ORTHOGONAL INVARIANCE AND IDENTIFIABILITY

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

Abstract. Orthogonally invariant functions of symmetric matrices often inherit properties from their diagonal restrictions: von Neumann's theorem on matrix norms is an early example. We discuss the example of "identifiability", a common property of nonsmooth functions associated with the existence of a smooth manifold of approximate critical points. Identifiability (or its synonym, "partial smoothness") is the key idea underlying active set methods in optimization. Polyhedral functions, in particular, are always partly smooth, and hence so are many standard examples from eigenvalue optimization.

Key words. Eigenvalues, symmetric matrix, partial smoothness, identifiable set, polyhedra

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

1. Introduction. Nonsmoothness is inherently present throughout even classical mathematics and engineering - the spectrum of a symmetric matrix variable is a good example. The nonsmooth behavior is not, however, typically pathological, but on the contrary is highly structured. The theory of *identifiability* (or its synonym, *partial smoothness*) [25, 21, 36, 16] models this idea by positing existence of smooth manifolds capturing the full "activity" of the problem. Such manifolds, when they exist, are simply composed of approximate critical points of the minimized function. In the classical case of nonlinear programming, this theory reduces to the active set philosophy. Illustrating the ubiquity of the notion, the authors of [3] prove that identifiable manifolds exist generically for convex semi-algebraic optimization problems.

Identifiable manifolds are particularly prevalent in the context of eigenvalue optimization. One of our goals is to shed new light on this phenomenon. To this end, we will consider so-called *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 $X \in \mathbf{S}$ the vector of its eigenvalues $(\lambda_1(X), \ldots, \lambda_n(X))$ in non-increasing order, see [4, 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 very badly behaved, as far as say differentiability is concerned, the symmetry of f makes up for the fact, allowing powerful analytic results to become available.

In particular, the *Transfer Principle* asserts that F 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 impressive: convexity [24, 14], prox-regularity [10], Clarke-regularity [26, 24], smoothness [24, 23, 32, 11, 34, 33], algebraicity [11], and stratifiability [15, Theorem 4.8]. In this work, we add identifiability (and partial smoothness) to the list (Theorems 3.17 and 3.21). In particular, 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 convex *polyhedral* function. As a direct corollary of our results, we conclude that such functions always admit partly smooth structure! Furthermore, a "polyhedral-like" duality theory of partly smooth manifolds becomes available.

One of our intermediary theorems is of particular interest. We will give an elementary argument showing that a permutation-invariant set M is a $\mathbf{C}^\infty$ manifold if and only if the spectral set $\lambda^{-1}(M)$ is a $\mathbf{C}^\infty$ manifold (Theorem 2.7). The converse implication of our result is apparently new. On the other hand, the authors of [11] proved the forward implication even for $\mathbf{C}^k$ manifolds (for $k = 2, \dots, \infty$). This being said, their proof is rather long and dense, whereas the proof of our result is very accessible. The key idea of our approach is to consider the metric projection onto M.

The outline of the manuscript is as follows. In Section 2 we establish some basic notation and give an elementary proof of the spectral lifting property for $\mathbf{C}^\infty$ 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 in the world of 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 say that an extended-real-valued function is *proper* if it is never $\{-\infty\}$ and is 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 $\mathbf{B}$. 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 $x_1 \geq x_2 \geq \dots \geq x_n$. 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 $\langle X, Y \rangle = \operatorname{tr}(XY)$, 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 $h \colon \mathbf{S}^n \to \overline{\mathbf{R}}$ is said to be *spectral* if we have $h(X) = h(U.X)$ for every $X \in \mathbf{S}^n$ and every $U \in \mathbf{O}^n$.

2.2. Spectral functions and the 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 its 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 [24, 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. We will encounter a number of such properties in the current work. Evidently, analogous results hold 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.

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

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

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 smoothness satisfies the Transfer Principle [33, 34].

THEOREM 2.2 (Lifts of smoothness). *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 $\mathbf{C}^p$-smooth ($p = 1, \ldots, \infty$) around $\bar{x}$ if and only if the spectral function $f \circ \lambda$ is $\mathbf{C}^p$-smooth around $\overline{X}$.*

The *distance* of a point $\bar{x}$ 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

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

It will be important for us to relate properties of a set Q with those of the metric projection P_Q. To this end, the following notion arises naturally [29, 30].

DEFINITION 2.3 (Prox-regularity). *A set $Q \subset \mathbf{E}$ is prox-regular at a point $\bar{x} \in Q$ if Q is locally closed around $\bar{x}$ and the projection mapping P_Q is single-valued around $\bar{x}$.*

In particular, all $\mathbf{C}^2$-manifolds and all closed convex sets are prox-regular around any of their points. See for example [31, Example 13.30, Proposition 13.32]. Additionally, it is well-known that if $M \subset \mathbf{E}$ is a $\mathbf{C}^p$ smooth manifold (for $p \geq 2$) around a point $\bar{x} \in M$, then there exists a neighborhood U of $\bar{x}$ on which the projection P_M is single-valued and $\mathbf{C}^{p-1}$-smooth. Prox-regularity also satisfies the transfer principle [10, Proposition 2.3, Theorem 2.4].

THEOREM 2.4 (Lifts of prox-regularity). *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*

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

Furthermore, Q is prox-regular at $\bar{x}$ if and only if $\lambda^{-1}(Q)$ is prox-regular at $\overline{X}$.

If a set $Q \subset \mathbf{E}$ is prox-regular at $\bar{x}$, then the *proximal normal cone*

$$N_Q(\bar{x}) := \mathbf{R}_+\{v \in \mathbf{E} : \bar{x} \in P_Q(\bar{x} + v)\},$$

and the *tangent cone*

$$T_Q(\bar{x}) := \Big\{ \lim_{i \to \infty} \lambda_i(x_i - \bar{x}) : \lambda_i \uparrow \infty \text{ and } x_i \in Q \Big\}.$$

are closed convex cones and are polar to each other [31, Corollary 6.29]. Here, we mean polarity in the standard sense of convex analysis, namely for any closed convex cone $K \subset \mathbf{E}$, the polar of K is another closed convex cone defined by

$$K^o := \{v \in \mathbf{E} : \langle v, w \rangle \le 0 \text{ for all } w \in K\}.$$

2.3. Lifts of symmetric manifolds. It turns out (not surprisingly) that smoothness of the projection P_Q is inherently tied to smoothness of Q itself, which is the content of the following lemma.

For any mapping $F \colon \mathbf{E} \to \mathbf{E}$, the directional derivative of F at $\bar{x}$ in direction w (if it exists) will be denoted by

$$DF(\bar{x})(w) := \lim_{t \downarrow 0} \frac{F(\bar{x} + tw) - F(\bar{x})}{t},$$

while the Gâteaux derivative of F at $\bar{x}$ (if it exists) will be denoted by $DF(\bar{x})$.

LEMMA 2.5 (Smoothness of the metric projection). *Consider a set $Q \subset \mathbf{E}$ that is prox-regular at a point $\bar{x} \in Q$. Then*

$$DP_Q(\bar{x})(v) = 0, \quad \text{for any } v \in N_Q(\bar{x}). \tag{2.1}$$

If P_Q is directionally differentiable at $\bar{x}$, then we also have

$$DP_Q(\bar{x})(w) = w, \quad \text{for any } w \in T_Q(\bar{x}). \tag{2.2}$$

In particular, if P_Q is Gâteaux differentiable at $\bar{x}$, then $N_Q(\bar{x})$ and $T_Q(\bar{x})$ are orthogonal subspaces and $DP_Q(\bar{x}) = P_{T_Q(\bar{x})}$. If P_Q is $\mathbf{C}^k$ $(k = 1, \dots, \infty)$ smooth near $\bar{x}$, then P_Q automatically has constant rank near $\bar{x}$ and consequently Q is a $\mathbf{C}^k$ manifold around $\bar{x}$.

Proof. Observe that for any normal vector $\bar{v} \in N_Q(\bar{x})$ there exists $\epsilon > 0$ so that $P_Q(\bar{x} + \epsilon'\bar{v}) = \bar{x}$ for all nonnegative $\epsilon' < \epsilon$. Equation (2.1) is now immediate.

Suppose now that P_Q is directionally differentiable at $\bar{x}$ and consider a vector $w \in T_Q(\bar{x})$ with $\|w\| = 1$. Then there exists a sequence $x_i \in Q$ converging to $\bar{x}$ and satisfying $w = \lim_{i \to \infty} \frac{x_i - \bar{x}}{\|x_i - \bar{x}\|}$. Define $t_i := \|x_i - \bar{x}\|$ and observe that since P_Q is Lipschitz continuous, for some constant L we have

$$\frac{\|P_Q(\bar{x} + t_i w) - P_Q(x_i)\|}{t_i} \le L \Big\| w - \frac{x_i - \bar{x}}{t_i} \Big\|,$$

and consequently this quantity converges to zero. We obtain

$$DP_Q(\bar{x})(w) = \lim_{i \to \infty} \frac{P_Q(\bar{x} + t_i w) - \bar{x}}{t_i} = \lim_{i \to \infty} \frac{P_Q(x_i) - \bar{x}}{t_i} = w,$$

as claimed.

Suppose now that P_Q is Gâteaux differentiable at $\bar{x}$. Then clearly from (2.1) we have $N_Q(\bar{x}) \subset \ker DP_Q(\bar{x})$. If $N_Q(\bar{x})$ were a proper convex subset of $\ker DP_Q(\bar{x})$, then we would deduce

$$T_Q(\bar{x}) \cap \ker DP_Q(\bar{x}) = [N_Q(\bar{x})]^\circ \cap \ker DP_Q(\bar{x}) \neq \{0\},$$

thereby contradicting equation (2.2). Hence $N_Q(\bar{x})$ and $T_Q(\bar{x})$ are orthogonal subspaces and the equation $DP_Q(\bar{x}) = P_{T_Q(\bar{x})}$ readily follows from (2.1) and (2.2).

Suppose now that P_Q is $\mathbf{C}^k$-smooth (for $k = 1, \ldots, \infty$) around $\bar{x}$. Then clearly we have

$$\operatorname{rank} DP_Q(x) \geq \operatorname{rank} DP_Q(\bar{x}), \text{ for all } x \text{ near } \bar{x}.$$

Towards establishing equality above, we now claim that the set-valued mapping T_Q is outer-semicontinuous at $\bar{x}$. To see this, consider sequences $x_i \to \bar{x}$ and $w_i \in T_Q(x_i)$, with w_i converging to some vector $\bar{w} \in \mathbf{E}$. From equation (2.2), we deduce $w_i = DP_Q(x_i)(w_i)$. Passing to the limit, while taking into account the continuity of DP_Q, we obtain $\bar{w} = DP_Q(\bar{x})(\bar{w})$. On the other hand, since $DP_Q(\bar{x})$ is simply the linear projection onto $T_Q(\bar{x})$, we deduce the inclusion $\bar{w} \in T_Q(\bar{x})$, and thereby establishing outer-semicontinuity of T_Q at $\bar{x}$. It immediately follows that the inequality, $\dim T_Q(x) \leq \dim T_Q(\bar{x})$, holds for all $x \in Q$ near $\bar{x}$.

One can easily verify that for any point x near $\bar{x}$, the inclusion $N_Q(P_Q(x)) \subset \ker DP_Q(x)$ holds. Consequently we deduce

$$\operatorname{rank} DP_Q(x) \leq \dim T_Q(P_Q(x)) \leq \dim T_Q(\bar{x}) = \operatorname{rank} DP_Q(\bar{x}),$$

for all $x \in \mathbf{E}$ sufficiently close to $\bar{x}$. as claimed. Hence P_Q has constant rank near $\bar{x}$. By the constant rank theorem, for all sufficiently small $\epsilon > 0$, the set $P_Q(B_\epsilon(\bar{x}))$ is a $\mathbf{C}^k$ manifold. Observing that the set $P_Q(B_\epsilon(\bar{x}))$ coincides with Q near $\bar{x}$ completes the proof. $\square$

The following observation will be key. It shows that the metric projection map onto a prox-regular set is itself a gradient of a $\mathbf{C}^1$-smooth function. This easily follows from [30, Proposition 3.1]. In the convex case, this observation has been recorded and used explicitly for example in [19, Proposition 2.2] and [22, Preliminaries], and even earlier in [2] and [37].

LEMMA 2.6 (Projection as a derivative). *Consider a set $Q \subset \mathbf{E}$ that is prox-regular at $\bar{x}$. Then the function*

$$h(x) := \frac{1}{2}\|x\|^2 - \frac{1}{2}d_Q^2(x),$$

is $\mathbf{C}^1$-smooth on a neighborhood of $\bar{x}$, with $\nabla h(x) = P_Q(x)$ for all x near $\bar{x}$.

We are now ready to state and prove the main result of this section.

THEOREM 2.7 (Spectral lifts of manifolds). *Consider a matrix $\overline{X} \in \mathbf{S}^n$ and a set $M \subset \mathbf{R}^n$ that is locally symmetric around $\bar{x} := \lambda(\overline{X})$. Then M is a $\mathbf{C}^\infty$ manifold around $\bar{x}$ if and only if the spectral set $\lambda^{-1}(M)$ is a $\mathbf{C}^\infty$ manifold around $\overline{X}$.*

Proof. Consider the function

$$h(x) := \frac{1}{2}\|x\|^2 - \frac{1}{2}d_M^2(x).$$

Suppose that M is a $\mathbf{C}^\infty$ manifold around $\bar{x}$. In particular M is prox-regular, see [31, Example 13.30]. Then using Theorem 2.4 we deduce that h is locally symmetric around $\bar{x}$. In turn, Lemma 2.6 implies the equality $\nabla h = P_M$ near $\bar{x}$. Since M is a $\mathbf{C}^\infty$ manifold, the projection mapping P_M is $\mathbf{C}^\infty$-smooth near $\bar{x}$. Combining this with Theorem 2.2, we deduce that the spectral function $h \circ \lambda$ is $\mathbf{C}^\infty$-smooth near $\overline{X}$. Observe

$$(h \circ \lambda)(X) = \frac{1}{2}\|\lambda(X)\|^2 - \frac{1}{2}d_M^2(\lambda(X))$$
$$= \frac{1}{2}\|X\|_F^2 - \frac{1}{2}d_{\lambda^{-1}(M)}^2(X),$$

where the latter equality follows from Theorem 2.4. Applying Theorem 2.4, we deduce that $\lambda^{-1}(M)$ is prox-regular at $\overline{X}$. Combining this with Lemma 2.6, we obtain equality $\nabla(h \circ \lambda)(X) = P_{\lambda^{-1}(M)}(X)$ for all X near $\overline{X}$. Consequently the mapping $X \mapsto P_{\lambda^{-1}(M)}(X)$ is $\mathbf{C}^\infty$-smooth near $\overline{X}$. Appealing to Lemma 2.5, we conclude that $\lambda^{-1}(M)$ is a $\mathbf{C}^\infty$ manifold. The proof of the converse implication is analogous. $\square$

REMARK 2.8. The proof of Theorem 2.7 falls short of establishing the lifting property for $\mathbf{C}^k$ manifolds, with k is finite, but not by much. The reason for that is that $\mathbf{C}^k$ manifolds yield projections that are only $\mathbf{C}^{k-1}$ smooth. Nevertheless, the same proof shows that $\mathbf{C}^k$ manifolds do lift to $\mathbf{C}^{k-1}$ manifolds, and conversely $\mathbf{C}^k$ manifolds project down by λ to $\mathbf{C}^{k-1}$ manifolds.

2.4. Dimension of the lifted manifold. The proof of Theorem 2.7 is relatively simple and short, unlike the involved proof of [11]. 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 [11] and [12]. 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 $x \in \mathbf{R}_\geq^n$ there exists a sequence

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

such that

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

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

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

As explained in [11, Section 2.2], the set of all such $\Delta_\mathcal{P}$'s defines an affine stratification of $\mathbf{R}_\geq^n$. Observe further that for every point $x \in \mathbf{R}_\geq^n$ we have

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

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

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

as one can easily check. Since the orbit $\lambda^{-1}(x)$ is isomorphic to $\mathbf{O}^n / \mathbf{O}_X^n$, it follows that it is a submanifold of $\mathbf{S}^n$. A computation, which can be found in [11], then yields the equation

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

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 $\mathbf{R}_{\ge}^n$. Let us further denote by Δ_* an affine stratum of the aforementioned stratification of $\mathbf{R}_{\ge}^n$ 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 [11, 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

$$\dim \lambda^{-1}(M) \ = \ \dim M \ + \sum_{1 \le i < j \le \rho_*} |I_i^*| \, |I_j^*|, \tag{2.3}$$

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

REMARK 2.9. It's worth to point out that it is possible to have strata $\Delta_1 \ne \Delta_2$ of $\mathbf{R}_{\ge}^n$ 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 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-Zhu [5], Clarke-Ledyaev-Stern-Wolenski [9], Mordukhovich [27], and Rockafellar-Wets [31] for more details. Unless otherwise stated, we follow the terminology and notation of [31].

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

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

and the *epigraph* of f is

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

We will say that f is *lower semicontinuous* (*lsc* for short) at a point $\bar{x}$ provided that the inequality $\liminf_{x \to \bar{x}} f(x) \ge f(\bar{x})$ holds. If f is lower semicontinuous at every point,

then we will simply say that f is lower semicontinuous. For any set Q, the symbols $\mathrm{cl}\,Q$, $\mathrm{conv}\,Q$, and $\mathrm{aff}\,Q$ will denote the topological closure, the convex hull, and the affine span of Q respectively. The symbol $\mathrm{par}\,Q$ will denote the parallel subspace of Q, namely the set $\mathrm{par}\,Q := (\mathrm{aff}\,Q) - Q$. For convex sets $Q \in \mathbf{E}$, the symbols $\mathrm{ri}\,Q$ and $\mathrm{rb}\,Q$ will denote the relative interior and the relative boundary of Q, respectively.

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 $\mathbf{F}$, we say that f is $\mathbf{C}^p$-*smooth* if for each point $\bar{x} \in Q$, there is a neighborhood U of $\bar{x}$ and a $\mathbf{C}^p$-smooth mapping $\widehat{f} \colon \mathbf{E} \to \mathbf{F}$ that agrees with f on $Q \cap U$.

Recall that by Theorem 2.2, smoothness of functions satisfies the Transfer Principle. Shortly, we will need a slightly strengthened version of this result, where smoothness is considered only relative to a certain locally symmetric subset. We record it now.

COROLLARY 3.1 (Lifts of restricted smoothness). *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* $\mathbf{C}^p$-*smooth* $(p = 1, \ldots, \infty)$ *around* $\bar{x}$ *if and only if the restriction of* $f \circ \lambda$ *to* $\lambda^{-1}(M)$ *is* $\mathbf{C}^p$-*smooth around* $\overline{X}$.

Proof. Suppose that the restriction of f to M is $\mathbf{C}^p$-smooth around $\bar{x}$. Then there exists a $\mathbf{C}^p$-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 $\mathbf{C}^p$-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 $\mathbf{C}^p$-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 $\mathbf{C}^p$-smooth around $\overline{X}$. Then there exists a $\mathbf{C}^p$-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 $\mathbf{C}^p$-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 $\mathbf{C}^p$-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. $\square$

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

DEFINITION 3.2 (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 $(x_i, f(x_i), v_i) \to (\bar{x}, f(\bar{x}), v)$.

Let us now recall from [24, Proposition 2] the following lemma, which shows that subdifferentials behave as one would expect in presence of symmetry.

LEMMA 3.3 (Subdifferentials under symmetry). *Consider a function* $f \colon \mathbf{R}^n \to \overline{\mathbf{R}}$ *that is locally symmetric at* $\bar{x}$. *Then the equation*

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

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

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

REMARK 3.4. In particular, if $f \colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is locally symmetric around $\bar{x}$, then the sets $\hat{\partial} f(\bar{x})$, $\mathrm{ri}\, \hat{\partial} f(\bar{x})$, $\mathrm{rb}\, \hat{\partial} f(\bar{x})$, $\mathrm{aff}\, \hat{\partial} f(\bar{x})$, and $\mathrm{par}\, \hat{\partial} f(\bar{x})$ are invariant under the action of the group $\mathrm{Fix}(\bar{x})$.

The following result is the cornerstone for the variational theory of spectral mappings [24, Theorem 6].

THEOREM 3.5 (Subdifferential under local symmetry). *Consider a lsc function* $f \colon \mathbf{R}^n \to \overline{\mathbf{R}}$ *and a symmetric matrix* $X \in \mathbf{S}^n$, *and suppose that* f *is locally symmetric at* $\lambda(X)$. *Then we have*

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

where

$$\mathbf{O}_X^n = \{ U \in \mathbf{O}^n : X = U^T (\mathrm{Diag}\, \lambda(X)) U \}.$$

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 3.6 (Directional 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 $v \in \partial f(\bar{x})$.
A set $Q \subset \mathbf{R}^n$ is *prox-regular at* $\bar{x}$ *for* $\bar{v} \in N_Q(\bar{x})$ provided that the indicator function δ_Q is prox-regular at $\bar{x}$ for $\bar{v}$.

In particular $\mathbf{C}^2$-smooth functions and lsc, convex functions are prox-regular at each of their points [31, Example 13.30, Proposition 13.34].

REMARK 3.7. A set $Q \subset \mathbf{E}$ is prox-regular at $\bar{x}$, in the sense above, if and only if it is prox-regular in the sense of Definition 2.3. For a proof, see for example [31, Exercise 13.38].

The following theorem shows that directional prox-regularity also satisfies the Transfer Principle [10, Theorem 4.2].

THEOREM 3.8 (Directional prox-regularity under spectral lifts).
Consider a lsc function $f \colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and a symmetric matrix $\bar{X}$. 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 two standard results of Linear Algebra will be important for us [24, Proposition 3].

LEMMA 3.9 (Simultaneous Conjugacy). *Consider vectors $x, y, u, v \in \mathbf{R}^n$. Then there exists an orthogonal matrix $U \in \mathbf{O}^n$ with*

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

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

COROLLARY 3.10 (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 $U_1, U_2 \in \mathbf{O}_X^n$ we have*

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

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

Proof. Observe

$$(U_1 U_2^T)^T \mathrm{Diag}\, v_1 (U_1 U_2^T) = \mathrm{Diag}\, v_2,$$

$$(U_1 U_2^T)^T \mathrm{Diag}\, \lambda(X) (U_1 U_2^T) = \mathrm{Diag}\, \lambda(X).$$

The result follows by an application of Lemma 3.9. $\square$

3.2. Main results. In this section, we consider partly-smooth sets, introduced in [25]. This notion generalizes the idea of active manifolds of classical nonlinear programming to an entirely representation-independent setting.

DEFINITION 3.11 (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 $\mathbf{C}^p$-*partly smooth* ($p = 2, \ldots, \infty$) at $\bar{x}$ relative to M if
   (i) **(Smoothness)** M is a $\mathbf{C}^p$ manifold around $\bar{x}$ and f restricted to M is $\mathbf{C}^p$-smooth near $\bar{x}$,
   (ii) **(Regularity)** f is prox-regular at $\bar{x}$,
   (iii) **(Sharpness)** the affine span of ∂f is a translate of $N_M(x)$,
   (iv) **(Continuity)** ∂f restricted to M is continuous at $\bar{x}$.
If the above properties hold, then we will refer to M as the *partly smooth manifold* of f at $\bar{x}$.

REMARK 3.12. 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 [21, 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 3.1 for an illustration.
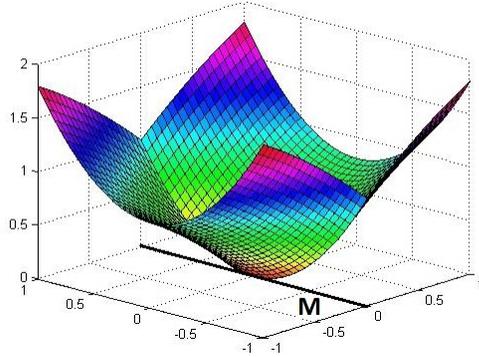

FIGURE 3.1. *The partly smooth manifold M for $f(x, y) := |x|(1 - |x|) + y^2$.*

It is reassuring to know that partly smooth manifolds are locally unique. This is the content of the following theorem [21, Corollary 4.2].

THEOREM 3.13 (Local uniqueness of partly smooth manifolds). *Consider a function $f \colon \mathbf{E} \to \overline{\mathbf{R}}$ that is $\mathbf{C}^p$-partly smooth $(p \geq 2)$ 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

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

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, 8, 7, 6, 18, 1, 20, 13], and has explicitly been studied in [16] for its own sake.

DEFINITION 3.14 (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 3.15. 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 3.16 (Partial smoothness and identifiability). *Consider a 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 $\mathbf{C}^p$ manifold $(p = 2, \ldots, \infty)$ containing $\bar{x}$, with the restriction $f\big|_M$ being $\mathbf{C}^p$-smooth near $\bar{x}$. Then the following are equivalent*

1. f is $\mathbf{C}^p$-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 3.17 (Spectral lifts of Identifiable sets). *Consider a 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 $\overline{U} \in \mathbf{O}^n_{\overline{X}}$. 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 3.10, there exists a permutation $\sigma \in \operatorname{Fix}(\bar{x})$ 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 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 $\overline{V} := \overline{U}^T(\operatorname{Diag} \bar{v})\overline{U}$. 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 $\sigma_i \in \operatorname{Fix}(\bar{x})$ satisfying $\sigma_i x_i \in \mathbf{R}_{\geq}$. Restricting to a subsequence, we may suppose that σ_i are equal to a fixed $\sigma \in \operatorname{Fix}(\bar{x})$. Define

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

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

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

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

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 $\overline{X} = (\overline{U}^T A_{\sigma^{-1}} \overline{U})^T \overline{X} (\overline{U}^T A_{\sigma^{-1}} \overline{U})$. Since $\lambda^{-1}(M)$ is identifiable (relative to F) at $\overline{X}$ for $(\overline{U}^T A_{\sigma^{-1}} \overline{U})^T \overline{V} (\overline{U}^T A_{\sigma^{-1}} \overline{U})$, 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. $\square$

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 3.18 (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*

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

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 $\mathbf{C}^\infty$-smooth around X and that $\lambda^{-1}(x+L)$ is a $\mathbf{C}^\infty$ manifold around X. Consequently $\hat{\partial}(g \circ \lambda)(X)$ is an affine subspace of $\mathbf{S}^n$. On the other hand, we have

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

thereby completing the proof. $\square$

PROPOSITION 3.19 (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*

$$\mathrm{aff}\, \hat{\partial}(f \circ \lambda)(X) = \{U^T (\mathrm{Diag}\, v) U : v \in \mathrm{aff}\, \hat{\partial} f(\lambda(X)) \text{ and } U \in \mathbf{O}_X^n\}, \qquad (3.1)$$

$$\mathrm{rb}\, \hat{\partial}(f \circ \lambda)(X) = \{U^T (\mathrm{Diag}\, v) U : v \in \mathrm{rb}\, \hat{\partial} f(\lambda(X)) \text{ and } U \in \mathbf{O}_X^n\}. \qquad (3.2)$$

$$\mathrm{ri}\, \hat{\partial}(f \circ \lambda)(X) = \{U^T (\mathrm{Diag}\, v) U : v \in \mathrm{ri}\, \hat{\partial} f(\lambda(X)) \text{ and } U \in \mathbf{O}_X^n\}. \qquad (3.3)$$

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

$$\hat{\partial}(f \circ \lambda)(X) \subset \{U^T(\mathrm{Diag}\, v)U : v \in \mathrm{aff}\, \hat{\partial}f(\lambda(X)) \text{ and } U \in \mathbf{O}_X^n\}.$$

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

$$\{U^T(\mathrm{Diag}\, v)U : v \in \mathrm{aff}\, \hat{\partial}f(\lambda(X)) \text{ and } U \in \mathbf{O}_X^n\},$$

is an affine subspace; but this is immediate from Remark 3.4 and Lemma 3.18. Hence (3.1) holds.

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

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

Consequently, by Corollary 3.10, there exists a permutation $\sigma \in \mathrm{Fix}(x)$ satisfying $\sigma \tilde{v} = v$. Since $\hat{\partial}f(x)$ is invariant under the action of $\mathrm{Fix}(x)$, it follows that $\tilde{v}$ lies in $\mathrm{rb}\, \hat{\partial}f(x)$, and consequently from Remark 3.4 we deduce $v \in \mathrm{rb}\, \hat{\partial}f(x)$. This establishes the inclusion $\subset$ of (3.2). To see the reverse inclusion, consider a sequence $v_i \in \mathrm{aff}\, \hat{\partial}f(x)$ converging to $v \in \hat{\partial}f(x)$ with $v_i \notin \hat{\partial}f(x)$ for each index i. Fix an arbitrary matrix $U \in \mathbf{O}_X^n$ and observe that the matrices $U^T(\mathrm{Diag}\, v_i)U$ lie in $\mathrm{aff}\, \hat{\partial}(f \circ \lambda)(x)$ and converge to $U^T(\mathrm{Diag}\, v)U$. We now claim that the matrices $U^T(\mathrm{Diag}\, v_i)U$ all lie outside of $\hat{\partial}(f \circ \lambda)(x)$. Indeed suppose this is not the case. Then there exist matrices $\widetilde{U}_i \in \mathbf{O}_X^n$ and subgradients $v_i \in \hat{\partial}f(x)$ satisfying

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

An application of Corollary 3.10 and Remark 3.4 then yields a contradiction. Therefore the inclusion $U^T(\mathrm{Diag}\, v)U \in \mathrm{rb}\, \hat{\partial}(f \circ \lambda)(X)$ 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 equation (3.2). To see the reverse inclusion, consider a matrix $U^T(\mathrm{Diag}\, v)U$, for some $U \in \mathbf{O}_X^n$ and $v \in \mathrm{ri}\, \hat{\partial}f(x)$. Again, an easy application of Corollary 3.10 and Remark 3.4 yields the inclusion $U^T(\mathrm{Diag}\, v)U \in \mathrm{ri}\, \hat{\partial}(f \circ \lambda)(X)$. We conclude that (3.3) holds. $\square$

LEMMA 3.20 (Symmetry of partly smooth manifolds). *Consider a lsc function* $f \colon \mathbf{R}^n \to \overline{\mathbf{R}}$ *that is locally symmetric at* $\bar{x}$. *Suppose that* f *is* $\mathbf{C}^p$*-partly smooth at* $\bar{x}$ *relative to* M. *Then* M *is locally symmetric around* $\bar{x}$.

Proof. Consider a permutation $\sigma \in \mathrm{Fix}(\bar{x})$. Then the function f is partly smooth at $\bar{x}$ relative to σM. On the other hand, partly smooth manifolds are locally unique Theorem 3.13. Consequently we deduce equality $M = \sigma M$ locally around $\bar{x}$. The claim follows. $\square$

The main result of this section is now immediate.

THEOREM 3.21 (Lifts of $\mathbf{C}^\infty$-partly smooth functions). *Consider a lsc 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* $\mathbf{C}^\infty$-*partly smooth at* $\bar{x}$ *relative to* M *if and only if* $f \circ \lambda$ *is* $\mathbf{C}^\infty$-*partly smooth at* $\overline{X}$ *relative to* $\lambda^{-1}(M)$.

Proof. Suppose that f is $\mathbf{C}^\infty$-partly smooth at $\bar{x}$ relative to M. In light of Lemma 3.20, we deduce that M is locally symmetric at $\bar{x}$. Consequently, Theorem 2.7 implies that the set $\lambda^{-1}(M)$ is a $\mathbf{C}^\infty$ manifold, while Corollary 3.1 implies that $f \circ \lambda$ is $\mathbf{C}^\infty$-smooth on $\lambda^{-1}(M)$ near $\overline{X}$. Applying Theorem 3.8, 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 3.19, there exists a vector $v \in \operatorname{ri} \partial f(\bar{x})$ and a matrix $U \in \mathbf{O}_{\overline{X}}^n$ satisfying

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

Observe by Proposition 3.16, the set M is identifiable at $\bar{x}$ for $\bar{v}$. Then applying Proposition 3.17, 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 3.16, we deduce that $f \circ \lambda$ is $\mathbf{C}^\infty$-partly smooth at $\overline{X}$ relative to $\lambda^{-1}(M)$, as claimed. The converse follows along the same lines. $\square$

The forward implication of Theorem 3.21 holds in the case of $\mathbf{C}^p$-partly smooth functions (for $p = 2, \ldots, \infty$). The proof is identical except one needs to use [11, Theorem 4.21] instead of Theorem 2.7. We record this result for ease of reference in future works.

THEOREM 3.22 (Lifts of $\mathbf{C}^p$-partly smooth functions). *Consider a lsc 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* $\bar{x} := \lambda(\overline{X})$. *If* f *is* $\mathbf{C}^p$-*partly smooth (for* $p = 2, \ldots, \infty$) *at* $\bar{x}$ *relative to* M, *then* $f \circ \lambda$ *is* $\mathbf{C}^\infty$-*partly smooth at* $\overline{X}$ *relative to* $\lambda^{-1}(M)$.

4. Partly smooth duality for polyhedrally generated spectral functions. Consider a lsc, convex function $f\colon \mathbf{E} \to \overline{\mathbf{R}}$. Then the *Fenchel conjugate* $f^*\colon \mathbf{E} \to \overline{\mathbf{R}}$ 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 $\mathcal{J}_f\colon \mathbb{P}(\mathbf{E}) \to \mathbb{P}(\mathbf{E})$ 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 [4, 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*

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

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

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

If the strata are open polyhedra, then $\mathcal{A}$ is a *polyhedral stratification*. If the strata are $\mathbf{C}^k$ manifolds, then $\mathcal{A}$ is a $\mathbf{C}^k$-*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 $\mathbf{C}^\infty$-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^*. Little thought, however shows an immediate obstruction: images of $\mathbf{C}^\infty$-smooth manifolds under the map $\mathcal{J}_f$ may fail to be even $\mathbf{C}^2$-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 the Figures 4.1 and 4.2 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^*.

In the sequel, let us denote by

$$M^{\operatorname{sym}} := \bigcup_{\sigma \in \Sigma} \sigma M$$
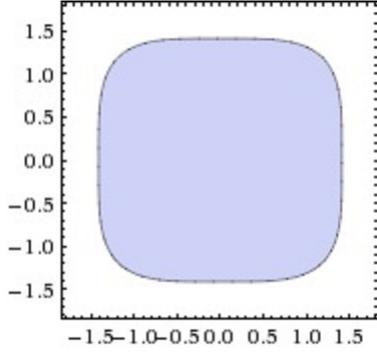
FIGURE 4.1. $\{(x, y) : x^4 + y^4 \leq 4\}$
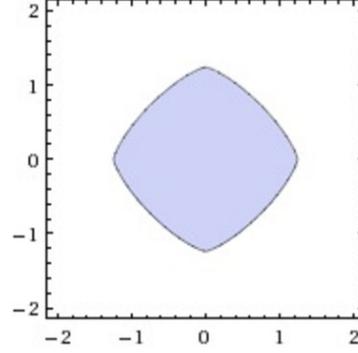
FIGURE 4.2. $\{(x, y) : |x|^{\frac{4}{3}} + |y|^{\frac{4}{3}} \leq \frac{4}{3}\}$

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 in 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 $\sigma_1 M = M$ 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. $\square$

Stratification duality for spectral lifts. Consider a symmetric, convex polyhedral function $f : \mathbf{R}^n \to \overline{\mathbf{R}}$ 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 $f : \mathbf{R}^n \to \overline{\mathbf{R}}$ and the induced stratification $\mathcal{A}_f$ of $\mathrm{dom}\, f$. Then the following are true.*

(i) *For any set $M_1, M_2 \in \mathcal{A}_f$ 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 $\mathbf{C}^\infty$ 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 $\mathbf{C}^\infty$ manifolds, which fail to be symmetric in general. In light of Lemma 4.5, the set M^{sym} is a $\mathbf{C}^\infty$ 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 $\mathrm{dom}\, f$, 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 $\mathrm{dom}\, f$ induced by f, and define the collection*

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

Then the following properties hold:

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

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

(iii) *The assignment $\mathcal{J}_F \colon \mathbb{P}(\mathbf{S}^n) \to \mathbb{P}(\mathbf{S}^n)$ 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:*

$$
\begin{array}{ccc}
\mathcal{A}_F & \xrightarrow{\ \mathcal{J}_F\ } & \mathcal{A}_{F^*} \\[2pt]
\Big\uparrow{\lambda^{-1}} & & \Big\uparrow{\lambda^{-1}} \\[2pt]
\mathcal{A}_f^{\mathrm{sym}} & \xrightarrow[\ \mathcal{J}_f\ ]{} & \mathcal{A}_{f^*}^{\mathrm{sym}}
\end{array}
$$

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 $\mathbf{C}^\infty$ manifold. The fact that $\mathcal{A}_F$ is a $\mathbf{C}^\infty$-stratification of $\mathrm{dom}\, F$ 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 3.21, we deduce that F is $\mathbf{C}^\infty$-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:

$$\mathcal{J}_F(\lambda^{-1}(M^{\mathrm{sym}})) = \bigcup_{X \in \lambda^{-1}(M^{\mathrm{sym}})} \mathrm{ri}\, \partial F(X)$$

$$= \bigcup_{X \in \lambda^{-1}(M^{\mathrm{sym}})} \{U^T(\mathrm{Diag}\, v)U : v \in \mathrm{ri}\, \partial f(\lambda(X)) \text{ and } U \in \mathbf{O}_X^n\},$$

and concurrently,

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

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 $v \in \mathrm{ri}\, \partial f(x)$, and a matrix $U \in \mathbf{O}^n$. We must show $V := U^T(\mathrm{Diag}\, v)U \in \mathcal{J}_F(\lambda^{-1}(M^{\mathrm{sym}}))$. To see this, fix a permutation $\sigma \in \Sigma$ with $\sigma x \in \mathbf{R}_\geq^n$, and observe

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

where A_σ denotes the matrix representing the permutation σ. Define a matrix $X := (A_\sigma U)^T(\mathrm{Diag}\, \sigma x)A_\sigma U$. Clearly, we have $V \in \mathrm{ri}\, \partial F(X)$ 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 $\mathcal{J}_F \colon \mathcal{A}_F \to \mathcal{A}_{F^*}$. $\square$

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

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

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 $\langle X, Y \rangle = \mathrm{tr}\,(X^T Y)$, 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

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

Recall that *singular values* of a matrix $A \in \mathbf{M}^{n \times m}$ are defined to be the square roots of the eigenvalues of the matrix $A^T A$. The *singular value mapping* $\sigma \colon \mathbf{M}^{n \times m} \to \mathbf{R}^m$ is simply the mapping taking each matrix X to its vector $(\sigma_1(X), \ldots, \sigma_m(X))$ of singular values in non-increasing 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 can necessarily be represented 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 essentially 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 $\bar{X} \in \mathbf{M}^{n \times m}$ and a set $M \subset \mathbf{R}^m$ that is locally absolutely permutation-invariant around $\bar{x} := \sigma(\bar{X})$. Then M is a $\mathbf{C}^\infty$ manifold around $\bar{x}$ if and only if the set $\sigma^{-1}(M)$ is a $\mathbf{C}^\infty$ manifold around $\bar{X}$.*

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

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

It is unknown whether the analogue of the latter theorem holds in the case of $\mathbf{C}^p$ partial smoothness, where $p < \infty$. This is so because it is unknown whether a nonsymmetric analogue of [11, Theorem 4.21] holds in case of functions that are differentiable only finitely many times.

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.

REFERENCES

[1] F. Al-Khayyal and J. Kyparisis. Finite convergence of algorithms for nonlinear programs and variational inequalities. *J. Optim. Theory Appl.*, 70(2):319–332, 1991.

[2] E. Asplund. Differentiability of the metric projection in finite-dimensional Euclidean space. *Proc. Amer. Math. Soc.*, 38:218–219, 1973.

[3] J. Bolte, A. Daniilidis, and A.S. Lewis. Generic optimality conditions for semialgebraic convex programs. *Math. Oper. Res.*, 36:55–70, 2011.

[4] J.M. Borwein and A.S. Lewis. *Convex analysis and nonlinear optimization.* CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 3. Springer-Verlag, New York, 2000. Theory and examples.

[5] J.M. Borwein and Q.J. Zhu. *Techniques of Variational Analysis.* Springer-Verlag, New York, 2005.

[6] J.V. Burke. On the identification of active constraints. II. The nonconvex case. *SIAM J. Numer. Anal.*, 27(4):1081–1103, 1990.

[7] J.V. Burke and J.J. Moré. On the identification of active constraints. *SIAM J. Numer. Anal.*, 25(5):1197–1211, 1988.

[8] P.H. Calamai and J.J. Moré. Projected gradient methods for linearly constrained problems. *Math. Program.*, 39(1):93–116, 1987.

[9] F.H. Clarke, Yu. Ledyaev, R.I. Stern, and P.R. Wolenski. *Nonsmooth Analysis and Control Theory.* Texts in Math. 178, Springer, New York, 1998.

[10] A. Daniilidis, A.S. Lewis, J. Malick, and H. Sendov. Prox-regularity of spectral functions and spectral sets. *J. Convex Anal.*, 15(3):547–560, 2008.

[11] A. Daniilidis, J. Malick, and H.S. Sendov. Locally symmetric submanifolds lift to spectral manifolds. *Preprint U.A.B.* **23**/*2009, 43 p., arXiv:1212.3936 [math.OC]*.

[12] A. Daniilidis, J. Malick, and H.S. Sendov. On the structure of locally symmetric manifolds. *(In preparation)*.

[13] A. Daniilidis, C. Sagastizábal, and M. Solodov. Identifying structure of nonsmooth convex functions by the bundle technique. *SIAM J. Optim.*, 20(2):820–840, 2009.

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

[15] D. Drusvyatskiy and M. Larsson. Approximating functions on stratifiable sets. *Under review, arXiv:1207.5258 [math.CA]*, 2012.

[16] D. Drusvyatskiy and A.S. Lewis. Optimality, identifiability, and sensitivity. *Under review, arXiv:1207.6628 [math.OC]*.

[17] J.C. Dunn. On the convergence of projected gradient processes to singular critical points. *J. Optim. Theory Appl.*, 55(2):203–216, 1987.

[18] M.C. Ferris. Finite termination of the proximal point algorithm. *Math. Program. Ser. A*, 50(3):359–366, 1991.

[19] S. Fitzpatrick and R.R. Phelps. Differentiability of the metric projection in Hilbert space. *Trans. Amer. Math. Soc.*, 270(2):483–501, 1982.

[20] S.D. Flåm. On finite convergence and constraint identification of subgradient projection methods. *Math. Program.*, 57:427–437, 1992.

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

[22] R.B. Holmes. Smoothness of certain metric projections on hilbert space. *Trans. Amer. Math. Soc.*, 184:pp. 87–100, 1973.

[23] A.S. Lewis. Derivatives of spectral functions. *Math. Oper. Res.*, 21(3):576–588, 1996.

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

[25] A.S. Lewis. Active sets, nonsmoothness, and sensitivity. *SIAM J. Optim.*, 13:702–725, 2002.

[26] A.S. Lewis and H.S. Sendov. Nonsmooth analysis of singular values. I. Theory. *Set-Valued Anal.*, 13(3):213–241, 2005.

[27] B.S. Mordukhovich. *Variational Analysis and Generalized Differentiation I: Basic Theory.* Grundlehren der mathematischen Wissenschaften, Vol 330, Springer, Berlin, 2006.

[28] R.R. Phelps. *Convex functions, monotone operators and differentiability*, volume 1364 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin, 2nd edition, 1993.

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

[30] R.A. Poliquin, R.T. Rockafellar, and L. Thibault. Local differentiability of distance functions. *Trans. Amer. Math. Soc.*, 352(11):5231–5249, 2000.

[31] R.T. Rockafellar and R.J-B. Wets. *Variational Analysis.* Grundlehren der mathematischen Wissenschaften, Vol 317, Springer, Berlin, 1998.

[32] H.S. Sendov. The higher-order derivatives of spectral functions. *Linear Algebra Appl.*, 424(1):240–281, 2007.

[33] M. Šilhavý. Differentiability properties of isotropic functions. *Duke Math. J.*, 104(3):367–373, 2000.

[34] J. Sylvester. On the differentiability of O(n) invariant functions of symmetric matrices. *Duke Math. J.*, 52(2):475–483, 1985.

[35] J. von Neumann. Some matrix inequalities and metrization of matrix-space. *Tomck. Univ. Rev.*, 1:286–300, 1937.

[36] S.J. Wright. Identifiable surfaces in constrained optimization. *SIAM J. Control Optim.*, 31:1063–1079, July 1993.

[37] E.H. Zarantonello. Projections on convex sets in Hilbert space and spectral theory. I. Projections on convex sets. In *Contributions to nonlinear functional analysis (Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1971)*, pages 237–341. Academic Press, New York, 1971.