SIAM J. OPTIM.
Vol. 13, No. 3, pp. 702–725

ACTIVE SETS, NONSMOOTHNESS, AND SENSITIVITY*

A. S. LEWIS†

Abstract. Nonsmoothness pervades optimization, but the way it typically arises is highly structured. Nonsmooth behavior of an objective function is usually associated, locally, with an *active manifold*: on this manifold the function is smooth, whereas in normal directions it is "vee-shaped." Active set ideas in optimization depend heavily on this structure. Important examples of such functions include the pointwise maximum of some smooth functions and the maximum eigenvalue of a parametrized symmetric matrix. Among possible foundations for practical nonsmooth optimization, this broad class of "partly smooth" functions seems a promising candidate, enjoying a powerful calculus and sensitivity theory. In particular, we show under a natural regularity condition that critical points of partly smooth functions are stable: small perturbations to the function cause small movements of the critical point on the active manifold.

Key words. active set, nonsmooth analysis, subdifferential, generalized gradient, sensitivity, $\mathcal{U}$-Lagrangian, eigenvalue optimization, spectral abscissa, identifiable surface

AMS subject classifications. Primary, 90C31, 49K40; Secondary, 65K10, 15A42

PII. S1052623401387623

1. Introduction.

Optimality conditions throughout the field of optimization are intimately bound up with nonsmoothness. As a simple example, consider the problem of minimizing a sum of Euclidean norms (cf. [1]):

$$\min_{x\in\mathbf{R}^n} h(x) := \sum_{i=1}^{k} \|A^i x - b^i\|$$

for given matrices A^i and vectors b^i. Except at the origin, the Euclidean norm is a *smooth* function, by which we will always mean twice continuously differentiable. Yet its nonsmoothness is crucial to any understanding of this problem. Associated with an optimal solution x_0 is an "active set" $\{i : A^i x_0 = b^i\}$, often nonempty, so the objective function h is nonsmooth at x_0. Furthermore, under reasonable conditions this active set is stable under small perturbations to the problem. (See [6, 20] for active set algorithms.)

This particular problem could be rephrased as a conic quadratic program, amenable to contemporary interior point techniques [1, 3]. Nonetheless, as in linear programming, the active set is an important tool for understanding the problem.

This phenomenon of nonsmoothness inducing a certain "activity" central to optimality conditions repeats many times throughout optimization. Consider the following examples.

(a) *Classical nonlinear programming and minimax*. At an optimal solution of a nonlinear constrained optimization problem, some subset of the inequality constraints is active (that is, those constraints hold with equality): under reasonable conditions (see, for example, [8]), this active set is stable under small perturbations to the problem.

*Received by the editors April 6, 2001; accepted for publication (in revised form) May 6, 2002; published electronically November 14, 2002. This research was supported by the NSERC.
http://www.siam.org/journals/siopt/13-3/38762.html
†Department of Mathematics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada (aslewis@sfu.ca, http://www.math.sfu.ca/∼aslewis).

Somewhat analogously, consider a nonlinear minimax problem

$$\min_{x \in \mathbf{R}^n} h(x) := \max_{i=1,2,\ldots,k} h_i(x,u),$$

where each function h_i is smooth and u denotes a vector of parameters. Under reasonable conditions the active set at an optimal solution x_0,

$$I(x_0,u) := \{i : h_i(x_0,u) = h(x_0)\},$$

is stable under small changes in u.

(b) *Sums of norms.* Rather more generally than our initial example, we could consider the problem

$$\min_{x \in \mathbf{R}^n} h(x) := \sum_{i=1}^{k} \|F_i(x,u)\|,$$

where each function F_i is smooth and u denotes a vector of parameters. Under reasonable conditions the active set at an optimal x_0,

$$I(x_0,u) := \{i : F_i(x_0,u) = 0\},$$

is stable under small changes in u. Any smooth norm could be used in place of the Euclidean norm (cf. [7]).

(c) *Semidefinite programming and eigenvalue optimization.* The primal variable in a semidefinite program is a positive semidefinite matrix (see [16], for example). An optimal solution has a zero eigenvalue with a certain multiplicity: under reasonable conditions, this multiplicity is stable under small perturbations to the problem.

Relatedly, consider the eigenvalue optimization problem (see, for example, [14])

$$\min_{x \in \mathbf{R}^n} h(x) := \lambda_1(F(x,u)),$$

where the smooth function F takes real symmetric matrix values, u denotes a vector of parameters, and the function $\lambda_1(\cdot)$ is the largest eigenvalue. At an optimal solution this largest eigenvalue has a certain multiplicity, which under reasonable conditions is stable under small changes in u.

(d) *Spectral abscissa minimization.* More generally, consider the problem

$$\min_{x \in \mathbf{R}^n} h(x) := \alpha(F(x,u)),$$

where F now takes arbitrary square matrix values and the function $\alpha(\cdot)$ is the *spectral abscissa* (the largest real part of an eigenvalue). An optimal matrix generally has several distinct "active" eigenvalues with real part equal to its spectral abscissa, and each such eigenvalue has an associated algebraic multiplicity (the geometric multiplicity typically being one): under reasonable conditions this pattern of multiplicities is stable under small changes in u (see [5]).

Each of these problems has optimal solutions with a corresponding "activity," which is stable under small perturbations to the problem. In nonlinear minimax or sums of norms, the activity consists of subsets of indices; in eigenvalue optimization or spectral abscissa minimization, it consists of a certain pattern of multiplicities. These "activities" have powerful algorithmic significance: in each case, once the activity of an optimal solution is known, finding it (at least locally) is a *smooth* minimization problem.

Let us summarize. The problem of minimizing a nonsmooth function is common in practice. But the nonsmoothness of a typical such function is highly structured: it induces a certain "activity" at an optimal solution, which under reasonable conditions is stable under small perturbations to the problem. Once the activity is known, the optimization problem is locally smooth.

The central idea of this current work is that the "activity" corresponds to a *manifold*. Each of the functions h above is what we will call *partly smooth*. Specifically, in a neighborhood of the point of interest x_0 there is a manifold $\mathcal{M}$ (the *active manifold*) containing x_0, with certain properties. Loosely speaking, the function h behaves smoothly as we move on the active manifold $\mathcal{M}$ and "sharply" if we move normal to the manifold; furthermore, in any fixed direction its directional derivative behaves continuously as we move on $\mathcal{M}$ and upper semicontinuously if we allow perturbations off it. (For closed convex functions, for example, this latter semicontinuity property, known as "regularity," is automatic.) We give the precise description in Definition 2.7.

The idea of partial smoothness at first sight appears rather intricate, but we shall find many interesting examples in practice. Each of our four examples is partly smooth under reasonable conditions. Given the parameter vector u, the four active manifolds are defined near x_0 as follows:

(a) $\{x : h_i(x,u) = h_j(x,u) \text{ for all } i,j \in I(x_0,u)\}$;
(b) $\{x : F_i(x,u) = 0 \text{ for all } i \in I(x_0,u)\}$;
(c) $\{x : \lambda_1(F(x,u)) \text{ has same multiplicity as } \lambda_1(F(x_0,u))\}$;
(d) $\{x : F(x,u) \text{ has same active eigenvalue multiplicities as } F(x_0,u)\}$.

Furthermore, we shall see that partly smooth functions have a robust calculus. Thus they form a rich, practical class of nonsmooth functions.

The literature contains many classes of nonsmooth functions more open to analysis than general, potentially pathological nonsmooth functions. A useful example is "amenability" [23, Def. 10.23], a powerful notion for combining smooth and convex techniques, again with a robust calculus. As we shall see, the real function $\sqrt{|\cdot|}$ is partly smooth at the origin relative to the active manifold $\{0\}$, but it is not amenable at the origin (see [23, Ex. 10.25(a)]), and it is not hard to construct similar Lipschitz examples using the fact that amenable functions are locally regular [23, Ex. 10.25(b)]. On the other hand, the convex, piecewise linear-quadratic function $x \mapsto \|x\|_1^2$ is not partly smooth relative to any manifold containing the origin.

The distinctive feature of partial smoothness is the notion of the active manifold: it is this idea that decouples the smooth behavior of the function from its "sharp" behavior. The importance of this general structure was realized for convex functions in [22], although not rigorously developed. The notion of active manifold is also implicit in the approach to polyhedral minimization via "structure functionals" [17]. In the nonconvex case the notion of the active manifold is familiar from active set methods for classical nonlinear programming (see [9], for example). In eigenvalue optimization the role of the active manifold is well known; see [21] and [24], for example. In spectral abscissa minimization, the idea is used heavily in [5].

For *convex* functions, partial smoothness is closely related to the "$\mathcal{U}$-Lagrangian" techniques of [12]: the active manifold is the "gully-shaped valley" of that work, and the normal and tangent spaces to the manifold correspond to the "$\mathcal{U}-\mathcal{V}$ decomposition" originating with the earlier work in [13] and developed for the maximum eigenvalue in [18, 19]. The idea of a "fast track" [15] is also closely related. We link the $\mathcal{U}$-Lagrangian theory to partial smoothness towards the end of the current work. Notice, however, that many interesting examples of partly smooth functions are not convex: convexity is *not* the real driving force behind this theory.

Another closely related idea is the notion of an "identifiable surface" of a convex set [26], which is a subset of the boundary having a suitable "sharpness" property. In [26] it is shown that, if the solution of an optimization problem posed over such a set lies in an identifiable surface, then various standard constrained optimization algorithms "identify" the surface after a finite number of iterations. Hence the idea of identifiability is a powerful tool for algorithmic analysis.

Remarkably, as we shall see, for convex sets, the ideas of identifiability and partial smoothness coincide, reinforcing the power of this theory. By contrast with identifiability, however, partial smoothness is defined in a more geometric manner, and once again is *not* dependent on convexity.

To demonstrate the power of partly smooth techniques, our culminating result is a sensitivity theorem. In classical nonlinear programming, if a local minimizer has linearly independent active constraints and satisfies strict complementarity and a strong second-order condition, then the minimizer depends smoothly on the parameters of the problem (see [8], for example). An analogous result for eigenvalue optimization appears in [24] and for spectral abscissa minimization, in [5]. Our work here shows how partial smoothness unifies this work. To sketch the idea, suppose the function h is partly smooth at a point x_0 relative to the active manifold $\mathcal{M}$. If x_0 is a strong second-order minimizer of the smooth, restricted function $h|_{\mathcal{M}}$, and is a "sharp" minimizer of the restriction to the normal space $h|_{x_0+N_{\mathcal{M}}(x_0)}$, then the critical point x_0 varies smoothly over $\mathcal{M}$ as the parameters of the problem vary. Back in the context of nonlinear programming, our "sharp minimizer" condition corresponds to the usual strict complementarity condition, and the usual linear independence assumption becomes a transversality condition allowing us to apply a chain rule.

The proof of our sensitivity result amounts to local reduction to a smooth equality-constrained problem. Such a reduction is a standard approach to sensitivity results in nonlinear programming (see [4, Rem. 4.127], for example), and also works in semidefinite programming [4, p. 495]. By comparison, we are able here to consider rather general optimization problems, and without recourse to general nonsmooth second-order theory (such as [23, Chap. 13], for example), but the price of this generality is that we must settle for critical points in our sensitivity analysis, rather than local minimizers (see the example in section 7).

Partial smoothness seems a promising framework for practical nonsmooth optimization. Partly smooth functions form a wide and robust class, with many of the properties sought by previously cited researchers interested in algorithm development, stemming from the decoupling of the smooth and sharp behaviors. We defer algorithmic discussion to a later work.

2. Partial smoothness.

We begin with some elementary definitions. We follow the notation and terminology of [23] throughout.

We consider a fixed Euclidean space X (a finite-dimensional real inner product space). We denote the subspace parallel to a nonempty convex set $C \subset X$ by $\operatorname{par} C$. Thus for any point $x \in C$ we have

$$\operatorname{par} C = (\operatorname{aff} C) - x = \mathbf{R}(C - C) = \mathbf{R}_+(C - C),$$

where $\operatorname{aff} C$ is the affine span of C. Easy exercises show $\operatorname{par}(AC) = A \operatorname{par} C$ for any linear map A, and $\operatorname{par}(C_1 \times C_2) = \operatorname{par} C_1 \times \operatorname{par} C_2$ for arbitrary nonempty convex sets C_1 and C_2. We denote the extended reals by $\overline{\mathbf{R}} = [-\infty, +\infty]$. The *lineality space* of a sublinear function $f : X \to \overline{\mathbf{R}}$ is the subspace

$$\operatorname{lin} f = \{w \in X : f(w) = -f(-w)\}.$$

Let us consider a function $h : X \to \overline{\mathbf{R}}$, finite at a point $x \in X$. We review some definitions from [23]. The *subderivative* $dh(x)(\cdot) : X \to \overline{\mathbf{R}}$ is defined by

$$dh(x)(\bar{w}) = \liminf_{\tau \downarrow 0,\ w \to \bar{w}} \frac{h(x + \tau w) - h(x)}{\tau} \quad (\bar{w} \in X)$$

and the set of *regular subgradients* is (see [23, Ex. 8.4])

$$\hat{\partial} h(x) = \{v \in X : \langle v, w \rangle \leq dh(x)(w) \text{ for all } w \in X\}.$$

The set of *subgradients* is

$$\partial h(x) = \left\{ \lim_r v_r : v_r \in \hat{\partial} h(x_r),\ x_r \to x,\ h(x_r) \to h(x) \right\},$$

while the set of *horizon subgradients* is

$$\partial^\infty h(x) = \left\{ \lim_r \lambda_r v_r : v_r \in \hat{\partial} h(x_r),\ x_r \to x,\ h(x_r) \to h(x),\ \lambda_r \downarrow 0 \right\}.$$

Suppose in addition $\partial h(x) \neq \emptyset$. Then h is *(subdifferentially) regular* at x if h is locally lower semicontinuous around x, every subgradient is regular, and furthermore the recession cone (in the sense of convex analysis) $h(x)^\infty$ coincides with $\partial^\infty h(x)$ (see [23, Cor. 8.11]). In this case, the support function of $\partial h(x)$ is the subderivative $dh(x)$ [23, Thm. 8.30]. This is the case in particular for any closed convex function h, and in this case ∂h is the usual subdifferential in the sense of convex analysis.

PROPOSITION 2.1 (lineality space of subderivative). *If the function h is regular at the point $x \in X$, and has a subgradient there, then*

$$\operatorname{lin} dh(x) = (\operatorname{par} \partial h(x))^\perp.$$

Proof. We know $w \notin \operatorname{lin} dh(x)$ if and only if $dh(x)(w) + dh(x)(-w) > 0$, which by [23, Thm. 8.30] is equivalent to the existence of subgradients y and z of h at x satisfying $\langle y - z, w \rangle > 0$, or equivalently $w \notin (\partial h(x) - \partial h(x))^\perp$. The result follows. $\square$

Given a set $\mathcal{M} \subset X$ containing a point x, we call a function $f : \mathcal{M} \to \overline{\mathbf{R}}$ *smooth around* x if x has an open neighborhood V in X such that some smooth function $g : V \to \mathbf{R}$ agrees with f on $\mathcal{M} \cap V$. We call such a function g a *smooth representative* of f around x. Note that in this case f is also smooth around any nearby point in $\mathcal{M}$. We call x a *critical point* of f if

$$f(z) - f(x) = o(\|z - x\|) \text{ for } z \text{ close to } x \text{ in } \mathcal{M}.$$

We call the function f *smooth* if it is smooth around every point in $\mathcal{M}$.

A "manifold" in X, loosely speaking, is a set consisting locally of the solutions of some smooth equations with linearly independent gradients. To be more precise, we say that a set $\mathcal{M} \subset X$ *is a manifold (of codimension m) around* a point $x \in X$ if $x \in \mathcal{M}$ and there is an open set $V \subset X$ containing x such that

$$\mathcal{M} \cap V = \{x \in V : F(x) = 0\},$$

where the smooth function $F : V \to \mathbf{R}^m$ has surjective derivative throughout V. In this case, the *tangent space* to $\mathcal{M}$ at x is given by

$$T_{\mathcal{M}}(x) = \mathrm{Ker}(\nabla F(x))$$

(which is independent of the choice of F), and the *normal space* to $\mathcal{M}$ at x is the orthogonal complement of the tangent space, namely

$$N_{\mathcal{M}}(x) = R(\nabla F(x)^*)$$

(where $R(\cdot)$ denotes range). The set $\mathcal{M}$ is then Clarke regular at x, and its normal cone there is exactly the normal space [23, Ex. 6.8].

We call a set $\mathcal{M}$ a *manifold (of codimension m)* if $\mathcal{M}$ is a manifold of codimension m around every point in $\mathcal{M}$. (More precisely, $\mathcal{M}$ is an "m-codimensional manifold embedded in X"; see [25].) If $\mathcal{M}$ is a manifold around a point x, then $\mathcal{M} \cap U$ is a manifold for some open neighborhood $U \subset X$ of x.

If the function $f : \mathcal{M} \to \overline{\mathbf{R}}$ is smooth around x and $\mathcal{M}$ is a manifold around x, then x is a critical point of f if and only if

$$\nabla g(x) \in N_{\mathcal{M}}(x),$$

where g is any smooth representative of f around x. In particular, this holds if x is a local minimizer of f.

The indicator function $\delta_{\mathcal{M}}$ takes the value 0 on $\mathcal{M}$ and $+\infty$ otherwise.

PROPOSITION 2.2 (subgradients and smoothness). *Suppose the set $\mathcal{M} \subset X$ is a manifold around the point $x \in \mathcal{M}$. For a function $h : X \to \overline{\mathbf{R}}$, if the restriction $h|_{\mathcal{M}}$ is smooth around x, then*

$$\hat{\partial} h(x) \subset \nabla g(x) + N_{\mathcal{M}}(x) \tag{2.3}$$

for any smooth representative g of $h|_{\mathcal{M}}$ around x, and hence

$$\mathrm{par}\, \hat{\partial} h(x) \subset N_{\mathcal{M}}(x).$$

Proof. For some open neighborhood V of x we have $g + \delta_{\mathcal{M}\cap V} = h + \delta_{\mathcal{M}\cap V}$, so by [23, Cor. 10.9] we deduce

$$\nabla g(x) + N_{\mathcal{M}}(x) = \hat{\partial}(g + \delta_{\mathcal{M}})(x) = \hat{\partial}(h + \delta_{\mathcal{M}})(x) \supset \hat{\partial} h(x),$$

and the result follows. ∎

Putting this together with the previous result, we arrive at the following proposition.

PROPOSITION 2.4 (smoothness and lineality). *Suppose the set $\mathcal{M} \subset X$ is a manifold around the point x. Suppose also that the function $h : X \to \overline{\mathbf{R}}$ has a subgradient at x and is regular there, and furthermore that the restriction $h|_{\mathcal{M}}$ is smooth around x. Then the subderivative $dh(x)$ is linear on the tangent space, or in other words*

$$\mathrm{lin}\, dh(x) \supset T_{\mathcal{M}}(x), \tag{2.5}$$

and the horizon subdifferential satisfies

$$\partial^{\infty} h(x) \subset N_{\mathcal{M}}(x). \tag{2.6}$$

Furthermore, the following properties are equivalent:

(i) *The lineality and tangent spaces coincide:*

$$\operatorname{lin} dh(x) = T_{\mathcal{M}}(x).$$

(ii) *The subdifferential and normal space are parallel:*

$$\operatorname{par} \partial h(x) = N_{\mathcal{M}}(x).$$

(iii) *h is "sharp" in normal directions at x, by which we mean*

$$dh(x)(-w) > -dh(x)(w) \quad \text{whenever } 0 \neq w \in N_{\mathcal{M}}(x).$$

Finally, if any of the above three properties hold, then $\nabla g(x) \in \operatorname{aff} \partial h(x)$ for any smooth representative g of $h|_{\mathcal{M}}$ and hence the following properties are equivalent:

(a) *x is a critical point of $h|_{\mathcal{M}}$;*

(b) *$0 \in \operatorname{aff} \partial h(x)$;*

(c) *$\operatorname{aff} \partial h(x) = N_{\mathcal{M}}(x)$.*

Proof. The first inclusion follows from Propositions 2.1 and 2.2, and the second (2.6) follows from the fact that $\partial^\infty h(x)$ is the recession cone of $\partial h(x)$. The equivalence of statements (i) and (ii) is also a consequence of Proposition 2.1. On the other hand, by Proposition 2.2, statement (ii) fails if and only if there exists a nonzero vector w in $N_{\mathcal{M}}(x)$ orthogonal to $\operatorname{par} \partial h(x)$, or in other words satisfying

$$\langle w, u - v \rangle = 0 \quad \text{for all } u, v \in \partial h(x),$$

and since we have

$$dh(x)(w) + dh(x)(-w) = \sup\{\langle w, u - v \rangle : u, v \in \partial h(x)\},$$

this is in turn equivalent to statement (iii) failing.

For the last statement, note that inclusion (2.3), regularity, and property (ii) imply $\nabla g(x) \in \operatorname{aff} \partial h(x)$, and hence $\operatorname{aff} \partial h(x) = \nabla g(x) + N_{\mathcal{M}}(x)$. This shows that properties (a) and (b) are equivalent, and the equivalence of properties (b) and (c) follows from property (ii). ∎

We are now ready for the key definition.

DEFINITION 2.7. *Suppose that the set $\mathcal{M} \subset X$ contains the point x. The function $h : X \to \overline{\mathbf{R}}$ is* partly smooth at x relative to $\mathcal{M}$ *if $\mathcal{M}$ is a manifold around x and the following four properties hold:*

(i) (restricted smoothness) *the restriction $h|_{\mathcal{M}}$ is smooth around x;*

(ii) (regularity) *at every point close to x in $\mathcal{M}$, the function h is regular and has a subgradient;*

(iii) (normal sharpness) *$dh(x)(-w) > -dh(x)(w)$ for all nonzero directions w in $N_{\mathcal{M}}(x)$;*

(iv) (subgradient continuity) *the subdifferential map ∂h is continuous at x relative to $\mathcal{M}$.*

We say h is partly smooth relative to *a set $\mathcal{M}$ if $\mathcal{M}$ is a manifold and h is partly smooth at each point in $\mathcal{M}$ relative to $\mathcal{M}$.*

DEFINITION 2.8 (partly smooth sets). *A set $S \subset X$ is* partly smooth at *a point* x relative to *a set $\mathcal{M}$ if δ_S is partly smooth at x relative to $\mathcal{M}$. We say S is* partly smooth relative to *a set $\mathcal{M}$ if $\mathcal{M}$ is a manifold and S is partly smooth at each point in $\mathcal{M}$ relative to $\mathcal{M}$.*

NOTE 2.9 (equivalent properties). Some comments may help with this rather lengthy definition.

(a) By Propositions 2.1, 2.2, and 2.4, we could replace property (iii) (normal sharpness) by either of the following properties:

(iii*) (*tangent linearity of subderivative*)

$$\mathrm{lin}\, dh(x) \subset T_{\mathcal{M}}(x)$$

(or indeed the corresponding equality);

(iii**) (*normals parallel to subdifferential*)

$$N_{\mathcal{M}}(x) \subset \mathrm{par}\, \partial h(x)$$

(or again the corresponding equality).

(b) Property (i) ensures that h is continuous relative to $\mathcal{M}$, so the subdifferential mapping is always outer semicontinuous relative to $\mathcal{M}$, by [23, Prop. 8.7]. Hence we could replace property (iv) by the following property:

(iv*) (*subgradient inner semicontinuity*) The subdifferential ∂h is *inner semicontinuous* at x relative to $\mathcal{M}$: in other words, for any sequence of points x_r in $\mathcal{M}$ approaching x and any subgradient $y \in \partial h(x)$, there exist subgradients $y_r \in \partial h(x_r)$ approaching y.

Notice that if h is locally Lipschitz (or "strictly continuous" in the terminology of [23]), then the subdifferential $\partial h(x)$ is everywhere nonempty and compact [23, Thm. 9.13], so by [23, Cor. 11.35] we could replace condition (iv) by the following condition:

(iv) (*subderivative continuity*) for all directions $w \in X$, the function $x \in \mathcal{M} \mapsto dh(x)(w)$ is continuous at x_0.

Furthermore, in this case the subderivative reduces to

$$dh(x)(w) = \liminf_{t \downarrow 0} \frac{h(x+tw) - h(x)}{t},$$

and regularity at x amounts to upper semicontinuity of the function $dh(\cdot)(w)$ at x for all directions w [23, Ex. 9.15 and Cor. 8.19]. This justifies the description of partial smoothness we gave in the introduction.

(c) Although the definition of partial smoothness is for a function h defined everywhere on the space X, it extends unchanged to a function defined only close to the point of interest, since partial smoothness depends only on properties of h near that point.

For a partly smooth function, the "normal sharpness" condition (iii), or equivalently, conditions (iii*) (tangent linearity of subderivative) and (iii**) (normals parallel to subdifferential), are all "stable": the fact that they hold at the point x_0 implies that they also hold at all nearby points in the active manifold. That is the content of the following result.

PROPOSITION 2.10 (local normal sharpness). *If the function $h : X \to \overline{\mathbf{R}}$ is partly smooth at the point x_0 relative to the set $\mathcal{M} \subset X$, then all points $x \in \mathcal{M}$ close to x_0 satisfy the condition*

$$dh(x)(-w) > -dh(x)(w) \quad \textit{for all } 0 \neq w \in N_{\mathcal{M}}(x),$$

or equivalently, the condition

$$N_{\mathcal{M}}(x) = \mathrm{par}\, \partial h(x).$$

Proof. The two properties are equivalent by Note 2.9. By Proposition 2.2 (subgradients and smoothness) we know $N_{\mathcal{M}}(x) \supset \operatorname{par} \partial h(x)$, so if the result fails, then there is a sequence of points $x_r \in \mathcal{M}$ approaching x_0 and a sequence of unit vectors $y_r \in N_{\mathcal{M}}(x_r)$ orthogonal to $\operatorname{par} \partial h(x_r)$. Taking a subsequence, we can suppose that y_r approaches a unit vector $y_0 \in N_{\mathcal{M}}(x_0)$.

Now for arbitrary subgradients $u_0, v_0 \in \partial h(x_0)$, by the continuity of ∂h there exist sequences $u_r \in \partial h(x_r)$ approaching u_0 and $v_r \in \partial h(x_r)$ approaching v_0, and they must satisfy $\langle y_r, u_r - v_r \rangle = 0$. Taking the limit shows $\langle y_0, u_0 - v_0 \rangle = 0$, so since u_0 and v_0 were arbitrary we deduce y_0 is orthogonal to $\operatorname{par} \partial h(x_0) = N_{\mathcal{M}}(x_0)$, which contradicts the fact that y_0 is a unit vector in $N_{\mathcal{M}}(x_0)$. □

We end this section with a simple characterization of partly smooth sets.

PROPOSITION 2.11 (partly smooth sets). *Suppose that the set $\mathcal{M} \subset X$ contains the point x_0. A set $S \subset X$ is partly smooth at x_0 relative to $\mathcal{M}$ if and only if $\mathcal{M}$ is a manifold around x_0 and the following four properties hold:*

(i) *$S \cap \mathcal{M}$ is a neighborhood of x_0 in $\mathcal{M}$;*
(ii) *S is Clarke regular at each point in $\mathcal{M}$ close to x_0;*
(iii) *$N_{\mathcal{M}}(x_0) \subset N_S(x_0) - N_S(x_0)$;*
(iv) *the normal cone map $N_S(\cdot)$ is continuous at x_0 relative to $\mathcal{M}$.*

Proof. This is an easy exercise using the facts that the set S is Clarke regular at the point $x \in S$ if and only if δ_S is regular there, and that $\partial \delta_S(x) = N_S(x)$, and then applying property (iii**) (normals parallel to subdifferential) in Note 2.9. □

The definition of partial smoothness looks a little involved at first sight, but we shall see that there are many important examples.

3. Basic examples. In this section we describe a few basic examples of partly smooth functions. In the next section we describe some calculus rules for building more complex examples.

Example 3.1 (smooth functions). If the open set $\Omega \subset X$ contains the point x and the function $h : \Omega \to \mathbf{R}$ is smooth, then h is partly smooth at x relative to Ω.

Example 3.2 (indicator functions). If $\mathcal{M} \subset X$ is a manifold around the point x, then $\mathcal{M}$ is a partly smooth set at x relative to $\mathcal{M}$. This is an easy consequence of Proposition 2.11 (partly smooth sets).

Example 3.3 (distance functions). If $\mathcal{M} \subset X$ is a manifold around the point x_0, then the distance function $d_{\mathcal{M}} : X \to \mathbf{R}$ defined by

$$d_{\mathcal{M}}(x) = \inf\{\|y - x\| : y \in \mathcal{M}\}$$

is partly smooth at x_0 relative to $\mathcal{M}$. To see this, notice that $\delta_{\mathcal{M}}|_{\mathcal{M}}$ is identically zero, which is smooth. By [23, Ex. 8.53] we know that $d_{\mathcal{M}}$ is regular at each point $x \in \mathcal{M}$ and

$$\partial h(x) = B \cap N_{\mathcal{M}}(x)$$

(where B denotes the closed unit ball in X). Thus the normal space is again parallel to the subdifferential, and this subdifferential varies continuously as x varies in $\mathcal{M}$. In fact, the Euclidean norm could be replaced by any other norm in this example, providing we replace B in the subdifferential formula above with the dual ball.

Notice in particular that the norm $\|\cdot\|$ is partly smooth at the origin relative to the origin.

Example 3.4 (polyhedral functions). Given any function $h : X \to \overline{\mathbf{R}}$ that is *polyhedral* (that is, its epigraph is a polyhedral set) and any point x_0 at which h is finite, there is a natural manifold about x_0 relative to which h is partly smooth. To see this we express h in the form (see [23, Thm. 2.49])

$$h(x) = \begin{cases} \max\{\langle a^i, x\rangle + b_i : i \in I\} & \text{if } \langle c^j, x\rangle \le d_j \text{ for all } j \in J, \\ +\infty & \text{otherwise} \end{cases}$$

for some finite index sets $I \neq \emptyset$ and J and given vectors a^i and c^j in X and reals b_i and d_j (for $i \in I$ and $j \in J$). For any point $x \in X$, define "active" index sets

$$I(x) = \{i \in I : \langle a^i, x\rangle + b_i = h(x)\},$$
$$J(x) = \{j \in J : \langle c^j, x\rangle = d_j\}.$$

Define the set

$$\mathcal{M}_{x_0} = \{x \in X : I(x) = I_0 \text{ and } J(x) = J_0\},$$

where $I_0 = I(x_0)$ and $J_0 = J(x_0)$. It is easy to see that $\mathcal{M}_{x_0}$ is a manifold around x_0. We claim that h is partly smooth at x_0 relative to $\mathcal{M}_{x_0}$.

To see this observe first that for any index $i \in I_0$ we have

$$h(x) = \langle a^i, x\rangle + b_i \quad \text{for all } x \in \mathcal{M}_{x_0},$$

so $h|_{\mathcal{M}_{x_0}}$ is smooth. Second, h is lower semicontinuous and convex, and hence regular whenever it is finite [23, Ex. 7.27]. Now routine calculation (using [23, Thm. 6.46], for example) shows that at any point $x \in \mathcal{M}_{x_0}$ we have

$$N_{\mathcal{M}_{x_0}}(x) = \left\{ \sum_{i\in I_0} \lambda_i a^i + \sum_{j \in J_0} \mu_j c^j : \sum_{i \in I_0} \lambda_i = 0 \right\},$$

$$\partial h(x) = \left\{ \sum_{i\in I_0} \lambda_i a^i + \sum_{j \in J_0} \mu_j c^j : \sum_{i \in I_0} \lambda_i = 1, \quad \lambda_i \ge 0 \quad (i \in I_0), \right.$$
$$\left. \mu_j \ge 0 \quad (j \in J_0) \right\}.$$

Thus the normal space is parallel to the subdifferential, which is constant on $\mathcal{M}_{x_0}$.

In particular, the *basic max function* $\mathrm{mx} : \mathbf{R}^n \to \mathbf{R}$ defined by $\mathrm{mx}\, x = \max_i x_i$ is partly smooth at any point $x_0 \in \mathbf{R}^n$ relative to the set

$$\mathcal{M}_{x_0} = \{x \in \mathbf{R}^n : I(x) = I(x_0)\}, \tag{3.5}$$

where

$$I(x) = \left\{ j : x_j = \max_i x_i \right\}.$$

Example 3.6 (largest eigenvalue). The Euclidean space $\mathbf{S}^n$ consists of the n-by-n real symmetric matrices with the inner product $\langle x, y\rangle = \mathrm{trace}(xy)$, for $x, y \in \mathbf{S}^n$. The functions $\lambda_1(x) \ge \lambda_2(x) \ge \cdots \ge \lambda_n(x)$ denote the eigenvalues of x (listed in decreasing order by multiplicity). Then the largest eigenvalue is partly smooth relative to the manifold

$$\mathcal{M}_m = \{x \in \mathbf{S}^n : \lambda_1(x) \text{ has multiplicity } m\} \quad (1 \le m \le n).$$

To see this, note first that the set $\mathcal{M}_m$ above is indeed a manifold (see [18], for example). Furthermore we can write the maximum eigenvalue as

$$\lambda_1(x) = m^{-1} \sum_{j=1}^{m} \lambda_j(x) \quad \text{for all } x \in \mathcal{M}_m,$$

and the right-hand side is a smooth function of x on $\mathcal{M}_m$ (see [11], for example). Second, λ_1 is convex (see [10], for example) and so is regular everywhere. Now, by [18], as x varies in $\mathcal{M}_m$ there is an n-by-m matrix $Q(x)$, depending continuously on x, whose columns are a basis for the eigenspace of x corresponding to $\lambda_1(x)$, and then we have

$$N_{\mathcal{M}_m}(x) = Q(x)\{w \in \mathbf{S}^n : \operatorname{trace} w = 0\}Q(x)^T,$$
$$\partial\lambda_1(x) = Q(x)\{w \in \mathbf{S}^n_+ : \operatorname{trace} w = 1\}Q(x)^T,$$

where $\mathbf{S}^n_+$ denotes the positive semidefinite matrices [18, Thm. 4.7]. It is easy to see from this that the normal space is parallel to the subdifferential, which varies continuously on $\mathcal{M}_m$.

Example 3.7 (spectral abscissa). The Euclidean space $\mathbf{M}^n$ consists of the n-by-n complex matrices with the (real) inner product $\langle x, y\rangle = \operatorname{Re}\operatorname{trace}(x^* y)$ for $x, y \in \mathbf{M}^n$. The *spectral abscissa* $\alpha(x)$ is the largest of the real parts of the eigenvalues of x.

Given any list $\phi = (n_1, n_2, \ldots, n_r)$ of positive integers with sum no greater than n, let $\mathcal{M}_\phi$ denote the subset of $\mathbf{M}^n$ consisting of matrices x satisfying the following properties:

(i) x has r distinct "active" eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_r$ with real part $\alpha(x)$, and all its other eigenvalues have real part strictly less than $\alpha(x)$;

(ii) each active eigenvalue λ_j has algebraic multiplicity n_j and geometric multiplicity one.

Classic results of Arnold [2] show that $\mathcal{M}_\phi$ is a manifold.

In fact the spectral abscissa α is partly smooth relative to $\mathcal{M}_\phi$ (see [5]).

4. Calculus. In this section we show that partly smooth functions form a robust class by proving a variety of calculus rules. Our fundamental result considers the composition of a partly smooth function with a smooth function, and requires a transversality condition. Consider Euclidean spaces X and Z, an open set $W \subset Z$ containing a point z, a smooth map $\Phi : W \to X$, and a set $\mathcal{M} \subset X$. We say Φ is *transversal to* $\mathcal{M}$ *at* z if $\mathcal{M}$ is a manifold around $\Phi(z)$, and

$$R(\nabla\Phi(z)) + T_{\mathcal{M}}(\Phi(z)) = X$$

or equivalently

$$\operatorname{Ker}(\nabla\Phi(z)^*) \cap N_{\mathcal{M}}(\Phi(z)) = \{0\}. \tag{4.1}$$

THEOREM 4.2 (chain rule). *Given Euclidean spaces X and Z, an open set $W \subset Z$ containing a point z_0, a smooth map $\Phi : W \to X$, and a set $\mathcal{M} \subset X$, suppose Φ is* *transversal to $\mathcal{M}$ at z_0. If the function $h : X \to \overline{\mathbf{R}}$ is partly smooth at $\Phi(z_0)$ relative to $\mathcal{M}$, then the composition $h \circ \Phi$ is partly smooth at z_0 relative to $\Phi^{-1}(\mathcal{M})$.*

Proof. An immediate consequence of transversality is that the set $\Phi^{-1}(\mathcal{M})$ is a manifold around any point $z \in \Phi^{-1}(\mathcal{M})$ close to z_0, with normal space

$$N_{\Phi^{-1}(\mathcal{M})}(z) = \nabla\Phi(z)^* N_{\mathcal{M}}(\Phi(z)),$$

and transversality also holds at all such z.

Given a smooth representative g of $h|_{\mathcal{M}}$ around $\Phi(z_0)$, it is easy to see that $g \circ \Phi$ is a smooth representative of $(h \circ \Phi)|_{\Phi^{-1}(\mathcal{M})}$ around z_0, so this latter function is smooth around z_0.

Consider any point $z \in \Phi^{-1}(\mathcal{M})$ close to z_0. By inclusion (2.6) we know

$$\partial^\infty h(\Phi(z)) \subset N_{\mathcal{M}}(\Phi(z)). \tag{4.3}$$

Transversality at z therefore implies

$$\mathrm{Ker}(\nabla\Phi(z)^*) \cap \partial^\infty h(\Phi(z)) = \{0\},$$

so by [23, Thm. 10.6], $h \circ \Phi$ is regular at z, with subdifferential

$$\partial(h \circ \Phi)(z) = \nabla\Phi(z)^* \partial h(\Phi(z)) \neq \emptyset. \tag{4.4}$$

Now, the normal space is parallel to the subdifferential, since

$$\begin{aligned} &\mathrm{par}\,(\partial(h \circ \Phi)(z_0)) = \mathrm{par}\,(\nabla\Phi(z_0)^* \partial h(\Phi(z_0))) \\ &= \nabla\Phi(z_0)^* \mathrm{par}\,(\partial h(\Phi(z_0))) \supset \nabla\Phi(z_0)^* N_{\mathcal{M}}(\Phi(z_0)) = N_{\Phi^{-1}(\mathcal{M})}(z_0), \end{aligned}$$

so it remains only to check the inner semicontinuity property of the subdifferential.

Consider therefore a convergent sequence of points $z_r \to z_0$ in $\Phi^{-1}(\mathcal{M})$, and a subgradient $w \in \partial(h \circ \Phi)(z_0)$. By (4.4) there is a subgradient $y \in \partial h(\Phi(z_0))$ such that $\nabla\Phi(z_0)^* y = w$. Since $\Phi(z_r) \to \Phi(z_0)$ in $\mathcal{M}$ and ∂h is continuous on $\mathcal{M}$ at $\Phi(z_0)$, there must be subgradients $y_r \in \partial h(\Phi(z_r))$ approaching y. But Φ is smooth, so the vectors $\nabla\Phi(z_r)^* y_r \in \partial(h \circ \Phi)(z_r)$ approach w, as required. □

For example, suppose $\Phi(z_0) = 0$ and $\nabla\Phi(z_0)$ is surjective. Then the function $z \mapsto \|\Phi(z)\|$ is partly smooth at z_0 relative to $\Phi^{-1}(0)$.

By applying this result with $h = \delta_S$, we obtain conditions guaranteeing that the set $\Phi^{-1}(S)$ is partly smooth if the set S is smooth.

PROPOSITION 4.5 (separability). *For each $i = 1, 2, \ldots, k$, suppose that X_i is a Euclidean space, that the set $\mathcal{M}_i \subset X_i$ contains the point x_i^0, and that the function $h_i : X_i \to \overline{\mathbf{R}}$ is partly smooth at x_i^0 relative to $\mathcal{M}_i$. Then the function $h : X_1 \times X_2 \times \cdots \times X_k \to \overline{\mathbf{R}}$ defined by*

$$h(x_1, x_2, \ldots, x_k) = \sum_{i=1}^{k} h_i(x_i) \quad \text{for } x_i \in X_i, \quad i = 1, 2, \ldots, k,$$

is partly smooth at $(x_1^0, x_2^0, \ldots, x_k^0)$ relative to $\mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_k$.

Proof. This follows easily from the facts that $\mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_k$ is a manifold around $(x_1^0, x_2^0, \ldots, x_k^0)$, with normal space

$$N_{\mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_k}(x_1, x_2, \ldots, x_k) = N_{\mathcal{M}_1}(x_1) \times N_{\mathcal{M}_2}(x_2) \times \cdots \times N_{\mathcal{M}_k}(x_k),$$

and

$$\partial h(x_1, x_2, \ldots, x_k) = \partial h_1(x_1) \times \partial h_2(x_2) \times \cdots \times \partial h_k(x_k),$$

with regularity providing each h_i is regular at x_i [23, Prop. 10.5]. ∎

For example, the function

$$(x_1, x_2, \ldots, x_k) \mapsto \|x_1\| + \|x_2\| + \cdots + \|x_k\|$$

is partly smooth at the origin relative to the origin.

Applying this result to indicator functions shows that direct products of partly smooth sets are partly smooth.

COROLLARY 4.6 (sum rule). *Consider sets $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_k$ in a Euclidean space Z. Suppose the function $h_i : Z \to \overline{\mathbf{R}}$ is partly smooth at the point z_0 relative to $\mathcal{M}_i$ for each i. Assume furthermore the condition*

$$\sum_{i=1}^{k} y_i = 0 \quad \text{and} \quad y_i \in N_{\mathcal{M}_i}(z_0) \quad \text{for each } i \quad \Rightarrow \quad y_i = 0 \quad \text{for each } i.$$

Then the function $\sum_i h_i$ is partly smooth at z_0 relative to $\cap_i \mathcal{M}_i$.

Proof. We apply the chain rule (Theorem 4.2) and Proposition 4.5 (separability) with

$$\begin{aligned} X &= Z \times Z \times \cdots \times Z \quad (k \text{ copies}), \\ W &= Z, \\ \Phi(z) &= (z, z, \ldots, z) \ \text{ for } z \in Z, \\ \mathcal{M} &= \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_k, \\ h(z_1, z_2, \ldots, z_k) &= \sum_i h_i(z_i) \ \text{ for } z_i \in Z,\ i = 1, 2, \ldots, k. \end{aligned}$$

∎

Applying this result to indicator functions gives conditions guaranteeing that intersections of partly smooth sets are partly smooth.

COROLLARY 4.7 (smooth perturbation). *If the function $h : X \to \overline{\mathbf{R}}$ is partly smooth at the point x_0 relative to the set $\mathcal{M} \subset X$ and the function $f : X \to \overline{\mathbf{R}}$ is smooth on an open set containing x_0, then the function $h + f$ is partly smooth at x_0 relative to $\mathcal{M}$.*

COROLLARY 4.8 (smooth max function). *Suppose W is an open subset of the Euclidean space Z, and the function $\Phi : W \to \mathbf{R}^n$ is smooth. For any point $z \in W$, define the "active set"*

$$J(z) = \left\{ i : \Phi_i(z) = \max_j \Phi_j(z) \right\}.$$

If the point $z_0 \in W$ satisfies

$$\{\nabla \Phi_i(z_0) : i \in J(z_0)\} \ \text{ linearly independent}, \tag{4.9}$$

then the function $h : W \to \mathbf{R}$ defined by $h(z) = \max_j \Phi_j(z)$ is partly smooth at z_0 relative to the set

$$\mathcal{M} = \{z \in W : J(z) = J(z_0)\}.$$

Proof. We apply the chain rule (Theorem 4.2) with $X = \mathbf{R}^n$, $\mathcal{M} = \mathcal{M}_{\Phi(z_0)}$ as in (3.5), and $h = \text{mx}$, the basic max function of Example 3.4. The transversality condition follows easily from condition (4.9). ∎

To apply the idea of partial smoothness to optimization problems with constraints, we need conditions to recognize partly smooth level sets. That is the aim of the last result of this section.

THEOREM 4.10 (level sets). *Consider a point x_0 in a set $\mathcal{M} \subset X$. Suppose that the function $h : X \to \overline{\mathbf{R}}$ is partly smooth at x_0 relative to $\mathcal{M}$, and that x_0 is not a critical point of $h|_{\mathcal{M}}$. Then the level set*

$$L = \{x \in \mathcal{M} : h(x) \leq 0\}$$

is partly smooth at x_0 relative to the set

$$\mathcal{M}_0 = \{x \in \mathcal{M} : h(x) = 0\}.$$

Proof. We can choose an open neighborhood V of x_0, and smooth functions $g : V \to \mathbf{R}$ and $F : V \to \mathbf{R}^m$, such that g agrees with h on the set

$$\mathcal{M} \cap V = \{x \in V : F(x) = 0\}$$

and F has surjective derivative throughout V. If we choose a sufficiently small neighborhood V, then the set

$$\{x \in \mathcal{M} \cap V : h(x) = 0\} \ = \ \{x \in V : F(x) = 0 \text{ and } g(x) = 0\} \tag{4.11}$$

is a manifold around x_0 since

$$\nabla g(x_0) \notin N_{\mathcal{M}}(x_0) = R(\nabla F(x_0)^*).$$

Thus $\mathcal{M}_0$ is indeed a manifold around x_0.

We now need to check the four conditions of Proposition 2.11 (partly smooth sets). Clearly property (i) holds, since $\mathcal{M}_0 \subset \mathcal{M}$.

The assumption that x_0 is not a critical point of $h|_{\mathcal{M}}$ is equivalent to $0 \notin \operatorname{aff} \partial(x_0)$, by Proposition 2.4, so in particular we know $0 \notin \partial h(x_0)$. Since the subdifferential mapping ∂h is continuous relative to $\mathcal{M}$, it follows that $0 \notin \partial h(x)$ for all points $x \in \mathcal{M}$ close to x_0. (In fact this follows just from outer semicontinuity.)

Now consider a point $x \in \mathcal{M}_0$ close to x_0. Notice that h is regular at x_0 and thus locally lower semicontinuous. We can apply [23, Prop. 10.3] to deduce that the level set L is Clarke regular at x (which proves property (ii)), and

$$N_L(x) = (\mathbf{R}_+ \partial h(x)) \cup \partial^\infty h(x).$$

Notice that the right-hand side is closed (since the normal cone is always closed), and it contains $\mathbf{R}_+ \partial h(x)$ and hence also $\operatorname{cl} \mathbf{R}_+ \partial h(x)$. On the other hand, by regularity we have

$$\partial^\infty h(x) = \partial h(x)^\infty \subset \operatorname{cl} \mathbf{R}_+ \partial h(x).$$

Putting these observations together, we deduce the representation

$$N_L(x) = \operatorname{cl} \mathbf{R}_+ \partial h(x). \tag{4.12}$$

By (4.11) we have

$$N_{\mathcal{M}_0}(x_0) = N_{\mathcal{M}}(x_0) + \mathbf{R}\nabla g(x_0),$$

and since h is partly smooth at x_0 relative to $\mathcal{M}$ we also know

$$N_{\mathcal{M}}(x_0) = \mathbf{R}_+(\partial h(x_0) - \partial h(x_0)).$$

Furthermore, Proposition 2.4 implies

$$\nabla g(x_0) \in \operatorname{aff} \partial h(x_0) = \partial h(x_0) + \mathbf{R}_+(\partial h(x_0) - \partial h(x_0)).$$

Hence certainly we have

$$N_{\mathcal{M}_0}(x_0) \subset \mathbf{R}_+\partial h(x_0) - \mathbf{R}_+\partial h(x_0) \subset N_L(x_0) - N_L(x_0),$$

which proves property (iii).

It remains to prove that the normal cone mapping $N_L(\cdot)$ is inner semicontinuous at x_0 relative to $\mathcal{M}_0$, or in other words

$$N_L(x_0) \subset \liminf_{x \to x_0,\ x \in \mathcal{M}_0} N_L(x).$$

Using (4.12) we can rewrite this as

$$\operatorname{cl} \mathbf{R}_+\partial h(x_0) \subset \liminf_{x \to x_0,\ x \in \mathcal{M}_0} \operatorname{cl} \mathbf{R}_+\partial h(x).$$

Since the lim inf is always closed, it suffices to prove

$$\mathbf{R}_+\partial h(x_0) \subset \liminf_{x \to x_0,\ x \in \mathcal{M}_0} \mathbf{R}_+\partial h(x).$$

To this end, suppose that the sequence of points $x_r \in \mathcal{M}_0$ converges to x_0, and consider a vector $y = \mu z$ for some real $\mu \geq 0$ and subgradient $z \in \partial h(x_0)$. Since the subdifferential map ∂h is continuous at x_0 relative to $\mathcal{M}$, there exist subgradients $z_r \in \partial h(x_r)$ approaching z, and then we have vectors $\mu z_r \in \mathbf{R}_+\partial h(x_r)$ approaching y as required. ∎

Corollary 4.13 (smooth constraints). *Suppose W is an open subset of the Euclidean space Z, and the function $\Phi : W \to \mathbf{R}^n$ is smooth. For any point z in the set*

$$L = \{z \in W : \Phi(z) \leq 0\},$$

define the "active set"

$$K(z) = \{k : \Phi_k(z) = 0\}.$$

If the point $z_0 \in L$ satisfies the condition

$$\{\nabla\Phi_k(z_0) : k \in K(z_0)\} \quad \textit{linearly independent},$$

then the set L is partly smooth at z_0 relative to the set

$$\{z \in W : K(z) = K(z_0)\}.$$

Proof. We apply Theorem 4.10 (level sets) to the smooth max function h defined in Corollary 4.8. Notice

$$\partial h(z_0) = \text{conv}\,\{\nabla \Phi_k(z_0) : k \in K(z_0)\},$$

so $0 \notin \text{aff}\,\partial h(z_0)$ by the linear independence assumption, and hence z_0 is not a critical point of $h|_{\mathcal{M}}$ for the set $\mathcal{M}$ defined in Corollary 4.8. ◻

Example 4.14 (semidefinite cone). The convex cone $\mathbf{S}^n_-$ of negative semidefinite matrices is partly smooth relative to the manifold

$$\{x \in \mathbf{S}^n_- : \text{rank}\, x = k\}$$

for any integer $k = 0, 1, \ldots, n$. To see this, we simply apply Theorem 4.10 to the largest eigenvalue.

Example 4.15 (semistable matrices). A matrix $x \in \mathbf{M}^n$ is *semistable* if all its eigenvalues lie in the closed left half-plane, or in other words, with the notation of Example 3.7 (spectral abscissa), if $\alpha(x) \leq 0$. The (nonconvex) cone of semistable matrices is partly smooth relative to the manifold

$$\{x \in \mathcal{M}_\phi : \alpha(x) = 0\}$$

for any list of multiplicities ϕ. To see this, we apply Theorem 4.10 to the spectral abscissa, using the fact that any subgradient of the spectral abscissa at any point has trace one.

5. Sensitivity. This section considers the stability of critical points of parametric partly smooth functions. Throughout this section we make the following assumption.

Assumption 5.1 (transversal embedding). For Euclidean spaces Y and Z, the set $\mathcal{Q} \subset Y \times Z$ is a manifold containing the point (y_0, z_0) and satisfies the condition

$$(w, 0) \in N_{\mathcal{Q}}(y_0, z_0) \quad \Rightarrow \quad w = 0.$$

Notice that this assumption is "stable": if it holds at the point (y_0, z_0), then it also holds at all nearby points in $\mathcal{Q}$.

For each vector $y \in Y$ we define the set

$$\mathcal{Q}_y = \{z \in Z : (y, z) \in \mathcal{Q}\}.$$

Since the condition in Assumption 5.1 is exactly the transversality condition (4.1) for the map $\Phi : Z \to Y \times Z$ defined by $\Phi(z) = (y_0, z)$, the set $\mathcal{Q}_{y_0}$ is a manifold around z_0. In fact the following result, whose proof is immediate, shows that rather more is true: providing y is close to y_0, the set $\mathcal{Q}_y$ has the structure of a manifold close to z_0.

PROPOSITION 5.2. *If Assumption* 5.1 *holds, then there is an open neighborhood* U *of* z_0 *such that for all vectors* $y \in Y$ *close to* y_0 *the set* $\mathcal{Q}_y \cap U$ *is a manifold.*

Throughout this section we consider a function $p : Y \times Z \to \overline{\mathbf{R}}$, and we define a function $p_y : Z \to \overline{\mathbf{R}}$ by

$$p_y(z) = p(y, z) \quad \text{for } y \in Y \text{ and } z \in Z.$$

Clearly if the restriction $p|_{\mathcal{Q}}$ is smooth, then so is the restriction $p_y|_{\mathcal{Q}_y}$. The next result shows an analogous property for partial smoothness.

PROPOSITION 5.3 (partial smoothness with parameters). *Suppose Assumption* 5.1 *holds and the function* p *is partly smooth relative to the manifold* $\mathcal{Q}$. *Then there is an open neighborhood* U *of the point* z_0 *such that the function* p_y *is partly smooth relative to* $\mathcal{Q}_y \cap U$ *for all vectors* $y \in Y$ *close to* y_0.

Proof. There are open neighborhoods U of z_0 and V of y_0 such that $\mathcal{Q}_y \cap U$ is a manifold for all $y \in V$ and

$$y \in V,\ z \in U,\ (y,z) \in \mathcal{Q},\ (w,0) \in N_{\mathcal{Q}}(y,z) \quad \Rightarrow \quad w = 0.$$

Hence for any points $\hat{y} \in V$ and $\hat{z} \in \mathcal{Q}_{\hat{y}} \cap U$ we can apply the chain rule (Theorem 4.2) at $\hat{z}$ with the map $\Phi : Z \to Y \times Z$ defined by $\Phi(z) = (\hat{y}, z)$ to deduce that the function $p_{\hat{y}} = p \circ \Phi$ is partly smooth at $\hat{z}$ relative to the manifold $\mathcal{Q}_{\hat{y}} \cap U$. ☐

Our main aim in this work is to study sensitivity of critical points for partly smooth functions. Just as in classical sensitivity analysis for nonlinear programming, we need second-order conditions to make progress.

DEFINITION 5.4. *Given any subset* $\mathcal{M}$ *of a Euclidean space* X, *a point* x_0 *is a* strong local minimizer *of a function* $f : \mathcal{M} \to \overline{\mathbf{R}}$ *if there exists a real* $\delta > 0$ *such that* $f(x) \geq f(x_0) + \delta \|x - x_0\|^2$ *for all* $x \in \mathcal{M}$ *near* x_0.

We recall some classical sensitivity analysis (see, for example, [8]). Suppose $\mathcal{M} \subset X$ is a manifold around the point $x_0 \in \mathcal{M}$, and the restriction $h|_{\mathcal{M}}$ is smooth around x_0, for some function $h : X \to \overline{\mathbf{R}}$. Let g be any smooth representative of $h|_{\mathcal{M}}$. By definition, x_0 has an open neighborhood $V \subset X$ such that

$$\mathcal{M} \cap V = \{x \in V : F(x) = 0\}$$

for some smooth function $F : V \to \mathbf{R}^m$ with $\nabla F(x_0)$ surjective. The point x_0 is a critical point of $h|_{\mathcal{M}}$ if and only if $\nabla g(x_0) \in N_{\mathcal{M}}(x_0)$, which is equivalent to the existence of a multiplier vector $\mu \in \mathbf{R}^m$ (necessarily unique) such that x_0 is a critical point of the corresponding Lagrangian function $L = g + \mu^T F$. Furthermore, x_0 is a strong local minimizer of $h|_{\mathcal{M}}$ if and only if it is a critical point of $h|_{\mathcal{M}}$ and satisfies the second-order condition

$$y^T \nabla^2 L(x_0) y > 0 \quad \text{whenever } 0 \neq y \in \mathrm{Ker}(\nabla F(x_0)).$$

The following result is also classical.

THEOREM 5.5 (parametric strong minimizers). *Suppose that the function* $p|_{\mathcal{Q}}$ *is smooth around the point* (y_0, z_0), *that the point* z_0 *is a strong local minimizer of the function* $p_{y_0}|_{\mathcal{Q}_{y_0}}$, *and that Assumption* 5.1 *holds. Then there are open neighborhoods* $U \subset Z$ *of* z_0 *and* $V \subset Y$ *of* y_0 *and a continuously differentiable function* $\Psi : V \to U$ *such that* $\Psi(y_0) = z_0$ *and for all vectors* $y \in V$ *the function* $p_y|_{\mathcal{Q}_y \cap U}$ *has a unique critical point* $\Psi(y)$, *which is furthermore a strong local minimizer.*

To approach a more complete sensitivity theory, we combine the smooth analysis of a partly smooth function on its active manifold with a study of its behavior in normal directions. That is the idea of the following definition.

DEFINITION 5.6 (strong critical point). *For a Euclidean space* X, *suppose the function* $h : X \to \overline{\mathbf{R}}$ *is partly smooth at the point* x_0 *relative to the set* $\mathcal{M} \subset X$. *We call* x_0 *a* strong critical point *of* h *relative to* $\mathcal{M}$ *if*

(i) x_0 *is a strong local minimizer of* $h|_{\mathcal{M}}$, *and*

(ii) $0 \in \mathrm{ri}\, \partial h(x_0)$.

In the next section we see that the condition $0 \in \mathrm{ri}\, \partial h(x_0)$ could be written equivalently as x_0 being a "sharp" local minimizer of the function $h|_{x_0 + N_{\mathcal{M}}(x_0)}$.

We are now ready for the main result. Comparing it with the classical result Theorem 5.5 above, we see that the extra assumption of strong criticality implies that the parametrized minimizer is also a strong critical point.

THEOREM 5.7 (strong critical points and parameters). *Suppose Assumption* 5.1 *holds and the function p is partly smooth relative to the manifold $\mathcal{Q}$. If the point z_0 is a strong critical point of the function p_{y_0} relative to the set $\mathcal{Q}_{y_0}$, then there are open neighborhoods $U \subset Z$ of z_0 and $V \subset Y$ of y_0 and a continuously differentiable function $\Psi : V \to U$ satisfying $\Psi(y_0) = z_0$ and with the following properties for all vectors $y \in V$:*

(i) *the function $p_y|_{\mathcal{Q}_y \cap U}$ has a unique critical point $\Psi(y)$;*

(ii) *$\Psi(y)$ is a strong critical point of the function p_y relative to the manifold $\mathcal{Q}_y \cap U$.*

Proof. Theorem 5.5 shows the existence of a function Ψ having the required properties, with the exception of property (ii). Proposition 5.3 shows that p_y is partly smooth relative to the manifold $\mathcal{Q}_y \cap U$. Hence to prove property (ii), it suffices to show

$$0 \in \operatorname{ri} \partial p_y(\Psi(y)) \quad \text{for } y \in V \text{ close to } y_0.$$

To this end, as in the proof of Proposition 5.3, we define a map $\Phi_y : Z \to Y \times Z$ by $\Phi_y(z) = (y, z)$ for $z \in Z$, observe that $p_y = p \circ \Phi_y$, and note that Assumption 5.1 allows us to apply the chain rule (Theorem 4.2). By (4.4) we deduce

$$\partial p_y(\Psi(y)) = \operatorname{proj}_Z \partial p(y, \Psi(y)),$$

where $\operatorname{proj}_Z : Y \times Z \to Z$ is the natural projection, whereas a standard calculation shows

$$N_{\mathcal{Q}_y}(\Psi(y)) = \operatorname{proj}_Z N_{\mathcal{Q}}(y, \Psi(y)).$$

We therefore know

$$0 \in \operatorname{ri} (\operatorname{proj}_Z \partial p(y_0, \Psi(y_0))), \tag{5.8}$$

and we want to deduce

$$0 \in \operatorname{ri} (\operatorname{proj}_Z \partial p(y, \Psi(y))) \quad \text{for all } y \text{ close to } y_0.$$

Notice that, by definition, we know $\Psi(y)$ is a critical point of the restriction $p_y|_{\mathcal{Q}_y}$ for y close to y_0, so by partial smoothness, Proposition 2.10 (local normal sharpness) and Proposition 2.4 (smoothness and lineality) we have

$$\operatorname{aff} (\operatorname{proj}_Z \partial p(y, \Psi(y))) = \operatorname{aff} \partial p_y(\Psi(y)) = N_{\mathcal{Q}_y}(\Psi(y)) = \operatorname{proj}_Z N_{\mathcal{Q}}(y, \Psi(y)).$$

If the result fails, then there is a sequence of vectors y_r in Y approaching y_0 such that

$$0 \notin \operatorname{ri} (\operatorname{proj}_Z \partial p(y_r, \Psi(y_r))) \quad \text{for all } r.$$

For all large r we can separate in the subspace $\operatorname{proj}_Z N_{\mathcal{Q}}(y_r, \Psi(y_r))$ to deduce the existence of a unit vector z_r in this subspace, satisfying

$$\inf \langle z_r, \operatorname{proj}_Z \partial p(y_r, \Psi(y_r)) \rangle \geq 0.$$

After taking a subsequence, we can assume z_r approaches a nonzero vector $z \in Z$.

Now, since the point $(y_r, \Psi(y_r))$ converges to the point (y_0, z_0) in the manifold $\mathcal{Q}$, it follows that the subspace $N_{\mathcal{Q}}(y_r, \Psi(y_r))$ converges to the subspace $N_{\mathcal{Q}}(y_0, z_0)$, so Assumption 5.1 implies that the subspace $\mathrm{proj}_Z N_{\mathcal{Q}}(y_r, \Psi(y_r))$ converges to the subspace $\mathrm{proj}_Z N_{\mathcal{Q}}(y_0, z_0)$, by [23, Ex. 4.28]. Hence we deduce

$$z \in \mathrm{proj}_Z N_{\mathcal{Q}}(y_0, z_0).$$

We now claim

$$\inf \langle z, \mathrm{proj}_Z \partial p(y_0, z_0) \rangle \geq 0. \tag{5.9}$$

To see this, consider any vector $u \in \mathrm{proj}_Z \partial p(y_0, z_0)$. Partial smoothness implies that $\partial p(y_r, \Psi(y_r))$ converges to $\partial p(y_0, z_0)$, so again by [23, Ex. 4.28] we deduce that $\mathrm{proj}_Z \partial p(y_r, \Psi(y_r))$ converges to $\mathrm{proj}_Z \partial p(y_0, z_0)$. Thus there is a sequence of vectors $u_r \in \mathrm{proj}_Z \partial p(y_r, \Psi(y_r))$ converging to u. Since $\langle z_r, u_r \rangle \geq 0$ for all r, we deduce $\langle z, u \rangle \geq 0$, as we claimed.

Thus inequality (5.9) holds, so the origin is separated from the convex set $\mathrm{proj}_Z \partial p(y_0, z_0)$ in its affine span (the subspace $\mathrm{proj}_Z N_{\mathcal{Q}}(y_0, z_0)$). But this contradicts relation (5.8), so the proof is complete. □

6. $\mathcal{U}-\mathcal{V}$ decomposition and identifiable surfaces. As we remarked in the introduction, our development is closely related to the $\mathcal{U}$-Lagrangian theory for convex functions of Lemaréchal, Oustry, and Sagastizábal (see, for example, [12]). The key idea of that theory is, for a given *convex* function $h : X \to \overline{\mathbf{R}}$, to decompose X as a sum of two orthogonal subspaces, $\mathcal{U}$ and $\mathcal{V}$: h behaves "sharply" at the point of interest if we perturb in directions in the $\mathcal{V}$ space, whereas it behaves smoothly if we perturb in directions in the $\mathcal{U}$ space.

Our purpose in this section is to draw the connection between this idea and partial smoothness. The development is a nice illustration of various features of the theory of partial smoothness.

We call a local minimizer x of an arbitrary function $h : X \to \overline{\mathbf{R}}$ *sharp* if

$$\liminf_{z \to 0} \frac{h(x+z) - h(x)}{\|z\|} > 0,$$

or equivalently, if $0 \in \mathrm{int}\, \hat{\partial} h(x)$.

THEOREM 6.1 ($\mathcal{U}-\mathcal{V}$ decomposition). *Suppose the function $h : X \to \overline{\mathbf{R}}$ is partly smooth at the point x relative to the set $\mathcal{M} \subset X$. Define subspaces $\mathcal{U} = T_{\mathcal{M}}(x)$ and $\mathcal{V} = N_{\mathcal{M}}(x)$. Then there exists a function $v : \mathcal{U} \to \mathcal{V}$ with the following three properties:*

(i) *the function v is smooth near the origin;*

(ii) *for small vectors $u \in \mathcal{U}$ and $w \in \mathcal{V}$, $x + u + w \in \mathcal{M} \Leftrightarrow w = v(u)$;*

(iii) *$v(u) = O(\|u\|^2)$ for small $u \in \mathcal{U}$.*

Fix any vector $y \in \mathrm{ri}\, \partial h(x)$. Then for any small vector $u \in \mathcal{U}$, the function

$$w \in \mathcal{V} \mapsto h(x+u+w) - \langle y, x+u+w \rangle \tag{6.2}$$

has a sharp minimizer at the point $v(u)$.

Furthermore, the point x is a strong critical point of h relative to $\mathcal{M}$ if and only if it is a strong local minimizer of $h|_{\mathcal{M}}$ and a sharp local minimizer of $h|_{x+\mathcal{V}}$.

Proof. By first intersecting with an open set, we can assume $\mathcal{M}$ is a manifold. A standard argument using the implicit function theorem shows the existence of the function v with the properties (i), (ii), and (iii).

Define a map $\Phi : \mathcal{U} \times \mathcal{V} \to X$ by

$$\Phi(u, w) = x + u + w \quad \text{for } u \in \mathcal{U} \text{ and } w \in \mathcal{V}.$$

Clearly Φ is everywhere transversal to $\mathcal{M}$. Hence by the chain rule (Theorem 4.2), the function $h \circ \Phi$ is partly smooth relative to the manifold $\Phi^{-1}(\mathcal{M})$. Consequently, by smooth perturbation (Corollary 4.7) the function $p : \mathcal{U} \times \mathcal{V} \to \overline{\mathbf{R}}$ defined by

$$p(u, w) = h(x + u + w) - \langle y, x + u + w \rangle \quad \text{for } u \in \mathcal{U} \text{ and } w \in \mathcal{V}$$

is partly smooth relative to the manifold

$$\mathcal{Q} = \Phi^{-1}(\mathcal{M}) = \{(u, w) \in \mathcal{U} \times \mathcal{V} : x + u + w \in \mathcal{M}\}.$$

Notice that for a small vector $u \in \mathcal{U}$, the function (6.2) is exactly p_u, and property (ii) shows

$$\mathcal{Q}_u = \{v(u)\}.$$

It is easy to check

$$N_{\mathcal{Q}}(0, 0) = \{0\} \times \mathcal{V},$$

so Assumption 5.1 holds for our function p. Hence we can apply Theorem 5.7 (strong critical points and parameters) to deduce that $v(u)$ is a strong critical point of p_u relative to $\{v(u)\}$. Hence

$$0 \in \operatorname{ri} \partial p_u(v(u)) = \operatorname{int} \partial p_u(v(u)),$$

since $\operatorname{aff} \partial p_u(v(u)) = N_{\mathcal{Q}_u}(v(u)) = \mathcal{V}$.

To see the "only if" direction of the last statement, we simply consider the function (6.2) with $y = 0$ and $u = 0$. In the converse direction, since x is a local minimizer of $h|_{\mathcal{M}}$, we know $\operatorname{aff} \partial h(x) = \mathcal{V}$, by Proposition 2.4 (smoothness and lineality), and since the origin is a sharp local minimizer of the function p_0, we deduce $0 \in \operatorname{int} \partial p_0(0) = \operatorname{int} \operatorname{proj}_{\mathcal{V}} \partial h(x)$, just like the proof of Theorem 5.7. It follows that $0 \in \operatorname{ri} \partial h(x)$. ∎

The spaces $\mathcal{U}$ and $\mathcal{V}$ in the above result coincide with those in [12] in the convex case.

The idea of partial smoothness is also closely related to the notion of an *identifiable surface* [26] of a convex set. Given a closed convex set $S \subset X$, we call a connected manifold $\mathcal{M} \subset S$ a *(class-C^2) identifiable surface* if either $\mathcal{M}$ is open or for every point $x_0 \in \mathcal{M}$ and every vector $w_0 \in \operatorname{ri} N_S(x_0)$ there exists an open set $V \subset X$ containing x_0 and a smooth function $F : V \to \mathbf{R}^m$ (where m is the codimension of $\mathcal{M}$) such that ∇F is everywhere surjective, $\mathcal{M} \cap V = F^{-1}(0)$, $\nabla F(x)^* \mathbf{R}^m_+ \subset N_S(x)$ for all points $x \in \mathcal{M} \cap V$, and $w_0 \in \nabla F(x_0)^* \mathbf{R}^m_{++}$ (where $\mathbf{R}^m_{++} = \operatorname{int} \mathbf{R}^m_+$).

THEOREM 6.3 (identifiable surfaces). *Consider a closed convex set $S \subset X$ and a connected manifold $\mathcal{M} \subset S$. Then S is partly smooth relative to $\mathcal{M}$ if and only if $\mathcal{M}$ is an identifiable surface.*

Proof. The case when $\mathcal{M}$ is open is immediate, so assume $\mathcal{M}$ has codimension $m > 0$.

Suppose first that $\mathcal{M}$ is an identifiable surface. We need to check the conditions of Proposition 2.11 (partly smooth sets). Condition (i) is immediate, and condition (ii) holds since closed convex sets are everywhere regular. At any point $x_0 \in \mathcal{M}$ we can choose a vector w_0, a neighborhood V, and a function F as in the definition of an identifiable surface, and then we have

$$N_{\mathcal{M}}(x_0) = R(\nabla F(x_0)^*) = \nabla F(x_0)^*(\mathbf{R}^m_+ - \mathbf{R}^m_+) \subset N_S(x_0) - N_S(x_0),$$

so condition (iii) holds.

It remains to show that the normal cone mapping N_S is inner semicontinuous at x_0 relative to $\mathcal{M}$. Since $N_S(x_0)$ is the closure of its relative interior, it suffices to show, for our arbitrary choice $w_0 \in \mathrm{ri}\, N_S(x_0)$, that for any sequence $\{x_r\} \subset \mathcal{M}$ converging to x_0, there exist vectors $w_r \in N_S(x_r)$ converging to w_0. But since $w_0 \in \nabla F(x_0)^* \mathbf{R}^m_{++}$, there exists a vector $\mu \in \mathbf{R}^m_{++}$ such that $w_0 = \nabla F(x_0)^* \mu$, and then the vector

$$w_r = \nabla F(x_r)^* \mu \in \nabla F(x_r)^* \mathbf{R}^m_+$$

lies in $N_S(x_r)$ for all large r and converges to w_0, as required.

Conversely, suppose that the set S is partly smooth relative to the manifold $\mathcal{M}$, and consider a point $x_0 \in \mathcal{M}$ and a vector $w_0 \in \mathrm{ri}\, N_S(x_0)$. By Proposition 2.10 (local normal sharpness) we know $N_{\mathcal{M}}(x) = N_S(x) - N_S(x)$ for all points $x \in \mathcal{M}$ close to x_0, and hence the closed convex cone $N_S(x)$ has the same dimension as the subspace $N_{\mathcal{M}}(x)$, namely m. Thus there exist linearly independent vectors $w_1, w_2, \ldots, w_m \in \mathrm{ri}\, N_S(x_0)$ such that

$$w_0 \in \mathrm{ri}\,(\mathrm{conv}\,\{w_1, w_2, \ldots, w_m\}).$$

Since $\mathcal{M}$ is a manifold of codimension m around x_0, there exists an open set $V \subset X$ containing x_0 and a smooth function $G : V \to \mathbf{R}^m$ such that ∇G is everywhere surjective and $\mathcal{M} \cap V = F^{-1}(0)$. Hence for all points $x \in \mathcal{M} \cap V$ we have $N_{\mathcal{M}}(x) = R(\nabla G(x)^*)$. Since $\nabla G(x)^*$ is injective for all points $x \in V$, there exists a basis $\{a^1, a^2, \ldots, a^m\}$ of $\mathbf{R}^m$ satisfying

$$\nabla G(x_0)^* a^j = w_j \quad \text{for } j = 1, 2, \ldots, m.$$

Now the function $F : V \to \mathbf{R}^m$ defined by

$$(F(x))_j = \langle a^j, G(x) \rangle \quad \text{for } x \in V,\ j = 1, 2, \ldots, m,$$

satisfies $F^{-1}(0) = G^{-1}(0) = \mathcal{M} \cap V$ and

$$\nabla F(x)^* e^j = \nabla G(x)^* a^j \quad \text{for } x \in V,\ j = 1, 2, \ldots, m$$

(where $e^j \in \mathbf{R}^m$ denotes the jth unit vector). Thus $\nabla F(x)^*$ is injective, and so $\nabla F(x)$ is surjective for all points $x \in V$. Also,

$$\nabla F(x_0)^* e^j = \nabla G(x_0)^* a^j = w_j \quad \text{for } j = 1, 2, \ldots, m,$$

so $w_0 \in \nabla F(x_0)^* \mathbf{R}^m_{++}$, as required, and furthermore,

$$N_{\mathcal{M}}(x) = R(\nabla F(x)^*) \quad \text{for } x \in \mathcal{M} \cap V.$$

It remains to prove $\nabla F(x)^* \mathbf{R}^m_+ \subset N_S(x)$ for all points $x \in \mathcal{M}$ close to x_0. If this fails, then for some index j there is a sequence $\{x_r\} \subset \mathcal{M}$ approaching x_0 such that

$$\nabla F(x_r)^* e^j \notin N_S(x_r) \quad \text{for all } r.$$

Both the left- and right-hand sides above are contained in the subspace $N_{\mathcal{M}}(x_r)$, so by separating in this subspace, there exists a unit vector $y_r \in N_{\mathcal{M}}(x_r)$ satisfying

$$\langle y_r, \nabla F(x_r)^* e^j \rangle < \langle y_r, v \rangle \quad \text{for all } v \in N_S(x_r),\ r = 1, 2, \ldots.$$

We can assume, after taking a subsequence, that the sequence $\{y_r\}$ converges to some unit vector $y_0 \in N_{\mathcal{M}}(x_0)$, and since $w_j \in \operatorname{ri} N_S(x_0)$, there exists a real $\delta > 0$ such that $w_j - \delta y_0 \in N_S(x_0)$. Now, since the mapping N_S is continuous, there exist vectors $v_r \in N_S(x_r)$ approaching $w_j - \delta y_0$. But we know

$$\langle y_r, \nabla F(x_r)^* e^j \rangle < \langle y_r, v_r \rangle \quad \text{for } r = 1, 2, \ldots,$$

so taking the limit as $r \to \infty$ gives the contradiction

$$\langle y_0, w_j \rangle \leq \langle y_0, w_j - \delta y_0 \rangle. \qquad \square$$

7. Example. The idea of a strong critical point *decouples* behavior in the active manifold from behavior in directions normal to it. Restricting to the active manifold, a strong critical point is a strong local minimizer, whereas, as we saw in the previous section, any point in the active manifold is a sharp local minimizer with respect to perturbations in normal directions.

One might hope that these properties suffice to ensure that strong critical points of reasonable functions are local minimizers. Unfortunately, this is not the case. We present in this section a locally Lipschitz, everywhere regular function $f : \mathbf{R}^2 \to \mathbf{R}$, partly smooth relative to two distinct manifolds containing the origin. Relative to one manifold, the origin is a strong critical point. However, f restricted to the other manifold has a strong local *maximum* at the origin.

We partition $\mathbf{R}^2$ into four disjoint sets

$$\begin{aligned} S_1 &= \{(x, y) : y \leq 0\}, \\ S_2 &= \{(x, y) : 0 < y < 2x^2\}, \\ S_3 &= \{(x, y) : 0 < 2x^2 \leq y \leq 4x^2\}, \\ S_4 &= \{(x, y) : 4x^2 < y\}, \end{aligned}$$

and we define f by

$$f(x, y) = \begin{cases} x^2 - y & \text{on } S_1, \\ \sqrt{x^4 + 2x^2 y - y^2} & \text{on } S_2, \\ 3x^2 - y & \text{on } S_3, \\ y - 5x^2 & \text{on } S_4. \end{cases}$$

It is easy to check that f is everywhere continuous and in fact is continuously differentiable except on the manifolds

$$\begin{aligned} \mathcal{M}_1 &= \{(x, y) : y = 0\}, \\ \mathcal{M}_2 &= \{(x, y) : y = 4x^2\}. \end{aligned}$$

A calculation shows that $\hat{\partial} f(x,y)$ is given by

$$\begin{cases} \{(2x,-1)\} & \text{on int}\, S_1, \\ \big[(2x,-1),(2x,1)\big] & \text{on } \mathcal{M}_1, \\ \Big\{\Big(1+2\Big(\frac{y}{x^2}\Big)-\Big(\frac{y}{x^2}\Big)^2\Big)^{-1/2}\Big(2x\Big(1+\Big(\frac{y}{x^2}\Big)\Big),1-\frac{y}{x^2}\Big)\Big\} & \text{on } S_2, \\ \{(6x,-1)\} & \text{on } S_3 \setminus \mathcal{M}_2, \\ \big[(6x,-1),(-10x,1)\big] & \text{on } \mathcal{M}_2, \\ \{(-10x,1)\} & \text{on } S_4, \end{cases}$$

where $[u,v]$ denotes the line segment between the points $u,v \in \mathbf{R}^2$. The calculation at every point except the origin is routine, since f is either continuously differentiable at such points or can be written locally as the maximum of two continuously differentiable functions. At the origin we use the inequality

$$|3x^2 - y| - 2x^2 \le f(x,y) \le |3x^2 - y| \quad \text{for all } x,y.$$

The map

$$\beta \in [0,2] \mapsto (1+2\beta-\beta^2)^{-1/2}(1-\beta)$$

has range the interval $[-1,1]$, so for $x \ge 0$,

$$\nabla f(x,y) \in [2x,6x] \times [-1,1] \quad \text{on } S_2, \tag{7.1}$$

and a similar relation holds if $x \le 0$. Hence f is everywhere locally Lipschitz, even around the origin.

We next claim $\partial f = \hat{\partial} f$ everywhere, so f is everywhere regular. As above, this is routine everywhere except at the origin, where it follows using (7.1).

Now it is straightforward to check that the function f is partly smooth relative to both the manifolds $\mathcal{M}_1$ and $\mathcal{M}_2$, and that the origin is a strong critical point relative to $\mathcal{M}_1$. But

$$f(x,y) = -x^2 \quad \text{on } \mathcal{M}_2,$$

so the origin is *not* a local minimizer. In summary, although strong criticality is significant for sensitivity analysis, it is *not* a sufficient condition for optimality.

Acknowledgment. Thanks to Henry Wolkowicz for suggesting the connection between partly smooth sets and identifiable surfaces.

REFERENCES

[1] K. D. Anderson, E. Christiansen, A. R. Conn, and M. L. Overton, *An efficient primal-dual interior-point method for minimizing a sum of Euclidean norms*, SIAM J. Sci. Comput., 22 (2000), pp. 243–262.

[2] V. I. Arnold, *On matrices depending on parameters*, Russian Math. Surveys, 26 (1971), pp. 29–43.

[3] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, SIAM, Philadelphia, 2001.

[4] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.

[5] J. V. Burke, A. S. Lewis, and M. L. Overton, *Optimal stability and eigenvalue multiplicity*, Found. Comput. Math., 1 (2001), pp. 205–225.
[6] P. H. Calamai and A. R. Conn, *A stable algorithm for solving the multifacility location problem involving Euclidean distances*, SIAM J. Sci. Stat. Comput., 1 (1980), pp. 512–526.
[7] P. H. Calamai and A. R. Conn, *A projected Newton method for l_p norm location problems*, Math. Programming, 38 (1987), pp. 75–109.
[8] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, New York, 1968; reprinted, SIAM, Philadelphia, 1990.
[9] R. Fletcher, *Practical Methods of Optimization*, 2nd ed., Wiley, Chichester, UK, 1987.
[10] R. A. Horn and C. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.
[11] T. Kato, *A Short Introduction to Perturbation Theory for Linear Operators*, Springer, New York, 1982.
[12] C. Lemaréchal, F. Oustry, and C. Sagastizábal, *The $\mathcal{U}$-Lagrangian of a convex function*, Trans. Amer. Math. Soc., 352 (2000), pp. 711–729.
[13] C. Lemaréchal and C. Sagastizábal, *Practical aspects of the Moreau–Yosida regularization: Theoretical preliminaries*, SIAM J. Optim., 7 (1997), pp. 367–385.
[14] A. S. Lewis and M. L. Overton, *Eigenvalue optimization*, Acta Numer., 5 (1996), pp. 149–190.
[15] R. Mifflin and C. Sagastizábal, *Proximal Points Are on the Fast Track*, Technical report, Washington State University, Pullman, WA, 2000.
[16] Y. Nesterov and A. Nemirovskii, *Interior-Point Polynomial Algorithms in Convex Programming*, SIAM, Philadelphia, 1994.
[17] M. R. Osborne, *Simplicial Algorithms for Minimizing Polyhedral Functions*, Cambridge University Press, Cambridge, UK, 2001.
[18] F. Oustry, *The $\mathcal{U}$-Lagrangian of the maximum eigenvalue function*, SIAM J. Optim., 9 (1999), pp. 526–549.
[19] F. Oustry, *A second-order bundle method to minimize the maximum eigenvalue function*, Math. Program., 89 (2000), pp. 1–33.
[20] M. L. Overton, *A quadratically convergent method for minimizing a sum of Euclidean norms*, Math. Programming, 27 (1983), pp. 34–63.
[21] M. L. Overton and R. S. Womersley, *Second derivatives for optimizing eigenvalues of symmetric matrices*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 697–718.
[22] M. L. Overton and X. Ye, *Towards second-order methods for structured nonsmooth optimization*, in Advances in Optimization and Numerical Analysis, S. Gomez and J.-P. Hennart, eds., Kluwer, Amsterdam, 1994, pp. 97–109.
[23] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer, Berlin, 1998.
[24] A. Shapiro and M. K. H. Fan, *On eigenvalue optimization*, SIAM J. Optim., 5 (1995), pp. 552–569.
[25] I. Vaisman, *A First Course in Differential Geometry*, Marcel Dekker, New York, 1984.
[26] S. J. Wright, *Identifiable surfaces in constrained optimization*, SIAM J. Control Optim., 31 (1993), pp. 1063–1079.