PARTIAL SMOOTHNESS AND CONSTANT RANK*

A.S. LEWIS†, JINGWEI LIANG‡, AND TONGHUA TIAN§

Abstract. In optimization, the notion of a partly smooth objective function is powerful for applications in algorithmic convergence and post-optimality analysis, and yet is complex to define. A shift in focus to the first-order optimality conditions reduces the concept to a simple constant-rank condition. In this view, partial smoothness extends to more general variational systems, encompassing in particular the saddlepoint operators underlying popular primal-dual splitting algorithms. For a broad class of semi-algebraic generalized equations, partial smoothness holds generically.

Key words. partial smoothness, active set identification, nonsmooth optimization, subdifferential, primal-dual splitting, semi-algebraic

AMS subject classifications. 90C31, 49M05, 65K10

1. Introduction.

A variety of optimization algorithms, ranging from classical active set methods to contemporary first-order algorithms for machine learning and high-dimensional statistics, "identify" structure associated with optimal solutions. The structure in question might be the active constraint set in a classical nonlinear program or, more specifically, the sparsity pattern of solution vectors in a machine learning problem. In the underlying Euclidean variable space $\mathbf{U}$, the algorithm generates iterates that eventually lie in a subset $\mathcal{M} \subset \mathbf{U}$ of feasible solutions with that particular structure. That subset $\mathcal{M}$, in the case of nonlinear programming, is exactly the "identifiable surface" [37] of the feasible region defined by the constraints active at optimality.

A simple but quite extensive model of this phenomenon, following the philosophy of [37], is as follows. We consider minimizing a lower semicontinuous objective function $f\colon \mathbf{U} \to \overline{\mathbf{R}}$ (convex, for now), and assume that the set $\mathcal{M}$ of interest is a smooth surface, or more precisely a manifold around an optimal solution $\bar{u}$, meaning that locally it consists of solutions of a system of $C^{(2)}$-smooth equations with linearly independent gradients. Identification amounts to the property

$$v_k \in \partial f(u_k), \quad u_k \to \bar{u}, \quad v_k \to 0 \quad \Rightarrow \quad u_k \in \mathcal{M} \text{ eventually,} \tag{1.1}$$

where ∂f denotes the classical subdifferential operator. Earlier versions of this identifiability idea include [4, 14, 3, 2, 1, 16, 17].

Closely related to the identification property (1.1) is the idea that the function f is *partly smooth* at the point $\bar{u}$ relative to the manifold $\mathcal{M}$. This property combines smoothness conditions on f when restricted to $\mathcal{M}$ with a sharpness property of f in directions normal to $\mathcal{M}$. More precisely, around the point $\bar{u}$ the restrictions of the function f and its subdifferential ∂f to the manifold $\mathcal{M}$ should be $C^{(2)}$-smooth and continuous respectively, and the affine span of $\partial f(\bar{u})$ should be a translate of the normal space to $\mathcal{M}$ at $\bar{u}$. This property, along with the nondegeneracy assumption that zero lies in the relative interior of $\partial f(\bar{u})$, together suffice to ensure identifiability (1.1), as shown in [29, Thm 4.10].

*Submitted to the editors January 8, 2019.

†ORIE, Cornell University, Ithaca, NY (http://people.orie.cornell.edu/aslewis). Research supported in part by National Science Foundation Grant DMS-2006990.

‡Institute of Natural Sciences and School of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai China (jingwei.liang@sjtu.edu.cn).

§ORIE, Cornell University, Ithaca, NY (tt543@cornell.edu).

As a simple example, the function f on the space $\mathbf{R}^2$ defined by $f(x,y) = |x| + y^2$ is partly smooth at its minimizer $(0,0)$ relative to the manifold $\{0\} \times \mathbf{R}$, and zero lies in the relative interior of the subdifferential $\partial f(0,0) = [-1,1] \times \{0\}$. Hence the identifiability property (1.1) holds, as is easy to verify directly.

The terminology and original definition of partly smooth sets and functions originated in [26]. A closely related thread of research, known as "$\mathcal{V}U$ theory", emerged from original work in [25]: subsequent explorations include [24, 31, 34, 33, 19, 32]. Inevitably, it seems, the formal definition of partly smooth sets and functions, and their $\mathcal{V}U$ analogues, are rather involved. The definition of an identifiable surface in [37] is not simple either, despite the transparency of the identifiability property (1.1).

As an approach to identifiability, considering partly smooth functions seems roundabout: our aim, the property (1.1), involves only the subdifferential operator ∂f, and not the underlying function f. It turns out that we can indeed characterize identifiability more naturally through a simple and fundamental property of the underlying operator ∂f. Simply put, if the graph of the operator (in the product space $\mathbf{U} \times \mathbf{U}$) is a smooth manifold around the point $(\bar{u}, 0)$, and the canonical projection of nearby points (u,v) in the graph to $u \in \mathbf{U}$ is *constant rank* (meaning that the projected tangent spaces at those points have constant dimension), then the identifiability property (1.1) follows.

In summary, this work's main contribution is to simplify and clarify the notion of partial smoothness, and how it relates to identifiability, revealing these ideas in essence as constant-rank properties. By simplifying the definition, we hope to promote partial smoothness as a conceptual tool. After reviewing basic ideas about manifolds in section 2, we present the new definition and its consequences in section 3. In section 4, we refine a remarkable Sard-type result of Ioffe [21], showing that semi-algebraic generalized equations are "regular" for generic data values. Consequently, such equations reduce to intersection problems that are numerically stable due to a transversality condition. In common settings — subdifferential criticality conditions, variational inequalities and complementarity problems — we prove that partial smoothness also holds generically. Some technical representation tools follow in section 5, before we verify, in section 6, that the new definition does indeed capture the full force of previous notions of partial smoothness. This new perspective not only clarifies our understanding of these powerful tools, but broadens their potential application beyond the basic optimality condition $0 \in \partial f(\bar{u})$ to more general variational conditions. As an example, in section 7, we consider the saddlepoint optimality conditions associated with primal-dual splitting methods like the Chambolle-Pock algorithm [5]. We end, in section 8, with a brief discussion of the second-order optimality conditions in this new perspective.

2. Manifolds. Throughout this work we consider mappings between two Euclidean spaces $\mathbf{U}$ and $\mathbf{V}$. We begin our more formal development by summarizing some elementary ideas about manifolds, following the terminology of [22]. To this end, consider a set $\mathcal{M} \subset \mathbf{U}$ that has the structure of a smooth manifold locally, around a point $\bar{u} \in \mathcal{M}$. By "smooth", we mean $C^{(1)}$-smooth, unless we state otherwise. We can consider such sets $\mathcal{M}$ using "local coordinates", as follows.

We denote the open ball of radius $\delta > 0$ around the point $\bar{u}$ by $B_\delta(\bar{u})$. In elementary language, $\mathcal{M}$ is a *smooth manifold around* $\bar{u}$ when there exists a Euclidean space $\mathbf{W}$ and a map $H \colon \mathbf{W} \to \mathbf{U}$ that is smooth around 0, with the derivative $\nabla H(0) \colon \mathbf{W} \to \mathbf{U}$ injective and $H(0) = \bar{u}$, and such that, for all small $\delta > 0$,

$$\mathcal{M} = H\big(B_\delta(\bar{u})\big) \quad \text{around } \bar{u}.$$

More formally [22, Chapter 8], some open neighborhood of $\bar{u}$ in $\mathcal{M}$ is an *embedded submanifold* of $\mathbf{U}$. Any small vector $w \in \mathbf{W}$ constitutes the *local coordinates centered around* $\bar{u}$ for the point $H(w) \in \mathcal{M}$. The *tangent space* at such a point is given simply by

$$T_{\mathcal{M}}\big(H(w)\big) = \mathrm{Range}\big(\nabla H(w)\big).$$

Its dimension (the *dimension* of $\mathcal{M}$ around $\bar{u}$) is a constant, namely $\dim \mathbf{W}$. The *normal space* is the orthogonal complement:

$$N_{\mathcal{M}}\big(H(w)\big) = \mathrm{Null}\big(\nabla H(w)^*\big).$$

Given another Euclidean space $\mathbf{V}$, a map $F\colon \mathcal{M} \to \mathbf{V}$ is *smooth around* $\bar{u}$ when there exists a map $G\colon \mathbf{U} \to \mathbf{V}$ that is smooth around $\bar{u}$ and agrees with F on a neighborhood of $\bar{u}$ in $\mathcal{M}$. In that case, the *rank* of F at $\bar{u}$ is $\dim\big(\nabla G(\bar{u}) T_{\mathcal{M}}(\bar{u})\big)$. Equivalently, F is smooth around $\bar{u}$ when the composition $F \circ H$ is smooth around 0, and its rank at $\bar{u}$ is then rank of the derivative $\nabla(F \circ H)(0)\colon \mathbf{W} \to \mathbf{V}$ as a linear map.

The map H defines a diffeomorphism from the open ball $B_\delta(0) \subset \mathbf{W}$ (for small $\delta > 0$) to an open neighborhood of the point $\bar{u}$ in the manifold $\mathcal{M}$. We can describe the inverse of this diffeomorphism via a map $G\colon \mathbf{U} \to \mathbf{W}$, smooth around the point $\bar{u}$, and satisfying

$$G\big(H(w)\big) = w \quad \text{for all small vectors } w \in \mathbf{W}. \tag{2.1}$$

The restriction $G|_{\mathcal{M}}$, around $\bar{u}$, is the inverse of the diffeomorphism H.

Adopting a dual approach, we can equivalently define a set $\mathcal{M} \subset \mathbf{U}$ to be a smooth manifold around a point $\bar{u}$ when there exists a Euclidean space $\mathbf{X}$ and a map $P\colon \mathbf{U} \to \mathbf{X}$ that is smooth around $\bar{u}$, with the derivative $\nabla P(\bar{u})\colon \mathbf{U} \to \mathbf{X}$ surjective and $P(\bar{u}) = 0$, and such that

$$\mathcal{M} \;=\; P^{-1}(0) \;=\; \{u \in \mathbf{U} : P(u) = 0\} \quad \text{around } \bar{u}.$$

Then the tangent and normal spaces are given by

$$T_{\mathcal{M}}(u) = \mathrm{Null}\big(\nabla P(u)\big)$$
$$N_{\mathcal{M}}(u) = \mathrm{Range}\big(\nabla P(u)^*\big)$$

at all points $u \in \mathcal{M}$ near $\bar{u}$. The normal space has the same dimension as $\mathbf{X}$.

We can naturally decompose the space U as a direct sum:

$$\mathbf{U} = T_{\mathcal{M}}(\bar{u}) \oplus N_{\mathcal{M}}(\bar{u}).$$

With this decomposition, the two derivatives $\nabla H(0)\colon \mathbf{W} \to \mathbf{U}$ and $\nabla P(\bar{u})\colon \mathbf{U} \to \mathbf{X}$ are given by

$$\nabla H(0)w = (Dw, 0)$$
$$\nabla P(\bar{u})(r, s) = Es$$

for some invertible linear maps $D\colon \mathbf{W} \to T_{\mathcal{M}}(\bar{u})$ and $E\colon N_{\mathcal{M}}(\bar{u}) \to \mathbf{X}$. Furthermore, the derivative $\nabla G(\bar{u})\colon \mathbf{U} \to \mathbf{W}$, restricted to $T_{\mathcal{M}}(\bar{u})$, is just the inverse map D^{-1}.

3. Partly smooth mappings. We are now ready to define the property central to this work. We consider the canonical projection $\mathrm{proj}\colon \mathbf{U}\times\mathbf{V} \to \mathbf{U}$ defined by $\mathrm{proj}(u,v) = u$.

DEFINITION 3.1 (Partly smooth mappings). A set-valued mapping $\Phi\colon \mathbf{U} \rightrightarrows \mathbf{V}$ is called *partly smooth at* a point $\bar u \in \mathbf{U}$ *for* a value $\bar v \in \Phi(\bar u)$ when the graph $\mathrm{gph}\,\Phi$ is a smooth manifold around $(\bar u, \bar v)$ and the projection proj restricted to $\mathrm{gph}\,\Phi$ has constant rank around $(\bar u,\bar v)$. The *dimension* of Φ at $\bar u$ for $\bar v$ is then just the dimension of its graph around $(\bar u,\bar v)$.

Note. An example is when the inverse mapping $\Phi^{-1}\colon \mathbf{V} \rightrightarrows \mathbf{U}$ is locally single-valued, smooth and constant-rank around $\bar v$ for $\bar u$. In this case, Φ is in particular "strongly (metrically) regular" at $\bar u$ for $\bar v$, in the terminology of [35].

By definition, the constant rank condition means that the subspace

$$\mathrm{proj}\,\big(T_{\mathrm{gph}\,\Phi}(u,v)\big)$$

and its orthogonal complement (called, in variational analysis [35], the *coderivative* of the mapping Φ)

$$D^*\Phi(u,v)(0) \;=\; \big\{w\in\mathbf{U} : (w,0)\in N_{\mathrm{gph}\,\Phi}(u,v)\big\},$$

or equivalently, the subspace

$$N_{\mathrm{gph}\,\Phi}(u,v)\cap(\mathbf{U}\times\{0\})$$

all have constant dimension for points (u,v) near $(\bar u,\bar v)$.

Consider, for example, the set-valued mapping $\Phi\colon \mathbf{R}\rightrightarrows\mathbf{R}$ defined by

$$\Phi(u) = \begin{cases} \{\pm\sqrt{u}\} & (u\geq 0)\\ \emptyset & (u<0).\end{cases}$$

The graph is of Φ is the manifold $\{(u,v)\in\mathbf{R}^2 : u = v^2\}$. However, Φ is not partly smooth at 0 for 0, because the projection proj restricted to $\mathrm{gph}\,\Phi$ has rank zero at the point $(0,0)$ but rank one at all nearby points.

PROPOSITION 3.2. *If a set-valued mapping $\Phi\colon \mathbf{U}\rightrightarrows\mathbf{V}$ is partly smooth at a point $\bar u\in\mathbf{U}$ for a value $\bar v\in\Phi(\bar u)$, then there exists a set $\mathcal{M}\subset\mathbf{U}$, uniquely defined around $\bar u$, that is a smooth manifold around $\bar u$, and satisfies*

$$\mathcal{M} \;=\; \big\{u\in B_\epsilon(\bar u) : \exists v\in\Phi(u)\cap B_\epsilon(\bar v)\big\} \quad \textit{around } \bar u,$$

for all small $\epsilon>0$. We call any such set $\mathcal{M}$ the **active manifold***.*

Proof. For any small $\epsilon>0$, the set

$$G_\epsilon \;=\; \mathrm{gph}\,\Phi\cap\big(B_\epsilon(\bar u)\times B_\epsilon(\bar v)\big)$$

is a manifold, and the projection proj restricted to G_ϵ is a constant-rank map. By the Constant Rank Theorem [22], the resulting image

$$\mathcal{M}_\epsilon \;=\; \big\{u\in B_\epsilon(\bar u) : \exists v\in\Phi(u)\cap B_\epsilon(\bar v)\big\}$$

is a manifold of dimension $\dim\mathrm{proj}\,T_{G_\epsilon}(\bar u,\bar v)$. This dimension is constant, for small $\epsilon>0$, since the tangent space satisfies $T_{G_\epsilon}(\bar u,\bar v) = T_{\mathrm{gph}\,\Phi}(\bar u,\bar v)$. For any $\epsilon'\in(0,\epsilon)$, we know $\mathcal{M}_{\epsilon'}\subset\mathcal{M}_\epsilon$, but these sets are manifolds around $\bar u$ of the same dimension, so must be identical around $\bar u$. ∎

We use the following definition [12].

Definition 3.3. A set $\mathcal{M} \subset \mathbf{U}$ is *identifiable* for a set-valued mapping $\Phi\colon \mathbf{U} \rightrightarrows \mathbf{V}$ at a point $\bar{u} \in \mathbf{U}$ for a value $\bar{v} \in \Phi(\bar{u})$ when $\operatorname{gph} \Phi \subset \mathcal{M} \times \mathbf{V}$ around the point $(\bar{u}, \bar{v})$.

The following proposition is then immediate.

Proposition 3.4. *If a set-valued mapping $\Phi\colon \mathbf{U} \rightrightarrows \mathbf{V}$ is partly smooth at a point $\bar{u} \in \mathbf{U}$ for a value $\bar{v} \in \Phi(\bar{u})$, then the active manifold is an identifiable set.*

In fact, as we see later in our development, the active manifold is a locally minimal identifiable set.

4. A Sard-type theorem for generalized equations. A value $v \in \mathbf{V}$ is *regular* for a set-valued mapping $\Phi\colon \mathbf{U} \rightrightarrows \mathbf{V}$ if the graph of Φ intersects $\mathbf{U} \times \{v\}$ everywhere transversally: using the language of coderivatives [35],

$$0 \in D^*\Phi(u,v)(w) \;\Rightarrow\; w = 0 \qquad (u \in \mathbf{U},\; w \in \mathbf{V}).$$

The numerical solution of equations involving a sufficiently smooth *single*-valued map Φ is buttressed by the classical Sard theorem, which guarantees that almost all values $v \in \mathbf{V}$ are regular, or equivalently, that the derivative of Φ is surjective throughout the inverse image $\Phi^{-1}(v)$. In computational practice, we make no attempt to verify regularity in advance: rather it is a reassuring generic property.

Sard-type theorems also hold for a large class of concrete *set*-valued mappings $\Phi\colon \mathbf{U} \rightrightarrows \mathbf{V}$, and in particular for those with semi-algebraic graphs [21]. A subset of a Euclidean space is *semi-algebraic* if it is a finite union of sets defined by finitely-many polynomial inequalities [7]. Such sets comprise a rich and robust class, representative of many concrete variational settings [27]. Our next result incorporates partial smoothness into a semi-algebraic Sard-type result. Thus even without verifying partial smoothness in advance, for generic data we may reasonably assume it. For a more detailed analysis in the case of certain variational inequalities, see [23].

Theorem 4.1 (Generic partial smoothness). *If the graph of a set-valued mapping $\Phi\colon \mathbf{U} \rightrightarrows \mathbf{V}$ is semi-algebraic and of dimension no larger than $\dim \mathbf{V}$, then regularity holds for almost all values $v \in \mathbf{V}$, along with the property of partial smoothness throughout the inverse image $\Phi^{-1}(v)$. These two properties hold in particular on a full-measure union of open semi-algebraic sets $V_i \subset \mathbf{V}$ (for $i = 0, 1, 2, \ldots, k$), on each of which the inverse image $\Phi^{-1}(\cdot)$ has constant cardinality i and the representation*

$$\Phi^{-1}(v) \;=\; \{G_i^j(v) : j = 1, 2, 3, \ldots, i\}$$

for single-valued, analytic, semi-algebraic, constant-rank maps $G_i^j\colon V_i \to \mathbf{U}$.

Proof. By [11, Proposition 4.3], there exists an integer $k \geq 0$ such that the inverse image $\Phi^{-1}(v)$ has cardinality at most k for almost all values $v \in \mathbf{V}$. Hence the union of the following semi-algebraic sets is full-measure in $\mathbf{V}$:

$$\tilde{V}_i \;=\; \{v \in \mathbf{V} : \Phi^{-1}(v) \text{ has cardinality } i\} \qquad (i = 0, 1, 2, \ldots, k).$$

For vectors $x, y \in \mathbf{R}^n$, the lexicographic relationship $x \succ y$ means that for some $j = 1, 2, 3, \ldots, n$, we have $x_i = y_i$ for all $i < j$ and $x_j > y_j$. By identifying the space $\mathbf{V}$ with $\mathbf{R}^n$, we thereby arrive at a strict total order $\succ$ on $\mathbf{V}$, with semi-algebraic graph. Now define semi-algebraic maps $G_i^j\colon \tilde{V}_i \to \mathbf{U}$, for each $i = 0, 1, 2, \ldots, k$ and $j = 1, 2, 3, \ldots, i$, as follows. For any value $v \in \tilde{V}_i$, we write

$$\Phi^{-1}(v) \;=\; \{u_1, u_2, u_3, \ldots, u_i\} \quad \text{with} \quad u_1 \succ u_2 \succ u_3 \succ \cdots \succ u_i,$$

and then define $G_i^j(v) = u_j$.

Using a standard stratification argument (see [18] and [10, Theorem 3.4]), we can partition each set $\tilde{V}_i$ into a disjoint finite union of "strata" — semi-algebraic analytic manifolds of various dimensions — on each of which all the maps G_i^j (for $j = 1, 2, 3, \ldots, i$) are analytic and constant-rank. The union of all these strata (for $i = 0, 1, 2, \ldots, k$) is full-measure in $\mathbf{V}$, and hence so is the union of just the open strata, since lower-dimensional manifolds have measure zero. We can therefore define each set V_i to be the union of just the open strata for $\tilde{V}_i$.

Consider any value $\bar{v}$ in the set V_i, for some $i = 0, 1, 2, \ldots, k$. Any point $\bar{u}$ with $\bar{v} \in \Phi(\bar{u})$ can be written $\bar{u} = G_i^j(\bar{v})$ for some $j = 1, 2, 3, \ldots, i$. Since V_i is open, we then have

$$\operatorname{gph} \Phi \;=\; \{(G_i^j(v), v) : v \in V_i\} \quad \text{around} \quad (\bar{u}, \bar{v}).$$

Thus the graph of Φ is locally identical to the graph of an analytic map (after reordering the variables). A standard exercise shows that its tangent space at any nearby point $(u, v) \in \operatorname{gph} \Phi$ is

$$T_{\operatorname{gph} \Phi}(u, v) \;=\; \{(\nabla G_i^j(v) w, w) : w \in \mathbf{V}\},$$

so we deduce

$$\operatorname{proj}\big(T_{\operatorname{gph} \Phi}(u, v)\big) \;=\; \operatorname{Range}\big(\nabla G_i^j(v)\big).$$

Since the map G_i^j is constant-rank, Φ is partly smooth at $\bar{u}$ for $\bar{v}$.

It remains to check that the graph of Φ intersects $\mathbf{U} \times \{\bar{v}\}$ transversally at the point $(\bar{u}, \bar{v})$. This follows from the fact that the graph is locally a manifold, and the tangent spaces

$$T_{\operatorname{gph} \Phi}(\bar{u}, \bar{v}) = \{(\nabla G_i^j(\bar{v}) w, w) : w \in \mathbf{V}\}$$
$$T_{\mathbf{U} \times \{\bar{v}\}}(\bar{u}, \bar{v}) = \mathbf{U} \times \{0\},$$

together span $\mathbf{U} \times \mathbf{V}$. $\square$

Examples of set-valued mappings $\Phi \colon \mathbf{V} \rightrightarrows \mathbf{V}$ with semi-algebraic graphs of dimension no larger than $\dim \mathbf{V}$ include subdifferentials of semi-algebraic functions $f \colon \mathbf{V} \to \overline{\mathbf{R}}$ (see [9, 11]). However, such mappings also appear in settings beyond optimization. Generalized equations such as variational inequalities and complementarity problems [15, 8] often involve mappings Φ that decompose into a sum of a single-valued map $F \colon \mathbf{V} \to \mathbf{V}$ and a maximal monotone operator $\Psi \colon \mathbf{V} \rightrightarrows \mathbf{V}$. By Minty's Theorem [35, Theorem 12.15], we have

$$\operatorname{gph}(F + \Psi) \;=\; \big\{\big(R_\Psi(v), F\big(R_\Psi(v)\big) + R_{\Psi^{-1}}(v)\big) : v \in \mathbf{V}\big\},$$

where the resolvents $R_\Psi = (I + \Psi)^{-1}$ and $R_{\Psi^{-1}} = (I + \Psi^{-1})^{-1}$ are both single-valued. If F and Ψ are both semi-algebraic, then so are the two resolvents, and hence, being a single-valued semi-algebraic image of $\mathbf{V}$, the graph of the sum $F + \Psi$ is semi-algebraic and of dimension no larger than $\dim \mathbf{V}$.

Theorem 4.1 thus shows, for concrete variational problems with generic data, that we can expect both partial smoothness and regularity to hold at a solution. These properties together open the door to "active-set Newton methods", generalizing sequential quadratic programming approaches for traditional nonlinear programming: for a full development, see [28, Theorem 3]. Of course, in practice, data may not be generic, but we would likely not try to verify partial smoothness or regularity in advance any more than we would verify, before using a numerical solver, a nonzero derivative at the solution of a scalar equation. Nonetheless, both those assumptions may be natural first steps in sensitivity analysis and algorithm design.

5. Representations of partly smooth mappings. In this section we develop two useful concrete representations of partly smooth mappings. The proofs, while routine, are a little fussy, and the reader might wish to pass them over at first. The first result gives a representation of a partly smooth mapping using local coordinates.

THEOREM 5.1 (Coordinate representation). *A set-valued mapping $\Phi\colon \mathbf{U} \rightrightarrows \mathbf{V}$ is partly smooth at a point $\bar u \in \mathbf{U}$ for a value $\bar v \in \Phi(\bar u)$ if and only if it has a local representation of the following form: there exist Euclidean spaces $\mathbf{W}$ and $\mathbf{Z}$, maps $H\colon \mathbf{W} \to \mathbf{U}$, smooth around 0 with $H(0) = \bar u$ and $\nabla H(0)$ injective, and $G\colon \mathbf{W}\times\mathbf{Z} \to \mathbf{V}$, smooth around $(0,0)$ with $G(0,0) = \bar v$, such that,*

$$(5.1)\quad w \in \mathbf{W},\ z \in \mathbf{Z},\ \nabla H(0)w = 0 \text{ and } \nabla G(0,0)(w,z) = 0 \quad\Rightarrow\quad w = 0 \text{ and } z = 0,$$

and for all small $\delta > 0$,

$$(5.2)\qquad \operatorname{gph}\Phi \;=\; \big\{\big(H(w), G(w,z)\big) : w \in B_\delta(0),\ z \in B_\delta(0)\big\} \;\text{ around } (\bar u, \bar v).$$

In this case, the dimension of Φ at $\bar u$ for $\bar v$ is $\dim \mathbf{W} + \dim \mathbf{Z}$, and the active manifold is $H\big(B_\delta(0)\big)$ around $\bar u$, providing $\delta > 0$ is sufficiently small.

Proof. Assuming the local representation, we first prove that Φ is partly smooth at $\bar u$ for $\bar v$. Consider the map $P\colon \mathbf{W}\times\mathbf{Z} \to \mathbf{U}\times\mathbf{V}$ defined by $P(w,z) = \big(H(w), G(w,z)\big)$ for $w \in \mathbf{W}$ and $z \in \mathbf{Z}$. This map is smooth around the point $(0,0)$, with derivative

$$\nabla P(w,z)(r,s) = \big(\nabla H(w)r, \nabla G(w,z)(r,s)\big),$$

for all small $w \in \mathbf{W}$ and $z \in \mathbf{Z}$, and vectors $r \in \mathbf{W}$ and $s \in \mathbf{Z}$. By assumption, the derivative $\nabla P(0,0)$ is injective, so $\operatorname{gph}\Phi$ is a smooth manifold around $(0,0)$, with tangent space at such points (w,z) given by

$$T_{\operatorname{gph}\Phi}\big(H(w), G(w,z)\big) \;=\; \big\{\big(\nabla H(w)r, \nabla G(w,z)(r,s)\big) : r \in \mathbf{W},\ s \in \mathbf{Z}\big\}.$$

Its image under the projection map $\operatorname{proj}\colon \operatorname{gph}\Phi \to \mathbf{U}$ is simply the range of $\nabla H(w)$. Since $\nabla H(0)$ is injective, the projection has locally constant rank $\dim \mathbf{W}$. Partial smoothness follows, and the local description of the active manifold follows from Proposition 3.2.

Conversely, suppose $\Phi\colon \mathbf{U} \rightrightarrows \mathbf{V}$ is partly smooth at $\bar u$ for $\bar v \in \Phi(\bar u)$. By the Constant Rank Theorem, we can consider the projection map proj as having the form $(w,z) \mapsto (w,0) \in \mathbf{W}\times\mathbf{Y}$, where $(w,z) \in \mathbf{W}\times\mathbf{Z}$ (for Euclidean spaces $\mathbf{W}$ and $\mathbf{Z}$) defines local coordinates for the manifold $\operatorname{gph}\Phi$, centered at $(\bar u, \bar v)$, and $(w,y) \in \mathbf{W}\times\mathbf{Y}$ (for a Euclidean space $\mathbf{Y}$) defines local coordinates for $\mathbf{U}$ centered around $\bar u$.

More explicitly, there exist maps

$$\begin{aligned} F\colon \mathbf{W}\times\mathbf{Z} \to \mathbf{U}&, \text{ smooth around } (0,0), \text{with } F(0,0) = \bar u\\ G\colon \mathbf{W}\times\mathbf{Z} \to \mathbf{V}&, \text{ smooth around } (0,0), \text{with } G(0,0) = \bar v\\ Q\colon \mathbf{W}\times\mathbf{Y} \to \mathbf{U}&, \text{ smooth around } (0,0), \text{with } Q(0,0) = \bar u \end{aligned}$$

with

$$\begin{aligned} \big(\nabla F(0,0), \nabla G(0,0)\big)&\colon \mathbf{W}\times\mathbf{Z} \to \mathbf{U}\times\mathbf{V}\\ \nabla Q(0,0)&\colon \mathbf{W}\times\mathbf{Y} \to \mathbf{U} \end{aligned}$$

both injective, and for all small $\delta > 0$,

$$\begin{aligned} \operatorname{gph}\Phi &= \big\{\big(F(w,z), G(w,z)\big) : w \in B_\delta(0),\ z \in B_\delta(0)\big\} \;\text{ around } (\bar u, \bar v)\\ \mathbf{U} &= \big\{Q(w,y) : w \in B_\delta(0),\ y \in B_\delta(0)\big\} \;\text{ around } \bar v, \end{aligned}$$

and furthermore, $F(w,z) = Q(w,0)$ for all small $w \in \mathbf{W}$ and $z \in \mathbf{Z}$.

Now define a map $H\colon \mathbf{W} \to \mathbf{U}$ by $H(w) = Q(w,0)$, for $w \in \mathbf{W}$, and notice $\nabla F(0,0) = (\nabla H(0), 0)$. Then, for points $w \in \mathbf{W}$ and $z \in \mathbf{Z}$, whenever $0 = \nabla H(0)w = \nabla F(0,0)(w,z)$ and $\nabla G(0,0)(w,z) = 0$, we must have $w = 0$ and $z = 0$. The result now follows. ∎

One consequence is the locally minimal identifiability of active manifolds we mentioned above, as we show next. Along the way, we prove that partly smooth mappings have smooth selections.

Corollary 5.2 (Minimal identifiability). *If a set-valued mapping $\Phi\colon \mathbf{U} \rightrightarrows \mathbf{V}$ is partly smooth at a point $\bar{u} \in \mathbf{U}$ for a value $\bar{v} \in \Phi(\bar{u})$, then the active manifold $\mathcal{M}$ has the following properties.*

- *There exists a map $F\colon \mathcal{M} \to \mathbf{V}$, smooth around $\bar{u}$, such that $F(\bar{u}) = \bar{v}$ and $F(u) \in \Phi(u)$ for all points $u \in \mathcal{M}$ near $\bar{u}$.*
- *For any set $\mathcal{M}' \subset \mathbf{U}$ containing $\bar{u}$, and any map $F'\colon \mathcal{M}' \to \mathbf{V}$ that is continuous at '$\bar{u}$ and satisfies $F(\bar{u}) = \bar{v}$ and $F(u) \in \Phi(u)$ for all points $u \in \mathcal{M}'$ near $\bar{u}$, we must have $\mathcal{M}' \subset \mathcal{M}$ around $\bar{u}$.*
- *$\mathcal{M}$ is a locally minimal identifiable set at $\bar{u}$ for $\bar{v}$.*

Proof. To see the first property, we apply the coordinate representation guaranteed by Theorem 5.1, and define the map F by $F\big(H(w)\big) = G(w,0)$ for small vectors $w \in \mathbf{W}$. The last property follows, since we just need to show the following inner semicontinuity property (see [12, Proposition 2.8]): for any sequence of points $u_r \to \bar{u}$ in the active manifold $\mathcal{M}$, there exists a sequence of values $v_r \to \bar{v}$ with $v_r \in \Phi(u_r)$ for all large indices r. To see this, simply set $v_r = F(u_r)$.

To see the second property, consider any sequence $u_r \in \mathcal{M}'$ converging to $\bar{u}$. By assumption, the sequence $\big(u_r, F'(u_r)\big) \in \operatorname{gph}\Phi$ converges to the point $(\bar{u},\bar{v})$, so $u_r \in \mathcal{M}$ for all large indices r by Proposition 3.4. ∎

We also have the following calculus rule.

Corollary 5.3 (Sum rule). *Consider a set-valued mapping $\Phi\colon \mathbf{U} \rightrightarrows \mathbf{V}$ that is partly smooth at a point $\bar{u} \in \mathbf{U}$ for a value $\bar{v} \in \Phi(\bar{u})$. If the function $F\colon \mathbf{U} \to \mathbf{V}$ is smooth around $\bar{u}$, then the set-valued mapping $\Phi + F$ is partly smooth at $\bar{u}$ for $\bar{v} + F(\bar{u})$, with the same dimension and active manifold.*

Proof. In terms of the coordinate representation guaranteed by Theorem 5.1, we have

$$\operatorname{gph}(\Phi + F) \;=\; \big\{\big(H(w), \tilde{G}(w,z)\big) : w \in B_\delta(0),\ z \in B_\delta(0)\big\} \;\text{ around } (\bar{u},\bar{v}),$$

where the map $\tilde{G}\colon \mathbf{W} \times \mathbf{Z} \to \mathbf{V}$ is defined by

$$\tilde{G}(w,z) = G(w,z) + F\big(H(w)\big) \quad (w \in \mathbf{W},\ z \in \mathbf{Z}).$$

This map is smooth around the point $(0,0)$ with $\tilde{G}(0,0) = \bar{v} + F(\bar{u})$. Furthermore, by assumption,

$$w \in \mathbf{W},\ z \in \mathbf{Z},\ \nabla H(0)w = 0 \text{ and } \nabla\tilde{G}(0,0)(w,z) = 0 \quad\Rightarrow\quad w = 0 \text{ and } z = 0,$$

since

$$\nabla\tilde{G}(0,0)(w,z) = \nabla G(0,0)(w,z) + \nabla F(\bar{u})\nabla H(0)w.$$

The result now follows by Theorem 5.1. ∎

As with manifolds, a dual representation is sometimes more useful.

THEOREM 5.4 (Dual representation). *A set-valued mapping $\Phi\colon \mathbf{U} \rightrightarrows \mathbf{V}$ is partly smooth at a point $\bar{u} \in \mathbf{U}$ for a value $\bar{v} \in \Phi(\bar{u})$ if and only if it has a local representation of the following form: there exist Euclidean spaces $\mathbf{X}$ and $\mathbf{Y}$, maps $P\colon \mathbf{U} \to \mathbf{X}$, smooth around $\bar{u}$ with $P(\bar{u}) = 0$ and $\nabla P(\bar{u})$ surjective, and $Q\colon \mathbf{U} \times \mathbf{V} \to \mathbf{Y}$, smooth around $(\bar{u}, \bar{v})$ with $Q(\bar{u}, \bar{v}) = 0$ and $\nabla_v Q(\bar{u}, \bar{v})$ surjective, such that*

$$\operatorname{gph}\Phi \;=\; \big\{(u,v) \in \mathbf{U} \times \mathbf{V} : P(u) = 0,\ Q(u,v) = 0\big\} \quad \textit{around } (\bar{u}, \bar{v}).$$

The active manifold (around $\bar{u}$) is then the inverse image $P^{-1}(0)$.

Proof. Assuming the given representation, define a map $R\colon \mathbf{U} \times \mathbf{V} \to \mathbf{X} \times \mathbf{Y}$ by $R(u,v) = \big(P(u), Q(u,v)\big)$ for points $u \in U$ and values $v \in V$. Clearly R is smooth around the point $(\bar{u}, \bar{v})$, with $R(\bar{u}, \bar{v}) = (0,0)$. The derivative $\nabla R(\bar{u}, \bar{v})\colon \mathbf{U} \times \mathbf{V} \to \mathbf{X} \times \mathbf{Y}$ is surjective, because for any values $x \in X$ and $y \in Y$ we can first find $r \in \mathbf{U}$ satisfying $\nabla P(\bar{u})r = x$, and then find $s \in \mathbf{V}$ satisfying $\nabla_v Q(\bar{u}, \bar{v})s = y - \nabla_u Q(\bar{u}, \bar{v})r$, and in that case we have

$$\nabla R(\bar{u}, \bar{v})(r,s) \;=\; \big(\nabla P(\bar{u})r, \nabla_u Q(\bar{u}, \bar{v})r + \nabla_v Q(\bar{u}, \bar{v})s\big) \;=\; (x,y).$$

Since $\operatorname{gph}\Phi = R^{-1}(0,0)$ around the point $(\bar{u}, \bar{v})$, we deduce that the graph of Φ is a manifold around $(\bar{u}, \bar{v})$.

For points $(u,v) \in \operatorname{gph}\Phi$ near the point $(\bar{u}, \bar{v})$, we have

$$\begin{aligned} T_{\operatorname{gph}\Phi}(u,v) &= \operatorname{Null}\big(\nabla R(u,v)\big) \\ &= \big\{(r,s) \in \mathbf{U} \times \mathbf{V} : \nabla P(u)r = 0,\ \nabla_u Q(u,v)r + \nabla_v Q(u,v)s = 0\big\}, \end{aligned}$$

so, since the partial derivative $\nabla_v Q(u,v)$ is surjective, we deduce

$$\operatorname{proj}\big(T_{\operatorname{gph}\Phi}(u,v)\big) \;=\; \operatorname{Null}\big(\nabla P(u)\big).$$

Since the derivative $\nabla P(u)$ is surjective, this space has constant dimension for u near $\bar{u}$, namely $\dim \mathbf{U} - \dim \mathbf{X}$, so partial smoothness follows.

Clearly the active manifold is contained in the inverse image $P^{-1}(0)$ around $\bar{u}$. We claim these sets in fact agree around $\bar{u}$. If not, there exists a sequence of points $u_k \to \bar{u}$ in $P^{-1}(0)$ lying outside the active manifold. By the implicit function theorem, since the derivative $\nabla_v Q(\bar{u}, \bar{v})$ is surjective, there exists a sequence of values $v_k \to \bar{v}$ such that $Q(u_k, v_k) = 0$ and hence $v_k \in \Phi(u_k)$ for all large k. But this contradicts the definition of the active manifold.

Conversely, suppose the mapping Φ is partly smooth at the point $\bar{u} \in \mathbf{U}$ for the value $\bar{v} \in \Phi(\bar{u})$. Using Theorem 5.1 (Coordinate representation), there exists a Euclidean space $\mathbf{W}$ and a map $H\colon \mathbf{W} \to \mathbf{U}$, smooth around 0 with $H(0) = \bar{u}$ and derivative $\nabla H(0)$ injective, such that the active manifold is $\mathcal{M} = H\big(B_\delta(0)\big)$ around $\bar{u}$ providing $\delta > 0$ is sufficiently small.

Consider the map $G\colon \mathbf{U} \to \mathbf{W}$ discussed in section 2, satisfying the property (2.1), so its restriction $G|_{\mathcal{M}}$ is the inverse of the diffeomorphism H around the point $\bar{u}$. Since $\operatorname{gph}\Phi$ is a manifold and contained in $\mathcal{M} \times \mathbf{V}$ around the point $(\bar{u}, \bar{v})$, the set

$$\Lambda \;=\; \big\{\big(G(u), v\big) : (u,v) \in \operatorname{gph}\Phi,\ u \in B_\delta(\bar{u}),\ v \in B_\delta(\bar{v})\big\}$$

is a manifold around the point $(0, \bar{v}) \in \mathbf{W} \times \mathbf{V}$. Hence $\Lambda = S^{-1}(0)$ around $(0, \bar{v})$, for some map $S\colon \mathbf{W} \times \mathbf{V} \to \mathbf{Y}$ (a Euclidean space), smooth around the point $(0, \bar{v})$ with $S(0, \bar{v}) = 0$ and $\nabla S(0, \bar{v})$ surjective. Equivalently, we have

$$\operatorname{gph}\Phi \;=\; \big\{\big(H(w), v\big) : S(w,v) = 0,\ w \in B_\delta(0),\ v \in B_\delta(\bar{v})\big\} \;\text{ around } (\bar{u}, \bar{v}).$$

We claim, more precisely, that the partial derivative $\nabla_v S(0,\bar{v})\colon \mathbf{V} \to \mathbf{Y}$ is surjective. If not, there exists a nonzero vector $y \in \mathbf{Y}$ such that $\nabla_v S(0,\bar{v})^* y = 0$. By Corollary 5.2 (Minimal identifiability), there exists a function $F\colon \mathbf{W} \to \mathbf{V}$, smooth around 0, such that $F(0) = \bar{v}$ and $F(w) \in \Phi\big(H(w)\big)$ for all small vectors $w \in \mathbf{W}$. We deduce $S\big(w, F(w)\big) = 0$ for all small $w \in \mathbf{W}$, so

$$\nabla_w S(0,\bar{v}) + \nabla_v S(0,\bar{v}) \nabla F(0) = 0.$$

Taking adjoints shows $\nabla_w S(0,\bar{v})^* y = 0$, so in fact $\nabla S(0,\bar{v})^* y = 0$, contradicting the surjectivity of $\nabla S(0,\bar{v})$.

There exists a Euclidean space $\mathbf{X}$ and a map $P\colon \mathbf{U} \to \mathbf{X}$, smooth around the point $\bar{u}$, with $P(\bar{u}) = 0$ and $\nabla P(\bar{u})$ surjective, such that the active manifold is $\mathcal{M} = P^{-1}(0)$ around $\bar{u}$. Furthermore, if we define a map $Q\colon \mathbf{U} \times \mathbf{V} \to \mathbf{Y}$ by $Q(u,v) = S\big(G(u), v\big)$, then the desired representation now follows, since the partial derivative

$$\nabla_v Q(\bar{u},\bar{v}) = \nabla_v S(0,\bar{v})$$

is surjective. ∎

6. The normal bundle and partial smoothness. The canonical example of a partly smooth mapping is the normal space operator associated with a manifold. More precisely, given a manifold $\mathcal{M} \subset \mathbf{U}$ around a point $\bar{u} \in \mathcal{M}$, we can consider the normal space as a set-valued mapping $N_{\mathcal{M}}\colon \mathbf{U} \rightrightarrows \mathbf{U}$, where we define $N_{\mathcal{M}}(u) = \emptyset$ if $u \notin \mathcal{M}$.

THEOREM 6.1 (Normal space). *If the set $\mathcal{M} \subset \mathbf{U}$ is a $C^{(2)}$-smooth manifold around a point $\bar{u} \in \mathcal{M}$, then the normal space mapping $N_{\mathcal{M}}\colon \mathbf{U} \rightrightarrows \mathbf{U}$ is partly smooth at $\bar{u}$ for any value $\bar{v} \in N_{\mathcal{M}}(\bar{u})$, with dimension* $\dim \mathbf{U}$ *and active manifold $\mathcal{M}$.*

Proof. We apply Theorem 5.1 (Coordinate representation). Following the notation of Section 2, there exists a vector $\bar{x} \in \mathbf{X}$ satisfying $\nabla P(\bar{u})^* \bar{x} = \bar{v}$. We claim

$$\operatorname{gph} N_{\mathcal{M}} \;=\; \Big\{\Big(H(w), \nabla P\big(H(w)\big)^* x\Big) : w \in B_\delta(0),\ x \in B_\delta(\bar{x})\Big\}, \quad \text{around } (\bar{u},\bar{v}),$$

providing $\delta > 0$ is sufficiently small. The inclusion "$\supset$" is clear, so it suffices to prove the inclusion "$\subset$".

For sufficiently small $\delta > 0$, the map H gives a diffeomorphism between the open ball $B_\delta(0) \subset \mathbf{W}$ and an open neighborhood of the point $\bar{u}$ in the manifold $\mathcal{M}$. For such δ, if the desired inclusion fails, then there exists a sequence of points $u_r \to \bar{u}$ in $\mathcal{M}$ and a sequence of normals $v_r \in N_{\mathcal{M}}(u_r)$ approaching $\bar{v}$, such that the sequence (u_r, v_r) is disjoint from the right-hand side. There must therefore exist a sequence of vectors $w_r \to 0$ in $\mathbf{W}$ satisfying $H(w_r) = u_r$, and a sequence of vectors $x_r \in \mathbf{X}$ satisfying

$$\nabla P(u_r)^* x_r = v_r \to \bar{v} = \nabla P(\bar{u})^* \bar{x}.$$

Since the linear map $\nabla P(\bar{u})$ is surjective, we can represent it with respect to some orthonormal bases by the matrix $(A\ 0)$, where the matrix A is invertible. Denote the corresponding representation of $\nabla P(u_r)$ by $(A_r\ C_r)$, where $A_r \to A$ and $C_r \to 0$. The property above ensures $A_r^T x_r \to A^T \bar{x}$ and hence $x_r \to \bar{x}$, contradicting our assumption that $x_r \notin B_\delta(\bar{x})$.

Now define a map $G\colon \mathbf{W} \times \mathbf{X} \to \mathbf{U}$ by

$$G(w,z) \;=\; \nabla P\big(H(w)\big)^* (\bar{x} + z) \quad (\text{for } w \in \mathbf{W},\ z \in \mathbf{X}).$$

Clearly G is smooth around the point $(0,0)$, with $G(0,0)=\bar{v}$. Furthermore, around the point $(\bar{u},\bar{v})$, the graph of Φ has the representation (5.2), as we have just seen. It remains to verify the regularity condition (5.1). By assumption, $\mathrm{Null}\big(\nabla H(0)\big)=\{0\}$, so we just need to check that vectors $z\in\mathbf{X}$ satisfy the property

$$\nabla G(0,0)(0,z)=0 \;\Rightarrow\; z=0.$$

However, $\nabla G(0,0)(0,z)=\nabla P(\bar{u})^*z$, and $\nabla P(\bar{u})$ is surjective. Notice that the dimension of $N_{\mathcal{M}}$ is

$$\dim\mathbf{W}+\dim\mathbf{X}=\dim T_{\mathcal{M}}(\bar{u})+\dim N_{\mathcal{M}}(\bar{u})=\dim\mathbf{U},$$

so the result now follows. ◻

We can generalize this result substantially. In the variational analysis that follows, we follow the terminology and notation of [35]. The original definition of a partly smooth set appeared in [26]. Here we use a slightly modified directional version [12].

DEFINITION 6.2. Consider a closed set $Q\subset\mathbf{U}$, a point $\bar{u}\in Q$, and a normal vector $\bar{v}\in N_Q(\bar{u})$. We call Q *partly smooth at $\bar{u}$ for $\bar{v}$ with respect to* a set $\mathcal{M}\subset Q$ when all of the following properties hold.

- Q is prox-regular at $\bar{u}$ for $\bar{v}$.
- $\mathcal{M}$ is a $C^{(2)}$-smooth manifold around $\bar{u}$.
- The regular normal cone $\hat{N}_Q(\bar{u})$ spans the normal space $N_{\mathcal{M}}(\bar{u})$.
- For some neighborhood W of $\bar{v}$, the mapping $u\mapsto N_Q(u)\cap W$ is inner semicontinuous at $\bar{u}$ relative to $\mathcal{M}$.

Since this definition is rather technical, a more concrete model is helpful. Consider the *fully amenable* case [35] when the set Q coincides around $\bar{u}$ with an inverse image $F^{-1}(D)$ where F is a $C^{(2)}$-smooth mapping and D is a closed convex set satisfying $N_D\big(F(\bar{u})\big)\cap N\big(\nabla F(\bar{u})\big)=\{0\}$ (as holds in particular if Q is closed and convex). Then the prox-regularity condition holds, and the normal and regular normal cones, $N_Q(\bar{u})$ and $\hat{N}_Q(\bar{u})$, coincide. The inner semicontinuity condition means that, for any normal vector $v\in N_Q(\bar{u})$ near $\bar{v}$, and any sequence of points $u_r\to\bar{u}$ in $\mathcal{M}$, there exists a corresponding sequence of normals $v_r\in N_Q(u_r)$ approaching $\bar{v}$.

We then have the following result.

THEOREM 6.3 (Partly smooth sets). *Consider a closed set $Q\subset\mathbf{U}$, a point $\bar{u}\in Q$, a regular normal vector $\bar{v}\in\hat{N}_Q(\bar{u})$, and suppose that $\mathcal{M}\subset Q$ is a $C^{(2)}$-smooth manifold around $\bar{u}$. Then the following properties are equivalent for the normal cone mapping N_Q.*

(i) *N_Q is partly smooth at $\bar{u}$ for $\bar{v}$, with active manifold $\mathcal{M}$.*
(ii) *$\mathcal{M}$ is identifiable for N_Q at $\bar{u}$ for $\bar{v}$.*
(iii) *Q is partly smooth at $\bar{u}$ for $\bar{v}$ with respect to $\mathcal{M}$, and $\bar{v}\in\mathrm{ri}\,\hat{N}_Q(\bar{u})$.*
(iv) *$\mathrm{gph}\,N_Q=\mathrm{gph}\,N_{\mathcal{M}}$ around $(\bar{u},\bar{v})$.*

When these properties hold, the dimension of $N_{\mathcal{M}}$ at $\bar{u}$ for $\bar{v}$ is just $\dim\mathbf{U}$.

Proof. The implication (i) $\Rightarrow$ (ii) follows from Proposition 3.4. The equivalence of the properties (ii), (iii), and (iv) follows from [12, Proposition 8.4]. The implication (iv) $\Rightarrow$ (i) follows from Theorem 6.1. ◻

The definition of a partly smooth function parallels that for sets. Again we use a directional version of the original idea in [26], following [13].

Definition 6.4. Consider a closed function $f\colon \mathbf{U} \to \overline{\mathbf{R}}$, a point $\bar{u} \in \mathbf{U}$, and a subgradient $\bar{v} \in \partial f(\bar{u})$. We call f *partly smooth at* $\bar{u}$ *for* $\bar{v}$ *with respect to* a set $\mathcal{M} \subset \mathbf{U}$ when all of the following properties hold.

- f is prox-regular at $\bar{u}$ for $\bar{v}$.
- The restriction $f|_{\mathcal{M}}$ is $C^{(2)}$-smooth around $\bar{u}$.
- The regular subdifferential $\hat{\partial} f(\bar{u})$ is a translate of the normal space $N_{\mathcal{M}}(\bar{u})$.
- For some neighborhood W of $\bar{v}$, the mapping $u \mapsto \partial f(u) \cap W$ is inner semicontinuous at $\bar{u}$ relative to $\mathcal{M}$.

We then have the following result.

Theorem 6.5 (Partly smooth functions). *Consider a closed function $f\colon \mathbf{U} \to \overline{\mathbf{R}}$, a point $\bar{u} \in \mathbf{U}$, and a regular subgradient $\bar{v} \in \hat{\partial} f(\bar{u})$. Suppose that f is subdifferentially continuous at $\bar{u}$ for $\bar{v}$. Suppose furthermore that $\mathcal{M} \subset Q$ is a $C^{(2)}$-smooth manifold around $\bar{u}$, and that the restriction $f|_{\mathcal{M}}$ is $C^{(2)}$-smooth around $\bar{u}$. Then there exists a function $\bar{f}\colon \mathbf{U} \to \mathbf{R}$ that is both $C^{(2)}$-smooth and satisfies $f|_{\mathcal{M}} = \bar{f}|_{\mathcal{M}}$ around $\bar{u}$, and for any such function the following properties are equivalent for the subdifferential mapping ∂f.*

(i) *The mapping ∂f is partly smooth at $\bar{u}$ for $\bar{v}$, with active manifold $\mathcal{M}$.*
(ii) *The manifold $\mathcal{M}$ is identifiable for ∂f at $\bar{u}$ for $\bar{v}$.*
(iii) *The function f is partly smooth at $\bar{u}$ for $\bar{v}$ with respect to $\mathcal{M}$, and $\bar{v} \in \operatorname{ri} \hat{\partial} f(\bar{u})$.*
(iv) *Around $(\bar{u}, \bar{v})$ we have*

$$\operatorname{gph} \partial f \;=\; \{(u, \nabla \bar{f}(u) + v) : u \in \mathcal{M},\ v \in N_{\mathcal{M}}(u)\}.$$

When these properties hold, the dimension of ∂f at $\bar{u}$ for $\bar{v}$ is just $\dim \mathbf{U}$.

Proof. The existence of the function $\bar{f}$ is just the definition smoothness of $f|_{\mathcal{M}}$. The implication (i) $\Rightarrow$ (ii) follows from Proposition 3.4. The equivalence of the properties (ii), (iii), and (iv) follows from [13, Proposition 10.12]. The implication (iv) $\Rightarrow$ (i) follows from Theorem 6.1 and Corollary 5.3 (Sum rule). ∎

Again, the assumptions are rather technical, so we illustrate with a more concrete model, sufficient to cover many objective functions in practice. Consider the *fully amenable* case when the function f is finite at $\bar{u}$ and agrees around $\bar{u}$ with a composite function $g \circ F$, where the mapping F is $C^{(2)}$-smooth around $\bar{u}$ and the function g is lower semicontinuous and convex, satisfying

$$N_{\mathrm{cl}(\mathrm{dom}\, g)}\big(F(\bar{u})\big) \cap N\big(\nabla F(\bar{u})\big) = \{0\}.$$

(When F is simply the identity mapping, we recover the case when f is lower semicontinuous and convex). Then both the subdifferential continuity and prox-regularity condition holds, and the normal and regular subdifferentials, $\partial f(\bar{u})$ and $\hat{\partial} f(\bar{u})$, coincide.

7. Identifiability for primal-dual splitting. As well as focusing attention on the fundamental feature underlying partly smooth geometry—the constant rank property—the formalism we have developed in this work extends beyond mere subdifferential mappings. We illustrate with a recent popular primal-dual setting.

We consider the saddlepoint problem

$$\inf_{x \in \mathbf{X}} \sup_{y \in \mathbf{Y}} \{(f+p)(x) + \langle Ax, y\rangle - (g+q)(y)\}$$

for Euclidean spaces $\mathbf{X}$ and $\mathbf{Y}$, lower semicontinuous convex functions $f\colon \mathbf{X} \to \overline{\mathbf{R}}$ and $g\colon \mathbf{Y} \to \overline{\mathbf{R}}$, $C^{(2)}$-smooth convex functions $p\colon \mathbf{X} \to \mathbf{R}$ and $q\colon \mathbf{Y} \to \mathbf{R}$, and a linear map $A\colon \mathbf{X} \to \mathbf{Y}$. Saddlepoints (x, y) satisfy the inclusion

$$(0, 0) \in \Phi(x, y)$$

where the set-valued mapping $\Phi\colon \mathbf{X} \times \mathbf{Y} \rightrightarrows \mathbf{X} \times \mathbf{Y}$ is defined by

$$\Phi(x, y) = \big(\partial f(x) + \nabla p(x) + A^* y\big) \times \big(- Ax + \partial g(y) + \nabla q(y)\big).$$

The following method (following [30]) covers a variety of primal-dual algorithms [5, 36, 6, 20]. As usual, we denote by $\mathrm{prox}_f(x)$ the unique minimizer of the function $f(\cdot) + \frac{1}{2}\|\cdot - x\|^2$.

Algorithm 7.1 Primal-dual splitting

Choose $\gamma, \mu > 0$. For $k = 0$, $x_0 \in \mathbf{X}$, $y_0 \in \mathbf{Y}$,
while not done **do**
  $x_{k+1} = \mathrm{prox}_{\gamma f}\big(x_k - \gamma \nabla p(x_k) - \gamma A^* y_k\big)$,
  $y_{k+1} = \mathrm{prox}_{\mu g}\big(y_k - \mu \nabla q(y_k) + \mu A(2x_{k+1} - x_k)\big)$,
  $k = k + 1$;
end while

Assuming suitable conditions [30, Theorem 3.3], there exists a saddlepoint $(\bar{x}, \bar{y})$ satisfying

$$(7.1) \qquad (x_k, y_k) \to (\bar{x}, \bar{y}) \quad \text{and} \quad \mathrm{dist}\big((0, 0), \Phi(x_k, y_k)\big) \to 0.$$

Assume furthermore, again following [30], that the function f is partly smooth at $\bar{x}$ for $-\nabla p(\bar{x}) - A^* \bar{y}$ with respect to some set $\mathcal{M} \subset \mathbf{X}$, that the function g is partly smooth at $\bar{y}$ for $-\nabla q(\bar{y}) + A\bar{x}$ with respect to some set $\mathcal{N} \subset \mathbf{Y}$, and that the nondegeneracy conditions

$$-\nabla p(\bar{x}) - A^* \bar{y} \in \mathrm{ri}\, \partial f(\bar{x}) \quad \text{and} \quad - \nabla q(\bar{y}) + A\bar{x} \in \mathrm{ri}\, \partial g(\bar{y})$$

hold. Theorem 6.5 implies that the mapping ∂f is partly smooth at $\bar{x}$ for $-\nabla p(\bar{x}) - A^* \bar{y}$ with respect to $\mathcal{M}$, and the mapping ∂g is partly smooth at $\bar{y}$ for $-\nabla q(\bar{y}) + A\bar{x}$ with respect to $\mathcal{N}$. It follows immediately that the set-valued mapping $(x, y) \mapsto \partial f(x) \times \partial g(y)$ is partly smooth at $(\bar{x}, \bar{y})$ for $(-\nabla p(\bar{x}) - A^* \bar{y}, -\nabla q(\bar{y}) + A\bar{x})$ with respect to $\mathcal{M} \times \mathcal{N}$ and hence by the sum rule that the set-valued mapping Φ is partly smooth at $(\bar{x}, \bar{y})$ for $(0, 0)$ with respect to $\mathcal{M} \times \mathcal{N}$. By Proposition 3.4, $\mathcal{M} \times \mathcal{N}$ is identifiable for Φ at $(\bar{x}, \bar{y})$ for $(0, 0)$, so the convergence property (7.1) implies $x_k \in \mathcal{M}$ and $y_k \in \mathcal{N}$ eventually: exactly the conclusion of [30, Theorem 3.3].

8. Example: smooth optimization on a manifold.

We end with a brief but representative example to illustrate the interplay between partial smoothness and the second-order sufficient conditions. This interplay sheds new light on local algorithms of active-set type, as discussed further in [28].

Suppose $\mathcal{M} \subset \mathbf{U}$ is a $C^{(2)}$-smooth manifold around a point $\bar{u} \in \mathcal{M}$, and $f\colon \mathcal{M} \to \mathbf{R}$ is a $C^{(2)}$-smooth function. We can consider a corresponding extended-valued function $\tilde{f}\colon \mathbf{U} \to \overline{\mathbf{R}}$ defined by

$$\tilde{f}(u) = \begin{cases} f(u) & (u \in \mathcal{M}) \\ +\infty & (u \notin \mathcal{M}). \end{cases}$$

Its subdifferential map is given by

$$\partial\tilde{f}(u) = \begin{cases} \nabla_{\mathcal{M}} f(u) + N_{\mathcal{M}}(u) & (u \in \mathcal{M}) \\ \emptyset & (u \notin \mathcal{M}), \end{cases}$$

where $\nabla_{\mathcal{M}} f(u) \in T_{\mathcal{M}}(u)$ denotes the covariant derivative (as discussed in [29]). By Corollary 5.3 (Sum rule), this set-valued mapping $\partial\tilde{f}$ is partly smooth at $\bar{u}$ for any value in the set $\nabla_{\mathcal{M}} f(\bar{u}) + N_{\mathcal{M}}(\bar{u})$. In particular, assuming the first-order necessary condition

$$\nabla_{\mathcal{M}} f(\bar{u}) = 0,$$

then $\partial\tilde{f}$ is partly smooth at $\bar{u}$ for 0, with dimension $\dim \mathbf{U}$ and active manifold $\mathcal{M}$.

Now suppose further that $\bar{u}$ is a local minimizer around which f grows quadratically: for some $\delta > 0$,

$$f(u) \geq f(\bar{u}) + \delta |u - \bar{u}|^2 \quad \text{for all } u \in \mathcal{M} \text{ near } \bar{u}.$$

Equivalently, in addition to the first-order condition, f satisfies the second-order sufficient condition: the covariant Hessian $\nabla^2_{\mathcal{M}} f(u) \colon T_{\mathcal{M}}(u) \to T_{\mathcal{M}}(u)$ (a self-adjoint linear map) is positive definite when $u = \bar{u}$. We also have (from [29]):

$$N_{\operatorname{gph} \partial\tilde{f}}(\bar{u}, 0) \; = \; \big\{(z, w) : w \in T_{\mathcal{M}}(\bar{u}) \text{ and } z + \nabla^2_{\mathcal{M}} f(\bar{u}) w \in N_{\mathcal{M}}(\bar{u})\big\}.$$

Hence $\operatorname{gph} \partial\tilde{f}$ intersects the subspace $\mathbf{U} \times \{0\}$ transversally at $(\bar{u}, 0)$. To see this, note

$$(z, w) \in N_{\operatorname{gph} \partial\tilde{f}}(\bar{u}, 0) \cap N_{\mathbf{U} \times \{0\}}(\bar{u}, 0)$$

if and only if

$$w \in T_{\mathcal{M}}(\bar{u}), \quad z + \nabla^2_{\mathcal{M}} f(\bar{u}) w \in N_{\mathcal{M}}(\bar{u}), \quad z = 0.$$

Since $\nabla^2_{\mathcal{M}} f(\bar{u})$ is positive definite, the latter property holds if and only if $z = 0$ and $w = 0$, as required. Consequently, $(\bar{u}, 0)$ is an isolated transversal point of intersection of the two manifolds $\operatorname{gph} \partial\tilde{f}$ and $\mathbf{U} \times \{0\}$.

To summarize, satisfying the first-order optimality conditions for minimizing the smooth function f on the manifold $\mathcal{M} \subset \mathbf{U}$ amounts to finding a point in the intersection of the space $\mathbf{U} \times \{0\}$ and the graph of the subdifferential of the corresponding extended-valued function $\tilde{f}$. Assuming the second-order sufficient conditions, the subdifferential is a partly smooth mapping of dimension $\dim \mathbf{U}$, and its graph (which is locally a manifold) intersects the subspace $\mathbf{U} \times \{0\}$ transversally at an isolated point. This simple geometry underlies many standard local algorithms of active-set type [28].

Acknowledgments. Thanks to Artur Gorokh for many helpful comments during the early development of these results, and to an anonymous referee for several suggestions, leading in particular to the discussion in section 4.

REFERENCES

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

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

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

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

[5] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.
[6] L. Condat, *A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, J. Optim. Theory Appl., 158 (2013), pp. 460–479.
[7] M. Coste, *An Introduction to Semialgebraic Geometry*, RAAG Notes, 78 pages, Institut de Recherche Mathématiques de Rennes, October 2002.
[8] A. Dontchev and R. Rockafellar, *Implicit Functions and Solution Mappings*, Springer-Verlag, Berlin, 2009.
[9] D. Drusvyatskiy, A. Ioffe, and A. Lewis, *The dimension of semialgebraic subdifferential graphs*, Nonlinear Anal., 75 (2012), pp. 1231–1245.
[10] D. Drusvyatskiy, A. Ioffe, and A. Lewis, *Generic minimizing behavior in semi-algebraic optimization*, SIAM J. Optimization, 26 (2016), pp. 513–534.
[11] D. Drusvyatskiy and A. Lewis, *Semi-algebraic functions have small subdifferentials*, Math. Program., 140 (2013), pp. 5–29.
[12] D. Drusvyatskiy and A. Lewis, *Optimality, identifiability, and sensitivity*, Math. Program., 147 (2014), pp. 467–498.
[13] D. Drusvyatskiy and A. Lewis, *Optimality, identifiability, and sensitivity (extended version).* arXiv:1207.6628, 2014.
[14] J. Dunn, *On the convergence of projected gradient processes to singular critical points*, J. Optim. Theory Appl., 55 (1987), pp. 203–216.
[15] F. Facchinei and J.-S. Pang, *Finite Dimensional Variational Inequalities and Complementarity Problems*, Springer-Verlag, New York, 2003.
[16] M. Ferris, *Finite termination of the proximal point algorithm*, Math. Program. Ser. A, 50 (1991), pp. 359–366.
[17] S. Flåm, *On finite convergence and constraint identification of subgradient projection methods*, Math. Program., 57 (1992), pp. 427–437.
[18] R. Hardt, *Semi-algebraic local-triviality in semi-algebraic mappings*, American Journal of Mathematics, 102 (1980), pp. 291–302.
[19] W. Hare and C. Sagastizábal, *Computing proximal points of nonconvex functions*, Math. Program., 116 (2009), pp. 221–258.
[20] B. He and X. Yuan, *Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective*, SIAM J. Imaging Sci., 5 (2012), pp. 119–149.
[21] A. Ioffe, *A Sard theorem for tame set-valued mappings*, J. Math. Anal. Appl., 335 (2007), pp. 882–901.
[22] J. Lee, *Introduction to Smooth Manifolds*, Springer, New York, 2003.
[23] J. H. Lee, G. M. Lee, and T.-S. Pham, *Genericity and Holder stability in semi-algebraic variational inequalities*, Journal of Optimation Theory and Applications, 178 (2018), pp. 56–77.
[24] C. Lemaréchal, F. Oustry, and C. Sagastizábal, *The U-lagrangian of a convex function*, Transactions of the American Mathematical Society, 352 (2000), pp. 711–729.
[25] C. Lemaréchal and C. Sagastizábal, *Practical aspects of the Moreau-Yosida regularization: theoretical preliminaries*, SIAM J. Optim., 7 (1997), pp. 367–385.
[26] A. Lewis, *Active sets, nonsmoothness, and sensitivity*, SIAM J. Optim., 13 (2003), pp. 702–725.
[27] A. Lewis, *Nonsmooth optimization: conditioning, convergence and semi-algebraic models*, in Proceedings of the International Congress of Mathematicians, Seoul, 2014, pp. 871–895.
[28] A. Lewis and C. Wylie, *Active-set Newton methods and partial smoothness*, Mathematics of Operations Research, 46 (2020), pp. 712–725.
[29] A. Lewis and S. Zhang, *Partial smoothness, tilt stability, and generalized Hessians*, SIAM J. Optim., 23 (2013), pp. 74–94.
[30] J. Liang, J. Fadili, and G. Peyré, *Local linear convergence analysis of primal-dual splitting methods*, Optimization, 67 (2018), pp. 821–853.
[31] R. Mifflin and C. Sagastizábal, *Proximal points are on the fast track*, Journal of Convex Analysis, 9 (2002), pp. 563—579.
[32] R. Mifflin and C. Sagastizábal, *Primal-dual gradient structured functions: second-order results; links to epi-derivatives and partly smooth functions*, SIAM J. Optim., 13 (2003), pp. 1174–1194.
[33] R. Mifflin and C. Sagastizábal, *$\mathcal{VU}$-smoothness and proximal point results for some nonconvex functions*, Optim. Methods Softw., 19 (2004), pp. 463–478.
[34] R. Mifflin and C. Sagastizábal, *A $\mathcal{VU}$-algorithm for convex minimization*, Math. Program., 104 (2005), pp. 583–608.
[35] R. Rockafellar and R.-B. Wets, *Variational Analysis*, Grundlehren der mathematischen Wissenschaften, Vol 317, Springer, Berlin, 1998.
[36] B. C. Vũ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*,

Adv. Comput. Math., 38 (2013), pp. 667–681.
[37] S. Wright, *Identifiable surfaces in constrained optimization*, SIAM J. Control Optim., 31 (1993), pp. 1063–1079.