*_j|, \tag{2.1}$$

where $\mathcal{P}_* = \{I_1^*, \ldots, I_\rho^*\}$ is the partition associated to $\bar{x}$ (or, equivalently, to any $x \in \Delta_*$).

Remark 2.8. It's worth pointing out that it is possible to have strata $\Delta_1 \neq \Delta_2$ of $\mathbf{R}^n_{\geq}$ of the same dimension, but giving rise to stabilizers of different dimension for their elements. The argument above shows that a connected locally symmetric manifold cannot intersect simultaneously with these strata. This also follows implicitly from the forthcoming Lemma 4.4, asserting the connectedness of $\lambda^{-1}(M)$, whenever M is connected.

3. Spectral lifts of identifiable sets and partly smooth manifolds. We begin this section by summarizing some of the basic tools used in variational analysis and nonsmooth optimization. We refer the reader to the monographs of Borwein and Zhu [4], Clarke et al. [8], Mordukhovich [27], and Rockafellar and Wets [31] for more details. We adhere closely to the terminology and notation of [31].

3.1. Variational analysis of spectral functions. For a function $f \colon \mathbf{E} \to \overline{\mathbf{R}}$, the *domain* of f is

$$\operatorname{dom} f := \{x \in \mathbf{E} : f(x) < +\infty\},$$

and the *epigraph* of f is

$$\operatorname{epi} f := \{(x, r) \in \mathbf{E} \times \mathbf{R} : r \geq f(x)\}.$$

We will say that f is *lower semicontinuous* (*lsc*) at a point $\bar{x}$ provided that the inequality $\liminf_{x \to \bar{x}} f(x) \geq f(\bar{x})$ holds. If f is lsc at every point, then we will simply say that f is lsc. For any set Q, the symbols $\operatorname{cl} Q$, $\operatorname{conv} Q$, and $\operatorname{aff} Q$ will denote the topological closure, the convex hull, and the affine span of Q, respectively. The symbol $\operatorname{par} Q$ will denote the parallel subspace of Q, namely the set $\operatorname{par} Q := (\operatorname{aff} Q) - Q$. For convex sets $Q \subset \mathbf{E}$, the symbols $\operatorname{ri} Q$ and $\operatorname{rb} Q$ will denote the relative interior and the relative boundary of Q, respectively.

Subdifferentials are the primary variation-analytic tools for studying general nonsmooth functions f on $\mathbf{E}$.

Definition 3.1 (subdifferentials). *Consider a function $f \colon \mathbf{E} \to \overline{\mathbf{R}}$ and a point $\bar{x}$ with $f(\bar{x})$ finite.*

(i) *The* Fréchet subdifferential *of f at $\bar{x}$, denoted $\hat{\partial} f(\bar{x})$, consists of all vectors $v \in \mathbf{E}$ satisfying*

$$f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle + o(\|x - \bar{x}\|).$$

(ii) *The* limiting subdifferential *of f at $\bar{x}$, denoted $\partial f(\bar{x})$, consists of all vectors $v \in \mathbf{E}$ for which there exist sequences $x_i \in \mathbf{E}$ and $v_i \in \hat{\partial} f(x_i)$ with $(x_i, f(x_i), v_i) \to (\bar{x}, f(\bar{x}), v)$.*

We now recall from [22, Proposition 2] the following lemma, which shows that subdifferentials behave as one would expect in the presence of symmetry.

Lemma 3.2 (subdifferentials under symmetry). *Consider a function $f \colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and a point $\bar{x} \in \mathbf{R}^n$. Then the equation*

$$\partial f(\sigma x) = \sigma \partial f(x) \quad \text{holds for any } \sigma \in \operatorname{Fix}(\bar{x}) \text{ and all } x \text{ near } \bar{x}.$$

Similarly, in terms of the spectral function $F := f \circ \lambda$, we have

$$\partial F(U.X) = U.(\partial F(X)) \quad \text{for any } U \in \mathbf{O}^n \text{ and any } X \in \mathbf{S}^n.$$

Remark 3.3. In particular, if $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is locally symmetric around $\bar{x}$, then the sets $\hat{\partial} f(\bar{x})$, $\operatorname{ri} \hat{\partial} f(\bar{x})$, $\operatorname{rb} \hat{\partial} f(\bar{x})$, $\operatorname{aff} \hat{\partial} f(\bar{x})$, and $\operatorname{par} \hat{\partial} f(\bar{x})$ are invariant under the action of the group $\operatorname{Fix}(\bar{x})$.

The following result is the cornerstone for the variational theory of spectral mappings [22, Theorem 6].

THEOREM 3.4 (subdifferential under local symmetry). *Consider an lsc function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and a matrix $X \in \mathbf{S}^n$, and suppose that f is locally symmetric at $\lambda(X)$. Then we have*

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

where

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

It is often useful to require a certain uniformity of the subgradients of the function. This is the content of the following definition [29, Definition 1.1].

DEFINITION 3.5 (prox-regularity). *A function $f\colon \mathbf{E} \to \overline{\mathbf{R}}$ is called* prox-regular at $\bar{x}$ for $\bar{v}$, *with $\bar{v} \in \partial f(\bar{x})$, if f is locally lsc at $\bar{x}$ and there exist $\epsilon > 0$ and $\rho > 0$ so that the inequality*

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

holds whenever $x, y \in B_\epsilon(\bar{x})$, $v \in B_\epsilon(\bar{v}) \cap \partial f(x)$, and $f(x) < f(\bar{x}) + \epsilon$. The function f is called prox-regular at $\bar{x}$, *if it is finite at $\bar{x}$ and f is prox-regular at $\bar{x}$ for every subgradient $v \in \partial f(\bar{x})$.*

In particular C^2-smooth functions and lsc, convex functions are prox-regular at each of their points [31, Example 13.30, Proposition 13.34]. In contrast, the negative norm function $x \mapsto -\|x\|$ is not prox-regular at the origin. The following theorem shows that prox-regularity also satisfies the transfer principle [9, Theorem 4.2].

THEOREM 3.6 (directional prox-regularity under spectral lifts). *Consider an lsc function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and a symmetric matrix $\overline{X}$. Suppose that f is locally symmetric around $\bar{x} := \lambda(\overline{X})$. Then f is prox-regular at $\bar{x}$ if and only if $f \circ \lambda$ is prox-regular at $\overline{X}$.*

The following standard result of linear algebra will be important for us [22, Proposition 3].

LEMMA 3.7 (simultaneous conjugacy). *Consider vectors $x, y, u, v \in \mathbf{R}^n$. Then there exists an orthogonal matrix $U \in \mathbf{O}^n$ with*

$$\operatorname{Diag} x = U^T(\operatorname{Diag} u)U \quad \text{and} \quad \operatorname{Diag} y = U^T(\operatorname{Diag} v)U,$$

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

The following is a simple consequence.

COROLLARY 3.8 (conjugations and permutations). *Consider vectors $v_1, v_2 \in \mathbf{R}^n$ and a matrix $X \in \mathbf{S}^n$. Suppose that for some $U_1, U_2 \in \mathbf{O}^n_X$ we have*

$$U_1^T(\operatorname{Diag} v_1)U_1 = U_2^T(\operatorname{Diag} v_2)U_2.$$

Then there exists a permutation $\sigma \in \operatorname{Fix}(\lambda(X))$ satisfying $\sigma v_1 = v_2$.

Proof. Observe

$$(U_1U_2^T)^T \operatorname{Diag} v_1 (U_1U_2^T) = \operatorname{Diag} v_2,$$

$$(U_1U_2^T)^T \operatorname{Diag} \lambda(X)(U_1U_2^T) = \operatorname{Diag} \lambda(X).$$

The result follows by an application of Lemma 3.7. ∎

3.2. Main results. In this section, we consider partly smooth sets, introduced in [23]. This notion generalizes the idea of active manifolds of classical nonlinear programming to an entirely representation-independent setting.

DEFINITION 3.9 (partial smoothness). *Consider a function* $f\colon \mathbf{E} \to \overline{\mathbf{R}}$ *and a set* $M \subset \mathbf{E}$ *containing a point* $\bar{x}$. *Then* f *is* C^p-partly smooth *(*$p = 2, \ldots, \infty, \omega$*) at* $\bar{x}$ *relative to* M *if the following hold:*

(i) (Smoothness) M *is a* C^p *manifold around* $\bar{x}$ *and* f *restricted to* M *is* C^p*-smooth near* $\bar{x}$*,*

(ii) (Regularity) f *is prox-regular at* $\bar{x}$*,*

(iii) (Sharpness) *the affine span of* ∂f *is a translate of the normal space* $N_M(x)$*,*

(iv) (Continuity) ∂f *restricted to* M *is continuous at* $\bar{x}$*.*

When the above properties hold, we call M *the* partly smooth manifold *of* f *at* $\bar{x}$*.*

Remark 3.10. Though the original definition of partial smoothness replaces the prox-regularity condition by Clarke-regularity, we feel that the prox-regularity is essential for the theory. In particular, without it, partly smooth manifolds are not even guaranteed to be locally unique and the basic property of identifiability may fail [20, section 7].

Some comments are in order. First, the continuity property of ∂f is meant in the Painlevé–Kuratowski sense. See, for example, [31, Definition 5.4]. The exact details of this notion will not be needed in our work, and hence we do not dwell on it further. Geometrically, partly smooth manifolds have a characteristic property in that the epigraph of f looks "valley-like" along the graph of $f\big|_M$. See Figure 1 for an illustration.

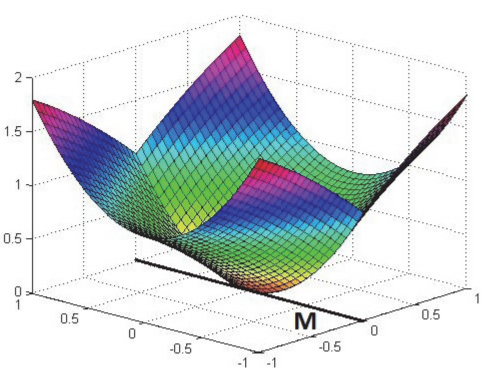

FIG. 1. *The partly smooth manifold* M *for* $f(x,y) := |x|(1-|x|) + y^2$.

It is reassuring to know that partly smooth manifolds are locally unique. This is the content of the following theorem [20, Corollary 4.2].

THEOREM 3.11 (local uniqueness of partly smooth manifolds). *Consider a function* $f\colon \mathbf{E} \to \overline{\mathbf{R}}$ *that is* C^2*-partly smooth at* $\bar{x}$ *relative to two manifolds* M_1 *and* M_2*. Then there exists a neighborhood* U *of* $\bar{x}$ *satisfying* $U \cap M_1 = U \cap M_2$*.*

Our goal in this section is to prove that partly smooth manifolds satisfy the transfer principle. However, proving this directly is rather difficult. This is in large part because the continuity of the subdifferential mapping $\partial(f \circ \lambda)$ seems to be intrinsically tied to continuity properties of the mapping

$$X \mapsto \mathbf{O}^n_X = \{U \in \mathbf{O}^n : X = U^T(\operatorname{Diag}\lambda(X))U\},$$

which are rather difficult to understand.

We, however, will side-step this problem entirely by instead focusing on a property that is seemingly different from partial smoothness—*finite identification.* This notion is of significant independent interest. It has been implicitly considered by a number of authors in connection with the possibility to accelerate various first-order numerical methods [36, 17, 7, 6, 5, 18, 1, 19, 13], and has explicitly been studied in [16] for its own sake.

DEFINITION 3.12 (identifiable sets). *Consider a function $f\colon \mathbf{E} \to \overline{\mathbf{R}}$, a point $\bar{x} \in \mathbf{R}^n$, and a subgradient $\bar{v} \in \partial f(\bar{x})$. A set $M \subset \operatorname{dom} f$ is* identifiable at $\bar{x}$ for $\bar{v}$ *if for any sequences $(x_i, f(x_i), v_i) \to (\bar{x}, f(\bar{x}), \bar{v})$, with $v_i \in \partial f(x_i)$, the points x_i must all lie in M for all sufficiently large indices i.*

Remark 3.13. It is important to note that identifiable sets are not required to be smooth manifolds. Indeed, as we will see shortly, identifiability is a more basic notion than partial smoothness.

The relationship between partial smoothness and finite identification is easy to explain. Indeed, as the following theorem shows, partial smoothness is in a sense just a "uniform" version of identifiability [16, Proposition 9.4].

PROPOSITION 3.14 (partial smoothness and identifiability). *Consider an lsc function $f\colon \mathbf{E} \to \overline{\mathbf{R}}$ that is prox-regular at a point $\bar{x}$. Let $M \subset \operatorname{dom} f$ be a C^p manifold ($p = 2, \ldots, \infty, \omega$) containing $\bar{x}$, with the restriction $f\big|_M$ being C^p-smooth near $\bar{x}$. Then the following are equivalent:*

1. *f is C^p-partly smooth at $\bar{x}$ relative to M,*
2. *M is an identifiable set (relative to f) at $\bar{x}$ for every subgradient $\bar{v} \in \operatorname{ri} \partial f(\bar{x})$.*

In light of the theorem above, our strategy for proving the transfer principle for partly smooth sets is two-fold: first, prove the analogous result for identifiable sets and then gain a better understanding of the relationship between the sets $\operatorname{ri} \partial f(\lambda(X))$ and $\operatorname{ri} \partial (f \circ \lambda)(X)$.

PROPOSITION 3.15 (spectral lifts of identifiable sets). *Consider an lsc function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and a symmetric matrix $\overline{X} \in \mathbf{S}^n$. Suppose that f is locally symmetric around $\bar{x} := \lambda(\overline{X})$ and consider a subset $M \subset \mathbf{R}^n$ that is locally symmetric around $\bar{x}$. Then M is identifiable (relative to f) at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$, if and only if $\lambda^{-1}(M)$ is identifiable (relative to $f \circ \lambda$) at $\overline{X}$ for $U^T(\operatorname{Diag} \bar{v})U \in \partial (f \circ \lambda)(\overline{X})$, where $U \in \mathbf{O}^n_{\overline{X}}$ is arbitrary.*

Proof. We first prove the forward implication. Fix a subgradient

$$\overline{V} := \overline{U}^T(\operatorname{Diag} \bar{v})\overline{U} \in \partial (f \circ \lambda)(\overline{X})$$

for an arbitrary transformation $\overline{U} \in \mathbf{O}^n_{\overline{X}}$ (see Theorem 3.4). For convenience, let $F := f \circ \lambda$ and consider a sequence $(X_i, F(X_i), V_i) \to (\overline{X}, F(\overline{X}), \overline{V})$. Our goal is to show that for all large indices i, the inclusion $\lambda(X_i) \in M$ holds. To this end, there exist matrices $U_i \in \mathbf{O}^n_{X_i}$ and subgradients $v_i \in \partial f(\lambda(X_i))$ with

$$U_i^T(\operatorname{Diag} \lambda(X_i))U_i = X_i \quad \text{and} \quad U_i^T(\operatorname{Diag} v_i)U_i = V_i.$$

Restricting to a subsequence, we may assume that there exists a matrix $\widetilde{U} \in \mathbf{O}^n_{\overline{X}}$ satisfying $U_i \to \widetilde{U}$, and consequently there exists a subgradient $\tilde{v} \in \partial f(\lambda(\overline{X}))$ satisfying $v_i \to \tilde{v}$. Hence we obtain

$$\widetilde{U}^T(\operatorname{Diag} \lambda(\overline{X}))\widetilde{U} = \overline{X} \quad \text{and} \quad \widetilde{U}^T(\operatorname{Diag} \tilde{v})\widetilde{U} = \overline{V} = \overline{U}^T(\operatorname{Diag} \bar{v})\overline{U}.$$

By Corollary 3.8, there exists a permutation $\sigma \in \mathrm{Fix}(\bar{x})$ with $\sigma\tilde{v} = \bar{v}$. Observe $(\lambda(X_i), f(\lambda(X_i)), v_i) \to (\bar{x}, f(\bar{x}), \tilde{v})$. Observe that since M is identifiable at $\bar{x}$ for $\bar{v}$ and f is locally symmetric around $\bar{x}$, the set $\sigma^{-1}M$ is identifiable (relative to f) at $\bar{x}$ for $\tilde{v}$. Consequently, for all large indices i, the inclusion $\lambda(X_i) \in \sigma^{-1}M$ holds. Since M is locally symmetric at $\bar{x}$, we deduce that all the points $\lambda(X_i)$ eventually lie in M.

To see the reverse implication, fix an orthogonal matrix $\overline{U} \in \mathbf{O}^n_{\overline{X}}$ and define $\overline{V} := \overline{U}^T(\mathrm{Diag}\,\bar{v})\overline{U} \in \partial(f \circ \lambda)(\overline{X})$. Consider a sequence $(x_i, f(x_i), v_i) \to (\bar{x}, f(\bar{x}), \bar{v})$ with $v_i \in \partial f(x_i)$. It is not difficult to see then that there exist permutations $\sigma_i \in \mathrm{Fix}(\bar{x})$ satisfying $\sigma_i x_i \in \mathbf{R}^n_{\geq}$. Restricting to a subsequence, we may suppose that σ_i are equal to a fixed $\sigma \in \mathrm{Fix}(\bar{\bar{x}})$. Define

$$X_i := \overline{U}^T(\mathrm{Diag}\,\sigma x_i)\overline{U} \quad \text{and} \quad V_i := \overline{U}^T(\mathrm{Diag}\,\sigma v_i)\overline{U}.$$

Letting $A_{\sigma^{-1}} \in \mathbf{O}^n$ denote the matrix representing the permutation σ^{-1}, we have

$$X_i := (\overline{U}^T A_{\sigma^{-1}}\overline{U})^T\big[\overline{U}^T(\mathrm{Diag}\,x_i)\overline{U}\big]\overline{U}^T A_{\sigma^{-1}}\overline{U} \quad \text{and}$$
$$V_i := (\overline{U}^T A_{\sigma^{-1}}\overline{U})^T[\overline{U}^T(\mathrm{Diag}\,v_i)\overline{U}]\overline{U}^T A_{\sigma^{-1}}\overline{U}.$$

We deduce $X_i \to (\overline{U}^T A_{\sigma^{-1}}\overline{U})^T\overline{X}(\overline{U}^T A_{\sigma^{-1}}\overline{U})$ and $V_i \to (\overline{U}^T A_{\sigma^{-1}}\overline{U})^T\overline{V}(\overline{U}^T A_{\sigma^{-1}}\overline{U})$. On the other hand, observe $\overline{X} = (\overline{U}^T A_{\sigma^{-1}}\overline{U})^T\overline{X}(\overline{U}^T A_{\sigma^{-1}}\overline{U})$. Moreover, we have $(\overline{U}^T A_{\sigma^{-1}}\overline{U})^T\overline{V}(\overline{U}^T A_{\sigma^{-1}}\overline{U}) = \overline{V}$ by Lemma 3.2. Since $\lambda^{-1}(M)$ is identifiable (relative to F) at $\overline{X}$ for $\overline{V}$, we deduce that the matrices X_i lie in $\lambda^{-1}(M)$ for all sufficiently large indices i. Since M is locally symmetric around $\bar{x}$, the proof is complete. ∎

Using the results of section 2, we can now describe in a natural way the affine span, relative interior, and relative boundary of the Fréchet subdifferential. We begin with a lemma.

LEMMA 3.16 (affine generation). *Consider a matrix $X \in \mathbf{S}^n$ and suppose that the point $x := \lambda(X)$ lies in an affine subspace $\mathcal{V} \subset \mathbf{R}^n$ that is invariant under the action of* $\mathrm{Fix}(x)$. *Then the set*

$$\{U^T(\mathrm{Diag}\,v)U : v \in \mathcal{V} \text{ and } U \in \mathbf{O}^n_X\},$$

is an affine subspace of $\mathbf{S}^n$.

Proof. Define the set $L := (\mathrm{par}\,\mathcal{V})^{\perp}$. Observe that the set $L \cap \mathcal{V}$ consists of a single vector; call this vector w. Since both L and $\mathcal{V}$ are invariant under the action of $\mathrm{Fix}(x)$, we deduce $\sigma w = w$ for all $\sigma \in \mathrm{Fix}(x)$.

Now define a function $g \colon \mathbf{R}^n \to \overline{\mathbf{R}}$ by declaring

$$g(y) = \langle w, y\rangle + \delta_{x+L}(y),$$

and note that the equation

$$\hat{\partial} g(x) := w + N_{x+L}(x) = \mathcal{V} \quad \text{holds.}$$

Observe that for any permutation $\sigma \in \mathrm{Fix}(x)$, we have

$$g(\sigma y) = \langle w, \sigma y\rangle + \delta_{x+L}(\sigma y) = \langle \sigma^{-1}w, y\rangle + \delta_{x+\sigma^{-1}L}(y) = g(y).$$

Consequently, g is locally symmetric at x. Observe

$$(g \circ \lambda)(Y) = \langle w, \lambda(Y)\rangle + \delta_{\lambda^{-1}(x+L)}(Y).$$

It is immediate from Theorems 2.2 and 2.7, that the function $Y \mapsto \langle w, \lambda(Y)\rangle$ is C^∞-smooth around X and that $\lambda^{-1}(x+L)$ is a C^∞ manifold around X. Consequently, $\hat{\partial}(g \circ \lambda)(X)$ is an affine subspace of $\mathbf{S}^n$. On the other hand, we have

$$\hat{\partial}(g \circ \lambda)(X) = \{U^T(\operatorname{Diag} v)U : v \in \mathcal{V} \text{ and } U \in \mathbf{O}^n_X\},$$

thereby completing the proof. ∎

Proposition 3.17 (affine span of the spectral Fréchet subdifferential). *Consider a function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and a matrix $X \in \mathbf{S}^n$. Suppose that f is locally symmetric at $\lambda(X)$. Then we have*

$$\operatorname{aff} \hat{\partial}(f \circ \lambda)(X) = \{U^T(\operatorname{Diag} v)U : v \in \operatorname{aff} \hat{\partial} f(\lambda(X)) \textit{ and } U \in \mathbf{O}^n_X\}, \tag{3.1}$$

$$\operatorname{rb} \hat{\partial}(f \circ \lambda)(X) = \{U^T(\operatorname{Diag} v)U : v \in \operatorname{rb} \hat{\partial} f(\lambda(X)) \textit{ and } U \in \mathbf{O}^n_X\}, \tag{3.2}$$

$$\operatorname{ri} \hat{\partial}(f \circ \lambda)(X) = \{U^T(\operatorname{Diag} v)U : v \in \operatorname{ri} \hat{\partial} f(\lambda(X)) \textit{ and } U \in \mathbf{O}^n_X\}. \tag{3.3}$$

Proof. Throughout the proof, let $x := \lambda(X)$. We prove the formulas in the order that they are stated. To this end, observe that the inclusion $\supset$ in (3.1) is immediate. Furthermore, the inclusion

$$\hat{\partial}(f \circ \lambda)(X) \subset \{U^T(\operatorname{Diag} v)U : v \in \operatorname{aff} \hat{\partial} f(\lambda(X)) \text{ and } U \in \mathbf{O}^n_X\}$$

clearly holds. Hence to establish the reverse inclusion in (3.1), it is sufficient to show that the set

$$\{U^T(\operatorname{Diag} v)U : v \in \operatorname{aff} \hat{\partial} f(\lambda(X)) \text{ and } U \in \mathbf{O}^n_X\}$$

is an affine subspace; but this is immediate from Remark 3.3 and Lemma 3.16. Hence (3.1) holds.

We now prove (3.2). Consider a matrix $U^T(\operatorname{Diag} v)U \in \operatorname{rb} \hat{\partial}(f \circ \lambda)(X)$ with $U \in \mathbf{O}^n_X$ and $v \in \hat{\partial} f(\lambda(X))$. Our goal is to show the stronger inclusion $v \in \operatorname{rb} \hat{\partial} f(x)$. Observe from (3.1), there exists a sequence $U_i^T(\operatorname{Diag} v_i)U_i \to U^T(\operatorname{Diag} v)U$ with $U_i \in \mathbf{O}^n_X$, $v_i \in \operatorname{aff} \hat{\partial} f(x)$, and $v_i \notin \hat{\partial} f(x)$. Restricting to a subsequence, we may assume that there exists a matrix $\widetilde{U} \in \mathbf{O}^n_X$ with $U_i \to \widetilde{U}$ and a vector $\tilde{v} \in \operatorname{aff} \hat{\partial} f(x)$ with $v_i \to \tilde{v}$. Hence the equation

$$\widetilde{U}^T(\operatorname{Diag} \tilde{v})\widetilde{U} = U^T(\operatorname{Diag} v)U \quad \text{holds.}$$

Consequently, by Corollary 3.8, there exists a permutation $\sigma \in \operatorname{Fix}(x)$ satisfying $\sigma\tilde{v} = v$. Since $\hat{\partial} f(x)$ is invariant under the action of $\operatorname{Fix}(x)$, it follows that $\tilde{v}$ lies in $\operatorname{rb} \hat{\partial} f(x)$, and consequently from Remark 3.3 we deduce $v \in \operatorname{rb} \hat{\partial} f(x)$. This establishes the inclusion $\subset$ of (3.2). To see the reverse inclusion, consider a sequence $v_i \in \operatorname{aff} \hat{\partial} f(x)$ converging to $v \in \hat{\partial} f(x)$ with $v_i \notin \hat{\partial} f(x)$ for each index i. Fix an arbitrary matrix $U \in \mathbf{O}^n_X$ and observe that the matrices $U^T(\operatorname{Diag} v_i)U$ lie in $\operatorname{aff} \hat{\partial}(f \circ \lambda)(X)$ and converge to $U^T(\operatorname{Diag} v)U$. We now claim that the matrices $U^T(\operatorname{Diag} v_i)U$ all lie outside of $\hat{\partial}(f \circ \lambda)(X)$. Indeed suppose this is not the case. Then there exist matrices $\widetilde{U}_i \in \mathbf{O}^n_X$ and subgradients $\tilde{v}_i \in \hat{\partial} f(x)$ satisfying

$$U^T(\operatorname{Diag} v_i)U = \widetilde{U}_i^T(\operatorname{Diag} \tilde{v}_i)\widetilde{U}_i.$$

An application of Corollary 3.8 and Remark 3.3 then yields a contradiction. Therefore, the inclusion $U^T(\operatorname{Diag} v)U \in \operatorname{rb} \hat{\partial}(f \circ \lambda)(X)$ holds, and the validity of (3.2) follows.

Finally, we aim to prove (3.3). Observe that the inclusion $\subset$ of (3.3) is immediate from (3.2). To see the reverse inclusion, consider a matrix $U^T(\operatorname{Diag} v)U$ for some $U \in \mathbf{O}^n_X$ and $v \in \operatorname{ri} \hat{\partial} f(x)$. Again, an easy application of Corollary 3.8 and Remark 3.3 yields the inclusion $U^T(\operatorname{Diag} v)U \in \operatorname{ri} \hat{\partial}(f \circ \lambda)(X)$. We conclude that (3.3) holds. ∎

LEMMA 3.18 (symmetry of partly smooth manifolds). *Consider an lsc function $f : \mathbf{R}^n \to \overline{\mathbf{R}}$ that is locally symmetric at $\bar{x}$. Suppose that f is C^2-partly smooth at $\bar{x}$ relative to M. Then M is locally symmetric around $\bar{x}$.*

Proof. Consider a permutation $\sigma \in \operatorname{Fix}(\bar{x})$. Then the function f is partly smooth at $\bar{x}$ relative to σM. On the other hand, partly smooth manifolds are locally unique by Theorem 3.11. Consequently, we deduce equality $M = \sigma M$ locally around $\bar{x}$. The claim follows. ∎

The main result of this section is now immediate.

THEOREM 3.19 (lifts of C^p-partly smooth functions). *Consider an lsc function $f : \mathbf{R}^n \to \overline{\mathbf{R}}$ and a matrix $\overline{X} \in \mathbf{S}^n$. Suppose that f is locally symmetric around $\bar{x} := \lambda(\overline{X})$. Then f is C^p-partly smooth ($p = 2, \ldots, \infty, \omega$) at $\bar{x}$ relative to M if and only if $f \circ \lambda$ is C^p-partly smooth at $\overline{X}$ relative to $\lambda^{-1}(M)$.*

Proof. Throughout the proof, we will be using Proposition 3.14, thereby going back and forth between identifiability and partial smoothness. Suppose that f is C^p-partly smooth at $\bar{x}$ relative to M. In light of Lemma 3.18, we deduce that M is locally symmetric at $\bar{x}$. Consequently, Theorem 2.7 implies that the set $\lambda^{-1}(M)$ is a C^p manifold, while Corollary 2.3 implies that $f \circ \lambda$ is C^p-smooth on $\lambda^{-1}(M)$ near $\overline{X}$. Applying Theorem 3.6, we conclude that $f \circ \lambda$ is prox-regular at $\overline{X}$. Consider now a subgradient $V \in \operatorname{ri} \partial(f \circ \lambda)(\overline{X})$. Then by Proposition 3.17, there exists a vector $v \in \operatorname{ri} \partial f(\bar{x})$ and a matrix $U \in \mathbf{O}^n_{\overline{X}}$ satisfying

$$V = U^T(\operatorname{Diag} v)U \quad \text{and} \quad \overline{X} = U^T(\operatorname{Diag} \bar{x})U.$$

Observe by Proposition 3.14, the set M is identifiable at $\bar{x}$ for v. Then applying Proposition 3.15, we deduce that $\lambda^{-1}(M)$ is identifiable (relative to $f \circ \lambda$) at $\overline{X}$ relative to V. Since V is an arbitrary element of $\operatorname{ri} \partial(f \circ \lambda)(\overline{X})$, applying Proposition 3.14, we deduce that $f \circ \lambda$ is C^p-partly smooth at $\overline{X}$ relative to $\lambda^{-1}(M)$, as claimed. The converse follows along the same lines. ∎

4. Partly smooth duality for polyhedrally generated spectral functions.

Consider an lsc, convex function $f : \mathbf{E} \to \overline{\mathbf{R}}$. Then the *Fenchel conjugate* $f^* : \mathbf{E} \to \overline{\mathbf{R}}$ is defined by setting

$$f^*(y) = \sup_{x \in \mathbf{R}^n} \{\langle x, y \rangle - f(x)\}.$$

Moreover, in terms of the powerset of $\mathbf{E}$, denoted $\mathbb{P}(\mathbf{E})$, we define a correspondence $\mathcal{J}_f : \mathbb{P}(\mathbf{E}) \to \mathbb{P}(\mathbf{E})$ by setting

$$\mathcal{J}_f(Q) := \bigcup_{x \in Q} \operatorname{ri} \partial f(x).$$

The significance of this map will become apparent shortly. Before proceeding, we recall some basic properties of the conjugation operation:

Biconjugation: $f^{**} = f$,

Subgradient inversion formula: $\partial f^* = (\partial f)^{-1}$,

Fenchel–Young inequality: $\langle x, y \rangle \leq f(x) + f^*(y)$ for every $x, y \in \mathbf{R}^n$.

Moreover, convexity and conjugation behave well under spectral lifts. See, for example, [3, section 5.2].

THEOREM 4.1 (lifts of convex sets and conjugation). *If $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is a symmetric function, then f^* is also symmetric and the formula*

$$(f \circ \lambda)^* = f^* \circ \lambda \quad \textit{holds.}$$

Furthermore, f is convex if and only if the spectral function $f \circ \lambda$ is convex.

The following definition is standard.

DEFINITION 4.2 (stratification). *A finite partition $\mathcal{A}$ of a set $Q \subset \mathbf{E}$ is a* stratification *provided that for any partitioning sets (called* strata*) M_1 and M_2 in $\mathcal{A}$, the implication*

$$M_1 \cap \operatorname{cl} M_2 \neq \emptyset \quad \Longrightarrow \quad M_1 \subset \operatorname{cl} M_2 \quad \textit{holds.}$$

If the strata are open polyhedra, then $\mathcal{A}$ is a polyhedral stratification. *If the strata are C^p manifolds, then $\mathcal{A}$ is a C^p*-stratification.

Stratification duality for convex polyhedral functions. We now establish the setting and notation for the rest of the section. Suppose that $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is a convex *polyhedral* function (epigraph of f is a closed convex polyhedron). Then f induces a finite polyhedral stratification $\mathcal{A}_f$ of $\operatorname{dom} f$ in a natural way. Namely, consider the partition of $\operatorname{epi} f$ into open faces $\{F_i\}$. Projecting all faces F_i, with $\dim F_i \leq n$, onto the first n-coordinates we obtain a stratification of the domain $\operatorname{dom} f$ of f that we denote by $\mathcal{A}_f$. In fact, one can easily see that f is C^ω-partly smooth relative to each polyhedron $M \in \mathcal{A}_f$.

A key observation for us will be that the correspondence $f \overset{*}{\longleftrightarrow} f^*$ is not only a pairing of functions, but it also induces a duality pairing between $\mathcal{A}_f$ and $\mathcal{A}_{f^*}$. Namely, one can easily check that the mapping $\mathcal{J}_f$ restricts to an invertible mapping $\mathcal{J}_f\colon \mathcal{A}_f \to \mathcal{A}_{f^*}$ with inverse given by $\mathcal{J}_{f^*}$.

Limitations of stratification duality. It is natural to ask whether for general (nonpolyhedral) lsc, convex functions $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, the correspondence $f \overset{*}{\longleftrightarrow} f^*$, along with the mapping $\mathcal{J}$, induces a pairing between *partly smooth manifolds* of f and f^*. A little thought, however, shows an immediate obstruction: images of C^ω-smooth manifolds under the map $\mathcal{J}_f$ may fail to be even C^2-smooth.

Example 4.3 (failure of smoothness). Consider the conjugate pair

$$f(x,y) = \frac{1}{4}(x^4 + y^4) \qquad \text{and} \qquad f^*(x,y) = \frac{3}{4}(|x|^{\frac{4}{3}} + |y|^{\frac{4}{3}}).$$

Clearly, f is partly smooth relative to $\mathbf{R}^2$, whereas any possible partition of $\mathbf{R}^2$ into partly smooth manifolds relative to f^* must consist of at least three manifolds (one manifold in each dimension: one, two, and three). Hence no duality pairing between partly smooth manifolds is possible. See Figures 2 and 3 for an illustration.

Indeed, this is not very surprising, since the convex duality is really a duality between smoothness and *strict* convexity. See, for example, [28, section 4] or [31, Theorem 11.13]. Hence, in general, one needs to impose tough strict convexity conditions in order to hope for this type of duality to hold. Rather than doing so, and more in line with the current work, we consider the spectral setting. Namely, we will show that in the case of spectral functions $F := f \circ \lambda$, with f symmetric and polyhedral—functions of utmost importance in eigenvalue optimization—the mapping $\mathcal{J}$ does induce a duality correspondence between partly smooth manifolds of F and F^*.

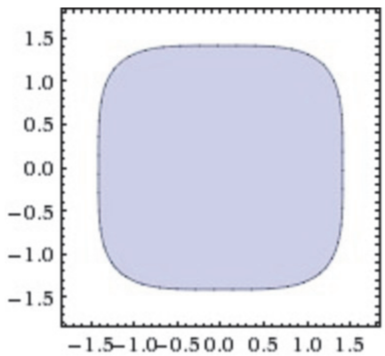

FIG. 2. $\{(x,y) : x^4 + y^4 \le 4\}$.

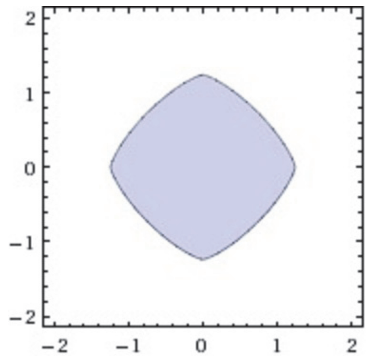

FIG. 3. $\{(x,y) : |x|^{\frac{4}{3}} + |y|^{\frac{4}{3}} \le \frac{4}{3}\}$.

In what follows, let us denote by

$$M^{\mathrm{sym}} := \bigcup_{\sigma \in \Sigma} \sigma M,$$

the symmetrization of any subset $M \subset \mathbf{R}^n$. Before we proceed, we will need the following result.

LEMMA 4.4 (path-connected lifts). *Let $M \subseteq \mathbf{R}^n$ be a path-connected set and assume that for any permutation $\sigma \in \Sigma$, we either have $\sigma M = M$ or $\sigma M \cap M = \emptyset$. Then $\lambda^{-1}(M^{\mathrm{sym}})$ is a path-connected subset of $\mathbf{S}^n$.*

Proof. Let X_1, X_2 be in $\lambda^{-1}(M^{\mathrm{sym}})$, and set $x_i = \lambda(X_i) \in M^{\mathrm{sym}} \cap \mathbf{R}^n_{\geq}$, for $i \in \{1,2\}$. It is standard to check that the sets $\lambda^{-1}(x_i)$ are path-connected manifolds for $i = 1, 2$. Consequently, the matrices X_i and $\mathrm{Diag}(x_i)$ can be joined via a path lying in $\lambda^{-1}(x_i)$. Thus, in order to construct a path joining X_1 to X_2 and lying in $\lambda^{-1}(M^{\mathrm{sym}})$, it would be sufficient to join x_1 to x_2 inside M^{sym}. This in turn will follow immediately if both $\sigma x_1, \sigma x_2$ belong to M for some $\sigma \in \Sigma$. To establish this, we will assume without loss of generality that x_1 lies in M. In particular, we have $M \cap \mathbf{R}^n_{\geq} \neq \emptyset$ and we will establish the inclusion $x_2 \in M$.

To this end, consider a permutation $\sigma \in \Sigma$ satisfying $x_2 \in \sigma M \cap \mathbf{R}^n_{\geq}$. Our immediate goal is to establish $\sigma M \cap M \neq \emptyset$, and thus $\sigma M = M$ thanks to our assumption. To this end, consider the point $y \in M$ satisfying $x_2 = \sigma y$. If y lies in $\mathbf{R}^n_{\geq}$, then we deduce $y = x_2$ and we are done. Therefore, we can assume $y \notin \mathbf{R}^n_{\geq}$. We can then consider the decomposition $\sigma = \sigma_k \cdots \sigma_1$ of the permutation σ into 2-cycles σ_i, each of which permutes exactly two coordinates of y that are not in the right (decreasing) order. For the sake of brevity, we omit details of the construction of such a decomposition; besides, it is rather standard. We claim now $\sigma_1 M = M$. To see this, suppose that σ_1 permutes the i and j coordinates of y where $y_i < y_j$ and $i > j$. Since x_1 lies in $\mathbf{R}^n_{\geq}$ and M is path-connected, there exists a point $z \in M$ satisfying $z_i = z_j$. Then $\sigma_1 z = z$, whence by assumption $\sigma_1 M = M$ and $\sigma_1 y \in M$. Applying the same argument to $\sigma_1 y$ and $\sigma_1 M$ with the 2-cycle σ_2 we obtain $\sigma_2 \sigma_1 M = M$ and $\sigma_2 \sigma_1 y \in M$. By induction, $\sigma M = M$. Thus $x_2 \in M$ and the assertion follows. ∎

Stratification duality for spectral lifts. Consider a symmetric, convex polyhedral function $f \colon \mathbf{R}^n \to \overline{\mathbf{R}}$ together with its induced stratification $\mathcal{A}_f$ of $\mathrm{dom}\, f$. Then with each polyhedron $M \in \mathcal{A}_f$, we may associate the symmetric set M^{sym}. We record some properties of such sets in the following lemma.

LEMMA 4.5 (properties of $\mathcal{A}_f$). *Consider a symmetric, convex polyhedral function*

$f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and the induced stratification $\mathcal{A}_f$ of $\operatorname{dom} f$*. Then the following are true:*

(i) *For any sets $M_1, M_2 \in \mathcal{A}_f$ and any permutation $\sigma \in \Sigma$, the sets σM_1 and M_2 either coincide or are disjoint.*

(ii) *The action of Σ on $\mathbf{R}^n$ induces an action of Σ on*

$$\mathcal{A}_f^k := \{M \in \mathcal{A}_f : \dim M = k\}$$

for each $k = 0, \ldots, n$. In particular, the set M^{sym} is simply the union of all polyhedra belonging to the orbit of M under this action.

(iii) *For any polyhedron $M \in \mathcal{A}_f$, and every point $x \in M$, there exists a neighborhood U of x satisfying $U \cap M^{\mathrm{sym}} = U \cap M$. Consequently, M^{sym} is a C^ω manifold of the same dimension as M.*

Moreover, $\lambda^{-1}(M^{\mathrm{sym}})$ is connected, whenever M is.

The last assertion follows from Lemma 4.4. The remaining assertions are straightforward and hence we omit their proof.

Notice that the strata of the stratification $\mathcal{A}_f$ are connected C^ω manifolds, which fail to be symmetric in general. In light of Lemma 4.5, the set M^{sym} is a C^ω manifold and a disjoint union of open polyhedra. Thus the collection

$$\mathcal{A}_f^{\mathrm{sym}} := \{M^{\mathrm{sym}} : M \in \mathcal{A}_f\}$$

is a stratification of $\operatorname{dom} f$, whose strata are now symmetric manifolds. Even though the new strata are disconnected, they give rise to connected lifts $\lambda^{-1}(M^{\mathrm{sym}})$. One can easily verify that, as before, $\mathcal{J}_f$ restricts to an invertible mapping $\mathcal{J}_f\colon \mathcal{A}_f^{\mathrm{sym}} \to \mathcal{A}_{f^*}^{\mathrm{sym}}$ with inverse given by the restriction of $\mathcal{J}_{f^*}$.

We now arrive at the main result of the section.

Theorem 4.6 (lift of the duality map). *Consider a symmetric, convex polyhedral function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and define the spectral function $F := f \circ \lambda$. Let $\mathcal{A}_f$ be the finite polyhedral partition of* $\operatorname{dom} f$ *induced by f, and define the collection*

$$\mathcal{A}_F := \Big\{\lambda^{-1}(M^{\mathrm{sym}}) : M \in \mathcal{A}_f\Big\}.$$

Then the following properties hold:

(i) *$\mathcal{A}_F$ is a C^ω-stratification of* $\operatorname{dom} F$ *comprised of connected manifolds,*

(ii) *F is C^ω-partly smooth relative to each set $\lambda^{-1}(M^{\mathrm{sym}}) \in \mathcal{A}_F$.*

(iii) *The map $\mathcal{J}_F\colon \mathbb{P}(\mathbf{S}^n) \to \mathbb{P}(\mathbf{S}^n)$ restricts to an invertible mapping $\mathcal{J}_F\colon \mathcal{A}_F \to \mathcal{A}_{F^*}$ with inverse given by the restriction of $\mathcal{J}_{F^*}$.*

(iv) *The following diagram commutes:*

$$\begin{tikzcd}
\mathcal{A}_F \arrow[r, "\mathcal{J}_F"] & \mathcal{A}_{F^*} \\
\mathcal{A}_f^{\mathrm{sym}} \arrow[u, "\lambda^{-1}"] \arrow[r, "\mathcal{J}_f"'] & \mathcal{A}_{f^*}^{\mathrm{sym}} \arrow[u, "\lambda^{-1}"']
\end{tikzcd}$$

That is, the equation $(\lambda^{-1} \circ \mathcal{J}_f)(M^{\mathrm{sym}}) = (\mathcal{J}_F \circ \lambda^{-1})(M^{\mathrm{sym}})$ holds for every set $M^{\mathrm{sym}} \in \mathcal{A}_f^{\mathrm{sym}}$.

Proof. In light of Lemma 4.5, each set $M^{\mathrm{sym}} \in \mathcal{A}_f^{\mathrm{sym}}$ is a symmetric C^ω manifold. The fact that $\mathcal{A}_F$ is a C^ω-stratification of $\operatorname{dom} F$ now follows from the transfer principle for stratifications [15, Theorem 4.8], while the fact that each manifold $\lambda^{-1}(M^{\mathrm{sym}})$

is connected follows immediately from Lemma 4.5. Moreover, from Theorem 3.19, we deduce that F is C^ω-partly smooth relative to each set in $\mathcal{A}_F$.

Consider now a set $M^{\mathrm{sym}} \in \mathcal{A}_f^{\mathrm{sym}}$ for some $M \in \mathcal{A}_f$. Then we have

$$\begin{aligned}\mathcal{J}_F(\lambda^{-1}(M^{\mathrm{sym}})) &= \bigcup_{X \in \lambda^{-1}(M^{\mathrm{sym}})} \operatorname{ri} \partial F(X) \\ &= \bigcup_{X \in \lambda^{-1}(M^{\mathrm{sym}})} \{U^T (\operatorname{Diag} v) U : v \in \operatorname{ri} \partial f(\lambda(X)) \text{ and } U \in \mathbf{O}^n_X\},\end{aligned}$$

and concurrently,

$$\lambda^{-1}(\mathcal{J}_f(M^{\mathrm{sym}})) = \lambda^{-1}\left(\bigcup_{x \in M^{\mathrm{sym}}} \operatorname{ri} \partial f(x)\right) = \bigcup_{x \in M^{\mathrm{sym}},\ v \in \operatorname{ri} \partial f(x)} \mathbf{O}^n.(\operatorname{Diag} v).$$

We claim that the equality $\lambda^{-1}(\mathcal{J}_f(M^{\mathrm{sym}})) = \mathcal{J}_F(\lambda^{-1}(M^{\mathrm{sym}}))$ holds. The inclusion "$\supset$" is immediate. To see the converse, fix a point $x \in M^{\mathrm{sym}}$, a vector $v \in \operatorname{ri} \partial f(x)$, and a matrix $U \in \mathbf{O}^n$. We must show $V := U^T (\operatorname{Diag} v) U \in \mathcal{J}_F(\lambda^{-1}(M^{\mathrm{sym}}))$. To see this, fix a permutation $\sigma \in \Sigma$ with $\sigma x \in \mathbf{R}^n_{\geq}$, and observe

$$U^T (\operatorname{Diag} v) U = (A_\sigma U)^T (\operatorname{Diag} \sigma v) A_\sigma U,$$

where A_σ denotes the matrix representing the permutation σ. Define a matrix $X := (A_\sigma U)^T (\operatorname{Diag} \sigma x) A_\sigma U$. Clearly, we have $V \in \operatorname{ri} \partial F(X)$ and $X \in \lambda^{-1}(M^{\mathrm{sym}})$. This proves the claimed equality. Consequently, we deduce that the assignment $\mathcal{J}_F \colon \mathbb{P}(\mathbf{S}^n) \to \mathbb{P}(\mathbf{S}^n)$ restricts to a mapping $\mathcal{J}_F \colon \mathcal{A}_F \to \mathcal{A}_{F^*}$, and that the diagram commutes. Commutativity of the diagram along with the fact that $\mathcal{J}_{f^*}$ restricts to be the inverse of $\mathcal{J}_f \colon \mathcal{A}_f^{\mathrm{sym}} \to \mathcal{A}_{f^*}^{\mathrm{sym}}$ implies that $\mathcal{J}_{F^*}$ restricts to be the inverse of $\mathcal{J}_F \colon \mathcal{A}_F \to \mathcal{A}_{F^*}$. ∎

Example 4.7 (constant rank manifolds). Consider the closed convex cones of positive (respectively, negative) semidefinite matrices $\mathbf{S}^n_+$ (respectively, $\mathbf{S}^n_-$). Clearly, we have equality $\mathbf{S}^n_\pm = \lambda^{-1}(\mathbf{R}^n_\pm)$. Define the constant rank manifolds

$$M_k^\pm := \{X \in \mathbf{S}^n_\pm : \operatorname{rank} X = k\} \quad \text{ for } k = 0, \ldots, n.$$

Then using Theorem 4.6 one can easily check that the manifolds $M_k^\pm$ and $M_{n-k}^\mp$ are dual to each other under the conjugacy correspondence $\delta_{\mathbf{S}^n_+} \overset{*}{\longleftrightarrow} \delta_{\mathbf{S}^n_-}$.

5. Extensions to nonsymmetric matrices. Consider the space of $n \times m$ real matrices $\mathbf{M}^{n \times m}$, endowed with the trace inner-product $\langle X, Y \rangle = \operatorname{tr}(X^T Y)$, and the corresponding Frobenius norm. We will let the group $\mathbf{O}^{n,m} := \mathbf{O}^n \times \mathbf{O}^m$ act on $\mathbf{M}^{n \times m}$ simply by defining

$$(U, V).X = U^T X V \text{ for all } (U, V) \in \mathbf{O}^{n,m} \quad \text{and} \quad X \in \mathbf{M}^{n \times m}.$$

Recall that *singular values* of a matrix $A \in \mathbf{M}^{n \times m}$ are the square roots of the eigenvalues of the matrix $A^T A$. The *singular value mapping* $\sigma : \mathbf{M}^{n \times m} \to \mathbf{R}^m$ is simply the mapping taking each matrix X to its vector $(\sigma_1(X), \ldots, \sigma_m(X))$ of singular values in nonincreasing order. We will be interested in functions $F \colon \mathbf{M}^{n \times m} \to \overline{\mathbf{R}}$ that are invariant under the action of $\mathbf{O}^{n,m}$. Such functions F are representable as a composition $F = f \circ \sigma$, where the outer-function $f \colon \mathbf{R}^m \to \overline{\mathbf{R}}$ is *absolutely permutation-invariant*, meaning invariant under all signed permutations of coordinates. As in the symmetric case, it is useful to localize this notion. Namely, we will say that a function f is *locally absolutely permutation-invariant* around a point $\bar{x}$ provided that for each signed permutation σ fixing $\bar{x}$, we have $f(\sigma x) = f(x)$ for all x near $\bar{x}$. Then all of the results presented in the symmetric case have natural analogues in this setting (with nearly identical proofs).

THEOREM 5.1 (the nonsymmetric case: lifts of manifolds). *Consider a matrix $\overline{X} \in \mathbf{M}^{n\times m}$ and a set $M \subset \mathbf{R}^m$ that is locally absolutely permutation-invariant around $\bar{x} := \sigma(\overline{X})$. Then M is a C^p manifold $(p = 2, \ldots, \infty, \omega)$ around $\bar{x}$ if and only if the set $\sigma^{-1}(M)$ is a C^p manifold around $\overline{X}$.*

PROPOSITION 5.2 (the nonsymmetric case: lifts of identifiable sets). *Consider an lsc $f \colon \mathbf{R}^m \to \overline{\mathbf{R}}$ and a matrix $\overline{X} \in \mathbf{M}^{n\times m}$. Suppose that f is locally absolutely permutation-invariant around $\bar{x} := \sigma(\overline{X})$ and consider a subset $M \subset \mathbf{R}^m$ that is locally absolutely permutation-invariant around $\bar{x}$. Then M is identifiable (relative to f) at $\bar{x}$ for $\bar{v} \in \partial f(\bar{x})$, if and only if $\sigma^{-1}(M)$ is identifiable (relative to $f \circ \sigma$) at $\overline{X}$ for $U^T(\mathrm{Diag}\,\bar{v})V \in \partial(f \circ \sigma)(\overline{X})$, where $(U, V) \in \mathbf{O}^{n,m}$ is any pair satisfying $\overline{X} = U^T(\mathrm{Diag}\,\sigma(\overline{X}))V$.*

THEOREM 5.3 (the nonsymmetric case: lifts of partly smooth manifolds). *Consider an lsc function $f \colon \mathbf{R}^m \to \overline{\mathbf{R}}$ and a matrix $\overline{X} \in \mathbf{M}^{n\times m}$. Suppose that f is locally absolutely permutation-invariant around $\bar{x} := \sigma(\overline{X})$. Then f is C^p-partly smooth $(p = 2, \ldots, \infty, \omega)$ at $\bar{x}$ relative to M if and only if $f \circ \sigma$ is C^p-partly smooth at $\overline{X}$ relative to $\sigma^{-1}(M)$.*

Finally, we should note that section 4 also has a natural analogue in the nonsymmetric setting. For the sake of brevity, we do not record it here.

Acknowledgment. The first author thanks Nicolas Hadjisavvas for useful discussions leading to a simplification of the proof of Lemma 4.4.

REFERENCES

[1] F. Al-Khayyal and J. Kyparisis, *Finite convergence of algorithms for nonlinear programs and variational inequalities*, J. Optim. Theory Appl., 70 (1991), pp. 319–332.
[2] J. Bolte, A. Daniilidis, and A.S. Lewis, *Generic optimality conditions for semialgebraic convex programs*, Math. Oper. Res., 36 (2011), pp. 55–70.
[3] J.M. Borwein and A.S. Lewis, *Convex Analysis and Nonlinear Optimization, Theory, and Examples*, CMS Books Math./Ouvrages Math. SMC 3, Springer, New York, 2000.
[4] J.M. Borwein and Q.J. Zhu, *Techniques of Variational Analysis*, Springer-Verlag, New York, 2005.
[5] J.V. Burke, *On the identification of active constraints.* II. *The nonconvex case*, SIAM J. Numer. Anal., 27 (1990), pp. 1081–1103.
[6] J.V. Burke and J.J. Moré, *On the identification of active constraints*, SIAM J. Numer. Anal., 25 (1988), pp. 1197–1211.
[7] P.H. Calamai and J.J. Moré, *Projected gradient methods for linearly constrained problems*, Math. Programming, 39 (1987), pp. 93–116.
[8] F.H. Clarke, S. Yu. Ledyaev, R.J. Stern, and P.R. Wolenski, *Nonsmooth Analysis and Control Theory*, Grad. Texts in Math. 178, Springer-Verlag, New York, 1998.
[9] A. Daniilidis, A.S. Lewis, J. Malick, and H. Sendov, *Prox-regularity of spectral functions and spectral sets*, J. Convex Anal., 15 (2008), pp. 547–560.
[10] A. Daniilidis, J. Malick, and H.S. Sendov, *Locally symmetric submanifolds lift to spectral manifolds*, preprint U.A.B. 23/2009, arXiv:1212.3936 [math.OC], 2012.
[11] A. Daniilidis, J. Malick, and H.S. Sendov, *On the structure of locally symmetric manifolds*, submitted.
[12] A. Daniilidis, J. Malick, and H.S. Sendov, *Spectral (isotropic) manifolds and their dimension*, Journal d'Analyse Mathematique, to appear.

[13] A. Daniilidis, C. Sagastizábal, and M. Solodov, *Identifying structure of nonsmooth convex functions by the bundle technique*, SIAM J. Optim., 20 (2009), pp. 820–840.
[14] C. Davis, *All convex invariant functions of Hermitian matrices*, Arch. Math., 8 (1957), pp. 276–278.
[15] D. Drusvyatskiy and M. Larsson, *Approximating functions on stratifiable sets*, Trans. Amer. Math. Soc., to appear.
[16] D. Drusvyatskiy and A.S. Lewis, *Optimality, identifiability, and sensitivity*, Math. Program., (2013), pp. 1–32.
[17] J.C. Dunn, *On the convergence of projected gradient processes to singular critical points*, J. Optim. Theory Appl., 55 (1987), pp. 203–216.
[18] M.C. Ferris, *Finite termination of the proximal point algorithm*, Math. Programming, 50 (1991), pp. 359–366.
[19] S.D. Flåm, *On finite convergence and constraint identification of subgradient projection methods*, Math. Program., 57 (1992), pp. 427–437.
[20] W.L. Hare and A.S. Lewis, *Identifying active constraints via partial smoothness and prox-regularity*, J. Convex Anal., 11 (2004), pp. 251–266.
[21] A.S. Lewis, *Derivatives of spectral functions*, Math. Oper. Res., 21 (1996), pp. 576–588.
[22] A.S. Lewis, *Nonsmooth analysis of eigenvalues*, Math. Program., 84 (1999), pp. 1–24.
[23] A.S. Lewis, *Active sets, nonsmoothness, and sensitivity*, SIAM J. Optim., 13 (2002), pp. 702–725.
[24] A.S. Lewis and H.S. Sendov, *Nonsmooth analysis of singular values.* I. *Theory*, Set-Valued Anal., 13 (2005), pp. 213–241.
[25] A.S. Lewis and S.J. Wright, *A proximal method for composite minimization*, arXiv:0812.0423, 2010.
[26] A.S. Lewis and S.J. Wright, *Identifying activity*, SIAM J. Optim., 21 (2011), pp. 597–614.
[27] B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation* I: *Basic Theory*, Grundlehren Math. Wiss. 330, Springer-Verlag, Berlin, 2006.
[28] R.R. Phelps, *Convex Functions, Monotone Operators and Differentiability*, 2nd ed., Lecture Notes in Math. 1364, Springer-Verlag, Berlin, 1993.
[29] R.A. Poliquin and R.T. Rockafellar, *Prox-regular functions in variational analysis*, Trans. Amer. Math. Soc., 348 (1996), pp. 1805–1838.
[30] J.-B. Poly and G. Raby, *Fonction distance et singularités*, Bull. Sci. Math. (2), 108 (1984), pp. 187–195.
[31] R.T. Rockafellar and R.J-B. Wets, *Variational Analysis*, Grundlehren Math. Wiss. 317, Springer-Verlag, Berlin, 1998.
[32] H.S. Sendov, *The higher-order derivatives of spectral functions*, Linear Algebra Appl., 424 (2007), pp. 240–281.
[33] M. Šilhavý, *Differentiability properties of isotropic functions*, Duke Math. J., 104 (2000), pp. 367–373.
[34] J. Sylvester, *On the differentiability of* $\mathrm{O}(n)$ *invariant functions of symmetric matrices*, Duke Math. J., 52 (1985), pp. 475–483.
[35] J. von Neumann, *Some matrix inequalities and metrization of matrix-space*, Tomsk. Univ. Rev., 1 (1937), pp. 286–300.
[36] S.J. Wright, *Identifiable surfaces in constrained optimization*, SIAM J. Control Optim., 31 (1993), pp. 1063–1079.