Partial smoothness and constant rank

A.S. Lewis[*] Jingwei Liang[†]

July 10, 2018

Abstract

The idea of partial smoothness in optimization blends certain smooth and nonsmooth properties of feasible regions and objective functions. As a consequence, the standard first-order conditions guarantee that diverse iterative algorithms (and post-optimality analyses) identify active structure or constraints. However, by instead focusing directly on the first-order conditions, the formal concept of partial smoothness simplifies dramatically: in basic differential geometric language, it is just a constant-rank condition. In this view, partial smoothness extends to more general mappings, such as saddlepoint operators underlying primal-dual splitting algorithms.

Key words: partial smoothness, active set identification, nonsmooth optimization, subdifferential, primal-dual splitting

AMS 2000 Subject Classification: ...

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, exhibit "identification" properties. Iterates in the underlying Euclidean space $\mathbf{U}$ converging to an optimal solution $\bar{u}$ eventually lie in a subset $M \subset \mathbf{U}$ capturing the structure of the optimal solution. In traditional nonlinear programming, M might be the "identifiable surface" [27] of the feasible region defined by the constraints active at optimality; in machine learning applications, M might consist of vectors with a certain sparsity pattern [18].

[*]ORIE, Cornell University, Ithaca, NY 14853, U.S.A. people.orie.cornell.edu/aslewis.
Research supported in part by National Science Foundation Grant DMS-1208338 and by the US-Israel Binational Science Foundation Grant 2008261.
[†]DAMTP, University of Cambridge, U.K. jl993@cam.ac.uk

A simple but quite extensive model of this phenomenon, following the philosophy of [27], 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 M of interest is a smooth surface, or more precisely a manifold around $\bar{u}$, meaning that locally it consists of solutions of a system of $C^{(2)}$-smooth equations with linearly independent gradients. Identification amounts to the property

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

where ∂f denotes the classical subdifferential operator. Earlier versions of this identifiability idea include [1–4, 9–11].

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 M. This property combines smoothness conditions on f when restricted to M with a sharpness property of f in directions normal to M. More precisely, around the point $\bar{u}$ the restrictions of the function f and its subdifferential ∂f to the manifold 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 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 [19, Thm 4.10].

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 [17]. A closely related thread of research, known as "$\mathcal{V}U$ theory", originated with [16], and includes [12, 15, 21–24]. 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 [27] 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, the notion of partial smoothness, and the closely related idea of identifiability, are in essence constant-rank properties. This 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, we consider the saddlepoint optimality conditions associated with primal-dual splitting methods like the Chambolle-Pock algorithm [5].

2 Manifolds

We begin a more formal development by summarizing some elementary ideas about manifolds. Given a Euclidean space $\mathbf{U}$, we consider a set $M \subset \mathbf{U}$ that has the structure of a smooth manifold locally, around a point $\bar{u} \in M$. By "smooth", we mean $C^{(1)}$-smooth, unless we state otherwise. We can consider such sets 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, 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 $\delta > 0$,

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

More formally [14, Chapter 8], some open neighborhood of $\bar{u}$ in 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 $H(w) \in M$. The *tangent space* at such a point is given simply by

$$T_M\big(H(w)\big) = \mathrm{Range}\big(\nabla H(w)\big).$$

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

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

Given another Euclidean space $\mathbf{V}$, a map $F \colon M \to \mathbf{V}$ 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 M. In that case, the *rank* of F at $\bar{u}$ is $\dim\big(\nabla G(\bar{u}) T_M(\bar{u})\big)$. 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 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

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

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

Adopting a dual approach, we can equivalently define a set $M \subset \mathbf{U}$ 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

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

Then the tangent and normal spaces are given by

$$\begin{aligned} T_M(u) &= \mathrm{Null}(\nabla P(u)) \\ N_M(u) &= \mathrm{Range}(\nabla P(u)^*) \end{aligned}$$

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

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

$$\mathbf{U} = T_M(\bar{u}) \oplus N_M(\bar{u}).$$

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

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

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

3 Partly smooth mappings

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 regular" at $\bar{u}$ for $\bar{v}$.

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, 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 $M \subset \mathbf{U}$, uniquely defined around $\bar{u}$, that is a smooth manifold around $\bar{u}$, and satisfies*

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

for all small $\epsilon > 0$. We call any such set 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, the resulting image

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

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 $M_{\epsilon'} \subset M_\epsilon$, but these sets are manifolds around $\bar{u}$ of the same dimension, so must be identical around $\bar{u}$. $\square$

We use the following definition [7].

Definition 3.3 A set $M \subset \mathbf{U}$ 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 $\mathrm{gph}\,\Phi \subset M \times \mathbf{V}$ 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 shortly, the active manifold is a locally minimal identifiable set.

4 Representations of partly smooth mappings

The following result gives a representation of a partly smooth mapping using local coordinates.

Theorem 4.1 (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,*

$$(4.2)\quad w \in \mathbf{W},\ z \in \mathbf{Z},\ \nabla H(0)w = 0 \text{ and } \nabla G(0,0)(w,z) = 0 \quad \Rightarrow \quad w = 0 \text{ and } z = 0,$$

and for all small $\delta > 0$*,*

$$(4.3)\qquad \operatorname{gph}\Phi \;=\; \big\{\big(H(w), G(w,z)\big) : w \in B_\delta(0),\ z \in B_\delta(0)\big\} \ \text{ around } (\bar{u},\bar{v}).$$

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

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

with

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

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

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

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.

Corollary 4.4 (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 M has the following properties.*

- *There exists a map $F\colon M\to\mathbf{V}$, smooth around $\bar u$, such that $F(\bar u)=\bar v$ and $F(u)\in\Phi(u)$ for all points $u\in M$ near $\bar u$.*
- *For any set $M'\subset\mathbf{U}$ containing $\bar u$, and any map $F'\colon M'\to\mathbf{V}$ that is continuous at ’$\bar u$ and satisfies $F(\bar u)=\bar v$ and $F(u)\in\Phi(u)$ for all points $u\in M'$ near $\bar u$, we must have $M'\subset M$ around $\bar u$.*
- *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 4.1, 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 [7, Proposition 2.8]: for any sequence of points $u_r\to\bar u$ in the active manifold 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 $u_r \in M'$ 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 $u_r \in M$ for all large indices r by Proposition 3.4. □

We also have the following calculus rule.

Corollary 4.5 (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 4.1, 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

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

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 4.1. □

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

Theorem 4.6 (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*

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

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

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 4.1 (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 $M=H\big(B_\delta(0)\big)$ 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 $G|_M$ is the inverse of the diffeomorphism H around the point $\bar u$. Since $\operatorname{gph}\Phi$ is a manifold and contained in $M\times\mathbf{V}$ 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 4.4 (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 $M = P^{-1}(0)$ 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. □

5 The normal bundle and partial smoothness

Given a manifold $M \subset \mathbf{U}$ around a point $\bar{u} \in M$, we can consider the normal space as a set-valued mapping $N_M\colon \mathbf{U} \rightrightarrows \mathbf{U}$, where we define $N_M(u) = \emptyset$ if $u \notin M$.

Theorem 5.1 (Normal space) *If a set $M \subset \mathbf{U}$ is a $C^{(2)}$-smooth manifold around a point $\bar{u} \in M$, then the normal space mapping $N_M\colon \mathbf{U} \rightrightarrows \mathbf{U}$ is partly smooth at $\bar{u}$ for any value $\bar{v} \in N_M(\bar{u})$, with dimension* $\dim \mathbf{U}$ *and active manifold M.*

Proof We apply Theorem 4.1 (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

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

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 M. For such δ, if the desired inclusion fails, then there exists a sequence of points $u_r \to \bar{u}$ in M and a sequence of normals $v_r \in N_M(u_r)$ 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 (4.3), as we have just seen. It remains to verify the regularity condition (4.2). 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_M is

$$\dim \mathbf{W} + \dim \mathbf{X} = \dim T_M(\bar{u}) + \dim N_M(\bar{u}) = \dim \mathbf{U},$$

so the result now follows. □

We can generalize this result substantially. In the variational analysis that follows, we follow the terminology and notation of [25]. The original definition of a partly smooth set appeared in [17]. Here we use a slightly modified directional version [7].

Definition 5.2 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 $M \subset Q$ when all of the following properties hold.

- Q is prox-regular at $\bar{u}$ for $\bar{v}$.
- M is a $C^{(2)}$-smooth manifold around $\bar{u}$.
- $N_M(\bar{u}) = \operatorname{span} \hat{N}_Q(\bar{u})$.
- For some neighborhood W of $\bar{v}$, the mapping $u \mapsto N_Q(u) \cap W$ is inner semicontinuous at $\bar{u}$ relative to M.

Since this definition is rather technical, a more concrete model is helpful. Consider the *fully amenable* case 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 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 5.3 *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 $M \subset Q$ 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 M.*

(ii) *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 M, and $\bar{v} \in \operatorname{ri} \hat{N}_Q(\bar{u})$.*

(iv) $\operatorname{gph} N_Q = \operatorname{gph} N_M$ *around $(\bar{u}, \bar{v})$.*

When these properties hold, the dimension of N_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 [7, Proposition 8.4]. The implication (iv) $\Rightarrow$ (i) follows from Theorem 5.1. □

The definition of a partly smooth function parallels that for sets. Again we use a directional version of the original idea in [17], following [8].

Definition 5.4 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 $M \subset \mathbf{U}$ when all of the following properties hold.

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

We then have the following result.

Theorem 5.5 *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 $M \subset Q$ is a $C^{(2)}$-smooth manifold around $\bar{u}$, and that the restriction $f|_M$ 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 $f|_M = \bar{f}|_M$ 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* M*.*

(ii) *The manifold* 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* M*, and* $\bar{v} \in \operatorname{ri} \hat{\partial} f(\bar{u})$*.*

(iv) *Around* $(\bar{u}, \bar{v})$ *we have*

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

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 $f|_M$. The implication (i) $\Rightarrow$ (ii) follows from Proposition 3.4. The equivalence of the properties (ii), (iii), and (iv) follows from [8, Proposition 10.12]. The implication (iv) $\Rightarrow$ (i) follows from Theorem 5.1 and Corollary 4.5 (Sum rule). □

Again, the assumptions are rather technical, so we illustrate with a more concrete model. 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 $N_{\mathrm{cl(dom\,g)}}\big(F(\bar{u})\big) \cap N\big(\nabla F(\bar{u})\big) \;=\; \{0\}$. (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.

6 Identifiability for primal-dual splitting

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

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

The following method (following [20]) covers a variety of primal-dual algorithms [5, 6, 13, 26]. As usual, we denote by $\operatorname{prox}_f(x)$ the unique minimizer of the function $f(\cdot) + \frac{1}{2}\|\cdot - x\|^2$.

Algorithm 6.1 (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 [20, Theorem 3.3], there exists a saddlepoint $(\bar{x}, \bar{y})$ satisfying

$$(x_k, y_k) \to (\bar{x}, \bar{y}) \qquad \text{and} \qquad \mathrm{dist}\big((0,0), \Phi(x_k, y_k)\big) \to 0. \tag{6.2}$$

Assume furthermore, again following [20], that the function f is partly smooth at $\bar{x}$ for $-\nabla p(\bar{x}) - L^* \bar{y}$ with respect to some set $M \subset \mathbf{X}$, that the function g is partly smooth at $\bar{y}$ for $-\nabla q(\bar{y}) + L\bar{x}$ with respect to some set $N \subset \mathbf{Y}$, and that the nondegeneracy conditions

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

hold. Theorem 5.5 implies that the mapping ∂f is partly smooth at $\bar{x}$ for $-\nabla p(\bar{x}) - L^* \bar{y}$ with respect to M, and the mapping ∂g is partly smooth at $\bar{y}$ for $-\nabla q(\bar{y}) + L\bar{x}$ with respect to 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 $(-\nabla p(\bar{x}) - L^* \bar{y}, -\nabla q(\bar{y}) + L\bar{x})$ with respect to $M \times N$ 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 $M \times N$. By Proposition 3.4, $M \times N$ is identifiable for Φ at $(\bar{x}, \bar{y})$ for $(0, 0)$, so the convergence property (6.2) implies $x_k \in M$ and $y_k \in N$ eventually: exactly the conclusion of [20, Theorem 3.3].

7 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. Suppose $M \subset \mathbf{U}$ is a $C^{(2)}$-smooth manifold around a point $\bar{u} \in M$, and $f \colon M \to \mathbf{R}$ 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

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

Its subdifferential map is given by

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

where $\nabla_M f(u) \in T_M(u)$ denotes the covariant derivative. By Corollary 4.5 (Sum rule), this set-valued mapping $\partial\tilde{f}$ is partly smooth at $\bar{u}$ for any value in the set $\nabla_M f(\bar{u}) + N_M(\bar{u})$. In particular, assuming the first-order necessary condition

$$\nabla_M f(\bar{u}) = 0,$$

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

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

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

Equivalently, in addition to the first-order condition, f satisfies the second-order sufficient condition: the covariant Hessian $\nabla^2_M f(u) \colon T_M(u) \to T_M(u)$ (a self-adjoint linear map) is positive definite when $u = \bar{u}$. We also have (from [19]):

$$N_{\mathrm{gph}\,\partial\tilde{f}}(\bar{u}, 0) \ = \ \big\{(z, w) : w \in T_M(\bar{u}) \text{ and } z + \nabla^2_M f(\bar{u})w \in N_M(\bar{u})\big\}.$$

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

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

if and only if

$$w \in T_M(\bar{u}), \quad z + \nabla^2_M f(\bar{u})w \in N_M(\bar{u}), \quad z = 0.$$

Since $\nabla^2_M f(\bar{u})$ 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 $\mathrm{gph}\,\partial\tilde{f}$ and $\mathbf{U} \times \{0\}$.

To summarize, satisfying the first-order optimality conditions for minimizing the smooth function f on the manifold $M \subset \mathbf{U}$ 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 intersects the subspace $\mathbf{U} \times \{0\}$ transversally at an isolated point.

Acknowledgements

Many thanks to Artur Gorokh for many helpful comments during the development of these results.

References

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

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

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

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

[5] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision*, 40:120–145, 2011.

[6] L. Condat. A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *J. Optim. Theory Appl.*, 158:460–479, 2013.

[7] D. Drusvyatskiy and A.S. Lewis. Optimality, identifiability, and sensitivity. *Math. Program.*, 147(1-2, Ser. A):467–498, 2014.

[8] D. Drusvyatskiy and A.S. Lewis. Optimality, identifiability, and sensitivity. arXiv:1207.6628, 2014.

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

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

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

[12] W. Hare and C. Sagastizábal. Computing proximal points of nonconvex functions. *Math. Program.*, 116(1-2, Ser. B):221–258, 2009.

[13] B. He and X. Yuan. Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective. *SIAM J. Imaging Sci.*, 5:119–149, 2012.

[14] J.M. Lee. *Introduction to Smooth Manifolds.* Springer, New York, 2003.

[15] C. Lemaréchal, F. Oustry, and C. Sagastizábal. The U-lagrangian of a convex function. *Transactions of the American Mathematical Society*, 352:711–729, 2000.

[16] C. Lemaréchal and C. Sagastizábal. Practical aspects of the Moreau-Yosida regularization: theoretical preliminaries. *SIAM J. Optim.*, 7:367—385, 1997.

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

[18] A.S. Lewis and S.J. Wright. A proximal method for composite minimization. *Math. Program.*, 158:501–546, 2016.

[19] A.S. Lewis and S. Zhang. Partial smoothness, tilt stability, and generalized Hessians. *SIAM J. Optim.*, 23(1):74–94, 2013.

[20] J. Liang, J. Fadili, and G. Peyré. Local linear convergence analysis of primal-dual splitting methods. *Optimization*, 2018. DOI 10.1080/02331934.2018.1426584.

[21] R. Mifflin and C. Sagastizábal. Proximal points are on the fast track. *Journal of Convex Analysis*, 9:563—579, 2002.

[22] 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(4):1174–1194 (electronic), 2003.

[23] R. Mifflin and C. Sagastizábal. $\mathcal{VU}$-smoothness and proximal point results for some nonconvex functions. *Optim. Methods Softw.*, 19(5):463–478, 2004.

[24] R. Mifflin and C. Sagastizábal. A $\mathcal{VU}$-algorithm for convex minimization. *Math. Program.*, 104(2-3, Ser. B):583–608, 2005.

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

[26] B. C. Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comput. Math.*, 38:667–681, 2013.

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