Mathematics of Operations Research

Active-Set Newton Methods and Partial Smoothness

Adrian S. Lewis , Calvin Wylie

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.)
and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual
professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to
transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

Active-Set Newton Methods and Partial Smoothness

Adrian S. Lewis,[a] Calvin Wylie[a]

[a] School of Operations Research and Information Engineering, Cornell University, Ithaca, New York 14853
Contact: adrian.lewis@cornell.edu, 🔵 https://orcid.org/0000-0002-1785-1106 (ASL)

Abstract. Diverse optimization algorithms correctly identify, in finite time, intrinsic constraints that must be active at optimality. Analogous behavior extends beyond optimization to systems involving partly smooth operators, and in particular to variational inequalities over partly smooth sets. As in classical nonlinear programming, such active-set structure underlies the design of accelerated local algorithms of Newton type. We formalize this idea in broad generality as a simple linearization scheme for two intersecting manifolds.

1. Introduction

Active set methods for optimization and more general variational problems proceed, either explicitly or implicitly, in two phases. The first phase relies on a globally convergent process to identify activity in the underlying constraints at the solution. After correct identification, the problem simplifies locally: the algorithm transitions to a second phase with accelerated local convergence, typically as a result of linearizing the simplified problem. Standard texts like Nocedal and Wright [40] and Facchinei and Pang [10] include overviews of active set methods for traditional nonlinear programming and variational inequalities, but the algorithmic philosophy also extends to less classical large-scale settings (Lee and Wright [21], Lewis and Wright [26], Liang [28], Liang et al. [30]).

The idea of partial smoothness, introduced in Lewis [24], formalizes the property driving the *first* phase: "identifiability." This term originated in Wright [45], with precursors in Calamai and Moré [5], Dunn [9], Burke and Moré [4], Burke [3], Al-Khayyal and Kyparisis [1], Ferris [13], and Flåm [14]. Applications to convergence and sensitivity analysis for large-scale convex optimization include Liang et al. [29], Liang et al. [32], Liang et al. [31], Vaiter et al. [44], Liang et al. [30], Vaiter et al. [43], and Fadili et al. [12]. Rather than active sets of constraints, partly smooth geometry focuses on the "active manifold" defined by those constraints. The equivalence with identifiability is explored in detail in Drusvyatskiy and Lewis [8].

In contrast with most of this prior literature, in this work we concentrate on the *second* phase of active set methods: in partly smooth settings, how should we understand accelerated local convergence? A formalism parallel to partial smoothness with related algorithmic ambitions, "$\mathcal{VU}$ theory," emerged in Lemaréchal and Sagastizábal [22] and includes Lemaréchal et al. [23], Mifflin and Sagastizábal [33], Mifflin and Sagastizábal [36], Mifflin and Sagastizábal [35], Hare and Sagastizábal [16], Mifflin and Sagastizábal [34], and most recently, the survey Sagastizábal [42]. Algorithms exploiting the active manifold structure, and the resulting connections between partial smoothness, $\mathcal{VU}$ theory, and classical nonlinear programming were explored in Miller and Malick [37]. Our aims here, rooted in the partly smooth framework, are more general, formal, and geometric.

In essence, as explained in Lewis and Liang [25], partial smoothness is a constant rank property. Assuming this simple differential geometric condition, the underlying problem reduces, locally, to finding the unique point in two manifolds that intersect transversally. Just like the classical Newton method for solving equations, via linearization we arrive at fast local algorithms.

A standard illustration (see Lewis and Liang [25]) involves $C^{(2)}$-smooth functions $g_i : \mathbf{R}^n \to \mathbf{R}$, for indices $i = 1, 2, 3 \ldots, m$, and a corresponding set

$$F = \left\{ x : g_i(x) \le 0 \ \forall i \right\}.$$

Given a cost vector $\bar{c} \in \mathbf{R}^n$, the first-order optimality condition for the optimization problem $\sup_{x \in F} \langle \bar{c}, x \rangle$ is a generalized equation involving the normal cone to F:

$$\bar{c} \in N_F(x). \tag{1}$$

Consider any feasible point $x \in F$ for which the constraint gradients $\nabla g_i(x)$ for i in the active set

$$I(x) = \{i : g_i(x) = 0\}$$

are linearly independent: for such points, we can describe the normal cone operator using Lagrange multiplier vectors $\lambda \in \mathbf{R}^m$:

$$N_F(x) = \left\{ \sum_i \lambda_i \nabla g_i(x) : \lambda \geq 0, \ \lambda_i = 0 \ (i \notin I(x)) \right\}.$$

Let us focus on a particular such point, $\bar{x}$, that solves Equation (1), with corresponding multiplier $\bar{\lambda}$. Assuming strict complementarity (meaning $\bar{\lambda}_i > 0$ for all $i \in I(\bar{x})$), the operator N_F is partly smooth at $\bar{x}$ (for $\bar{c}$): around the pair $(\bar{x}, \bar{c})$, its graph, gph N_F, is a manifold, and the canonical projection is constant rank and maps gph N_F locally onto the active manifold

$$\mathcal{M} = \{x : g_i(x) = 0 \ (i \in I(\bar{x}))\}.$$

In particular, for any nearby pair (x, c) in the graph, the point x must lie in $\mathcal{M}$ and so, has the same active set: this is the property of identifiability. For more details in the setting of this standard example, see Lewis and Liang [25].

Now, however, suppose in addition that the given point $\bar{x}$ also satisfies the standard *second-order* sufficient conditions: the Lagrangian $\sum_i \bar{\lambda}_i \nabla g_i - c$ is not only critical at $\bar{x}$, but also, its Hessian is positive definite on the tangent space to $\mathcal{M}$:

$$\{d \in \mathbf{R}^n : \langle \nabla g_i(\bar{x}), d \rangle = 0 \ (i \in I(\bar{x}))\}.$$

In that case, locally around the pair $(\bar{x}, \bar{c})$, the graph of the normal cone operator, gph N_F, and the space $\mathbf{R}^n \times \{\bar{c}\}$ are manifolds intersecting at the unique point $(\bar{x}, \bar{c})$, and furthermore, the intersection is transversal. Based at a nearby point $(x, c) \in$ gph N_F, we could linearize gph N_F, calculate the intersection point of the resulting tangent space with $\mathbf{R}^n \times \{\bar{c}\}$, restore this point to gph N_F, and repeat. As we discuss later, such a process fundamentally underlies the philosophy of classical sequential quadratic programming.

In this classical example, we see that our theme just results in a standard methodology, namely sequential quadratic programming for nonlinear programming. Nonetheless, as should be apparent, even the concise summary we gave inevitably involves an extended setup involving active constraint indices, Lagrange multipliers, strict complementarity, and second-order sufficient conditions. Rather than algorithmic innovation, in this case our contribution is conceptual. By contrast with the traditional development, our theme is simpler and geometric, involving just fundamental ideas about manifolds: tangent spaces, transversality, and constant rank maps. In summary, although we hope that the ideas we develop here may help illuminate numerical methods in less traditional settings, our primary goal is not the immediate design of new algorithms. Rather, our aim is a simple unifying framework in which to study active-set-type acceleration techniques for a much broader class of optimization and variational problems.

2. Manifolds

We begin our more formal development with a brief summary of terminology for manifolds. We refer the reader to Lee [20] for basic background in differential geometry.

We consider manifolds in a Euclidean space $\mathbf{E}$, by which we mean sets defined, locally, by smooth equations with linearly independent gradients. More precisely, a set $\mathcal{X} \subset \mathbf{E}$ is a $C^{(r)}$-*manifold* around a point $z \in \mathcal{X}$ (for some degree $r = 1, 2, 3, \ldots$) if there exists a $C^{(r)}$-smooth map $F : \mathbf{E} \to \mathbf{R}^m$ (for some number m) with surjective derivative $DF(z)$ such that the inverse image $F^{-1}(0)$ is a neighborhood of z in $\mathcal{X}$. Equivalently, there exists a number k, an open neighborhood U of zero in $\mathbf{R}^k$, and a $C^{(r)}$-smooth map $G : U \to \mathbf{E}$ with value $G(0) = z$, everywhere injective derivative, and range $G(U)$ a neighborhood of z in $\mathcal{X}$. In that case the number $k = \dim \mathbf{E} - m$ is called the *dimension* of the manifold $\mathcal{X}$ around z, and we refer to G as a *local coordinate system*.

With the notation, the *tangent* and *normal* spaces, respectively, to $\mathcal{X}$ at z are given by

$$T_{\mathcal{X}}(x) = \text{Null}(DF(z)) = \text{Range}(DG(0))$$
$$N_{\mathcal{X}}(x) = \text{Range}(DF(z)^*) = \text{Null}(DG(0)^*).$$

These two orthogonal complements are independent of the choice of maps F and G. We say that two $C^{(2)}$-manifolds around z intersect *transversally* there if the two corresponding normal spaces intersect trivially.

3. A Newton Method for Intersecting Manifolds

The following simple scheme, illustrated in Figure 1, uses standard ideas. We study the intersection of two manifolds by linearizing just one of them.

Theorem 1 (Semilinearization Error). *In a Euclidean space, consider an isolated point of intersection z of two $C^{(2)}$-manifolds $\mathcal{X}$ and $\mathcal{Y}$ around z, and suppose that the intersection is transversal. Then, there exists an open neighborhood W of z such that, for all points $x \in \mathcal{X} \cap W$, the linear approximation $x + T_{\mathcal{X}}(x)$ intersects the set $\mathcal{Y} \cap W$ at a unique point $y(x)$, which satisfies*

$$|y(x) - z| = O(|x - z|^2) \quad as \; x \to z. \tag{2}$$

Proof. Because the manifolds $\mathcal{X}$ and $\mathcal{Y}$ intersect at a singleton, transversally their dimensions sum to dim $\mathbf{E}$. We can represent $\mathcal{X}$ using a local coordinate system: an open neighborhood U of zero in $\mathbf{R}^k$ and a $C^{(2)}$-smooth map $G : U \to \mathbf{E}$ with value $G(0) = z$, everywhere injective derivative, and range $G(U)$ an open neighborhood of z in $\mathcal{X}$. By definition, there also exists a $C^{(2)}$-smooth map $F : \mathbf{E} \to \mathbf{R}^k$ with surjective derivative $DF(z)$ such that the inverse image $F^{-1}(0)$ is a neighborhood of z in $\mathcal{Y}$.

The transversality condition is

$$\{0\} = N_{\mathcal{X}}(z) \cap N_{\mathcal{Y}}(z) = \text{Null}(DG(0)^*) \cap \text{Range}(DF(z)^*).$$

Hence, all vectors $v \in \mathbf{R}^k$ satisfy

$$DG(0)^*(DF(z)^* v) = 0 \quad \Leftrightarrow \quad DF(z)^* v = 0.$$

By assumption, the adjoint map $DF(z)^*$ is injective, so the right-hand side is equivalent to $v = 0$. Consequently, the composition $DG(0)^* \circ DF(z)^*$ is injective, so its adjoint $DF(z) \circ DG(0) : \mathbf{R}^k \to \mathbf{R}^k$ is surjective and hence, in fact, invertible. By the inverse function theorem, the composite map $F \circ G : \mathbf{R}^k \to \mathbf{R}^k$ has a $C^{(2)}$-smooth inverse map around zero.

We can express any point $x \in \mathcal{X}$ near z as $x = G(u)$ for some small vector $u \in \mathbf{R}^k$. In that case, the point we seek has the form $y(x) = G(u) + DG(u)v$ for some small vector $v \in \mathbf{R}^k$, where

$$F(G(u) + DG(u)v) = 0. \tag{3}$$

The function of (u, v) on the left-hand side, which we denote $H(u, v)$, is $C^{(1)}$-smooth. Its value at $(u, v) = (0, 0)$ is zero, and its derivative there with respect to v is $D_v H(0, 0) = DF(z) \circ DG(0)$. Because this derivative is invertible,

Figure 1. Semilinearized Newton method.
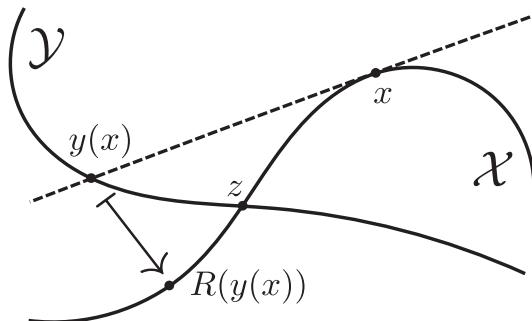
the implicit function theorem implies that, for all small vectors $u \in \mathbf{R}^k$, Equation (3) has a unique small solution $v(u) \in \mathbf{R}^k$, depending smoothly on u, and satisfying $v(0) = 0$.

Because the map G is $C^{(2)}$-smooth, we have

$$G(u + v) = G(u) + DG(u)v + O(|v|^2) \qquad (4)$$

for all small vectors u and v. Because the map F is Lipschitz, we deduce

$$
\begin{aligned}
F(G(u + v(u))) &= F\left(G(u) + DG(u)v(u) + O(|v(u)|^2)\right) \\
&= F(G(u) + DG(u)v(u)) + O(|v(u)|^2) \\
&= O(|v(u)|^2)
\end{aligned}
$$

as $u \to 0$, and hence,

$$u + v(u) = O(|v(u)|^2),$$

because the inverse map $(F \circ G)^{-1}$ is Lipschitz. However, because

$$|v(u)| \le |u| + |u + v(u)| = |u| + O(|v(u)|^2) \le |u| + \frac{1}{2}|v(u)|$$

for all small u, we deduce $|v(u)| \le 2|u|$. Equation (4) implies

$$
\begin{aligned}
y(x) - z = G(u) + DG(u)v(u) - z &= G(u + v(u)) + O(|v(u)|^2) - z \\
&= G(O(|v(u)|^2) - z + O(|v(u)|^2) \\
&= O(|v(u)|^2) = O(|u|^2) = O(|G(u) - z|^2) = O(|x - z|^2)
\end{aligned}
$$

as claimed. □

How we compute the point $y(x)$ in concrete settings depends on our representation of the manifolds $\mathcal{X}$ and $\mathcal{Y}$. For example, we discuss one particularly interesting case arising in the study of variational inequalities in Section 6.

If we can somehow restore the point $y(x)$ in the manifold $\mathcal{Y}$ to a nearby point in the manifold $\mathcal{X}$, by iterating we rapidly converge to the intersection point z. Based on this idea, illustrated in Figure 1, the following simple result is the foundation for all the Newton schemes we consider.

Corollary 1 (Semilinearization for Intersecting Manifolds). *Following the framework of Theorem 1, consider two manifolds $\mathcal{X}$ and $\mathcal{Y}$ intersecting transversally at their unique common point z and in addition, a Lipschitz map $R : \mathcal{Y} \to \mathcal{X}$ satisfying $R(z) = z$. Then, starting from any point $x \in \mathcal{X}$ near z, the iteration*

$$x \leftarrow R(y(x))$$

converges quadratically to z.

Proof. Using the fixed point and Lipschitz properties and Equation (2), we deduce

$$\left|R(y(x)) - z\right| = \left|R(y(x)) - R(z)\right| = O(|y(x) - z|) = O(|x - z|^2),$$

for all points $x \in \mathcal{X}$ near z, as required. □

Because $\mathcal{X}$ is a $C^{(2)}$-smooth manifold around the point z, a natural candidate for the restoration map R is $\mathrm{Proj}_{\mathcal{X}}$, the corresponding nearest-point projection map onto $\mathcal{X}$, which is $C^{(1)}$-smooth and hence, Lipschitz around z and which obviously leaves z fixed. However, in the broad scenario that we have in mind, this projection may not be the most natural choice. We describe this scenario next.

4. Generalized Equations

Henceforth, we fix a Euclidean space $\mathbf{U}$ and consider the problem of finding a zero of a set-valued mapping Φ or in other words, a solution of the generalized equation

$$0 \in \Phi(u).$$

Henceforth, we rely on standard terminology from variational analysis, as presented in monographs such as Rockafellar and Wets [41] and Mordukhovich [39].

Assuming that the graph of Φ is a manifold, locally we can frame this equation as an intersection problem as in the previous section and Theorem 1. The underlying Euclidean space is now $\mathbf{E} = \mathbf{U}^2$, the sets $\mathcal{X}$ and $\mathcal{Y}$ are the graphs of Φ and the zero mapping, respectively:

$$\mathcal{X} = \operatorname{gph}\Phi = \{(u,v) \in \mathbf{U}^2 : v \in \Phi(u)\}$$
$$\mathcal{Y} = \mathbf{U} \times \{0\},$$

and the intersection point is $z = (\bar{u}, 0)$. The transversality condition is

$$N_{\operatorname{gph}\Phi}(\bar{u}, 0) \cap N_{\mathbf{U}\times\{0\}}(\bar{u}, 0) = \{(0, 0)\}.$$

Automatically, then, we must have $\dim \operatorname{gph}\Phi = \dim \mathbf{U}$.

Using the language of *coderivatives*, we have by definition for any set-valued mapping Φ the relation

$$(v, -y) \in N_{\operatorname{gph}\Phi}(\bar{u}, 0) \quad \Leftrightarrow \quad v \in D^*\Phi(\bar{u}|0)(y).$$

In this definition, N denotes the *limiting normal cone*, which coincides with the classical normal space for manifolds. Hence, because

$$N_{\mathbf{U}\times\{0\}}(\bar{u}, 0) = \{0\} \times \mathbf{U},$$

we can rewrite the transversality condition as in the following assumption.

Assumption 1. *The point $\bar{u} \in \mathbf{U}$ is an isolated zero of a set-valued mapping $\Phi : \mathbf{U} \rightrightarrows \mathbf{U}$. Furthermore, the graph of Φ is a $C^{(2)}$-smooth manifold around the point $(\bar{u}, 0)$, and the transversality condition*

$$0 \in D^*\Phi(\bar{u}|0)(y) \quad \Rightarrow \quad y = 0. \tag{5}$$

holds.

We remark in passing that the coderivative criterion (5) plays a fundamental role throughout the variational analysis literature. For any set-valued mapping Φ with closed graph and satisfying $0 \in \Phi(\bar{u})$, it is equivalent to *metric regularity* (Rockafellar and Wets [41, theorem 9.43]): the ratio of two distances

$$\frac{\operatorname{dist}(u, \Phi^{-1}(v))}{\operatorname{dist}(v, \Phi(u))}$$

is uniformly bounded above for pairs (u, v) near $(\bar{u}, 0)$ (where we interpret $\frac{0}{0} = 0$).

Turning to the linearization step in Theorem 1, given a current point $u \in \mathbf{U}$ and a current value $v \in \Phi(u)$, using the language of *graphical derivatives*, we have by definition

$$(u', 0) \in (u, v) + T_{\operatorname{gph}\Phi}(u, v) \quad \Leftrightarrow \quad -v \in D\Phi(u|v)(u' - u). \tag{6}$$

In this definition, T denotes the standard *tangent cone*, which coincides with the classical tangent space for manifolds. Having calculated the new point $u' \in \mathbf{U}$ from this relationship, to apply Corollary 1 we need a Lipschitz map to restore the point $(u', 0)$ to a nearby point on gph Φ.

4.1. Example: Classical Newton Method

For illustration, consider the case where the mapping Φ is just a single-valued $C^{(2)}$-smooth map $F : \mathbf{U} \to \mathbf{U}$. The transversality condition (5) reduces to invertibility of the derivative map $DF(\bar{u})$. At the current pair $(u, F(u)) \in \operatorname{gph}F$, the linearization step (6) results in a new pair $(u', 0) \in \mathbf{U} \times \{0\}$ where $u' = u - (DF(u))^{-1}F(u)$. Rather than trying to project $(u', 0)$ onto $\operatorname{gph}F$, we can use the natural restoration map R defined by $(u', 0) \mapsto (u', F(u'))$,

which fixes the point $(\bar{u}, 0)$ and obviously inherits the Lipschitz property of F. Corollary 1 now simply recovers the quadratic convergence of the classical Newton iteration for solving $F(u) = 0$.

4.2. Example: Variational Inequalities

More generally, consider the case

$$\Phi = F + \Psi$$

for a single-valued $C^{(2)}$-smooth map $F : \mathbf{U} \to \mathbf{U}$ and a maximal monotone operator $\Psi : \mathbf{U} \rightrightarrows \mathbf{U}$. In particular, Ψ might be the subdifferential operator ∂g for a proper closed convex function $g : \mathbf{U} \to \overline{\mathbf{R}}$. This case includes two important examples: the special case when F is the derivative ∇f of a smooth function $f : \mathbf{U} \to \mathbf{R}$, corresponding to the optimization problem $\inf\{f + g\}$, and the special case when g is the indicator function for a closed convex set $K \subset \mathbf{U}$, corresponding to the *variational inequality*

$$\mathrm{VI}(K, F) : \quad 0 \in F(u) + N_K(u).$$

We assume that the point $\bar{u} \in \mathbf{U}$ is an isolated zero of the mapping $F + \Psi$ and that the graph of the operator Ψ is a $C^{(2)}$-smooth manifold around the point $(\bar{u}, -F(\bar{u}))$. Standard nonsmooth calculus (or direct calculation) implies

$$D^*(F + \Psi)(\bar{u}|0)(y) = DF(\bar{u})^* y + D^* \Psi(\bar{u}| - F(\bar{u}))(y)$$
$$D(F + \Psi)(u|F(u) + w)(s) = DF(u)s + D\Psi(u|w)(s)$$

for $w \in \Psi(u)$. Hence, the transversality condition (5) becomes

$$-DF(\bar{u})^* y \in D^* \Psi(\bar{u}| - F(\bar{u}))(y) \quad \Rightarrow \quad y = 0.$$

Given a current vector $u \in \mathbf{U}$ and a vector $w \in \Psi(u)$ corresponding to the choice $v = F(u) + w$ in the relation (6), the linearization step is defined by

$$-F(u) - w \in DF(u)(u' - u) + D\Psi(u|w)(u' - u).$$

It remains to find a Lipschitz map restoring the new point $(u', 0)$ to $\mathrm{gph}\,(F + \Psi)$. To that end we can use Minty's Theorem (Rockafellar and Wets [41, theorem 12.15]), which guarantees that, for any choice of $\lambda > 0$, the *resolvent* map $(I + \lambda\Psi)^{-1} : \mathbf{U} \to \mathbf{U}$ is single valued and nonexpansive. Define a Lipschitz map $Q : \mathbf{U} \to \mathbf{U}$ by

$$Q(u) = (I + \lambda\Psi)^{-1}(u - \lambda F(u)). \tag{7}$$

Clearly, zeros of the mapping $F + \Psi$ coincide with fixed points of Q, and simple calculations show

$$\left(Q(u), \frac{1}{\lambda}(u - Q(u)) - F(u) + F(Q(u))\right) \in \mathrm{gph}\,(F + \Psi) \quad \text{for all } u \in \mathbf{U}.$$

Hence, we can use the restoration map $R : \mathbf{U} \times \{0\} \to \mathrm{gph}\,(F + \Psi)$ defined by setting $R(u, 0)$ to the left-hand side of this inclusion. This map is clearly Lipschitz and fixes the point $(\bar{u}, 0)$, as required.

In the smooth case ($\Psi = 0$), we arrive at a slight variant of the classical Newton method for the smooth equation $F(u) = 0$ that we saw in the previous example (and coinciding with it in the limiting case $\lambda = 0$). A simple extension is when the mapping Ψ is the normal cone operator N_L for a linear subspace $L \subset \mathbf{U}$, in which case we arrive at a slight variant of the natural projected Newton scheme for solving $F(u) = 0$, $u \in L$.

5. Identifying Primal Activity: Partial Smoothness

Corollary 1 describes the basic Newton-type approach to the manifold intersection scenario

$$\mathcal{X} \cap \mathcal{Y} = \{z\}$$

and more particularly, for the special case associated generalized equation $0 \in \Phi(u)$:

$$\mathrm{gph}\,\Phi \cap (\mathbf{U} \times \{0\}) = \{(\bar{u}, 0)\}.$$

The Newton iteration depends fundamentally on a Lipschitz restoration map R from the manifold $\mathcal{X}$ into the manifold $\mathcal{Y}$ that fixes the intersection point z. In general, it may be unclear how to construct such a map,

and even in special cases when a closed form is available, as in the previous maximal monotone examples, the resulting conceptual algorithm is not transparent and involves simultaneous coupled updates to primal variables (in the domain of the mapping Φ) and dual variables (in the range of Φ).

In this section, we describe a systematic approach to the restoration map R, motivated by traditional active-set optimization algorithms. We decouple updates to the primal and dual variables, thereby decomposing R into a composition of two maps. The first map updates the primal variables by identifying an underlying active submanifold in the primal manifold $\mathcal{Y} = \mathbf{U} \times \{0\}$. The second map fixes the primal variables and updates the dual variables to restore the iterate to the manifold $\mathcal{X} = \mathrm{gph}\,\Phi$. This approach depends crucially on the existence of a constant rank map from $\mathcal{X}$ into $\mathcal{Y}$: in the special case of the generalized equation $0 \in \Phi(u)$, this assumption amounts to the "partial smoothness" of the mapping Φ, language that we argue elsewhere is convenient for the philosophy of active set identification.

Theorem 2 (Active Set Method for Manifold Intersection). *In a Euclidean space, consider an isolated point of intersection z of two $C^{(2)}$-manifolds $\mathcal{X}$ and $\mathcal{Y}$ around z, and suppose that the intersection is transversal. Suppose furthermore that there exists a map $P : \mathcal{X} \to \mathcal{Y}$ that fixes the point z and is $C^{(2)}$-smooth and constant rank nearby. Then, for any sufficiently small neighborhood W of z, the image*

$$\mathcal{M} = P(\mathcal{X} \cap W) \subset \mathcal{Y}$$

is a $C^{(2)}$-manifold around z, and for all points $x \in \mathcal{X} \cap W$, the linear approximation $x + T_\mathcal{X}(x)$ intersects the set $\mathcal{Y} \cap W$ at a unique point $y(x)$. Furthermore, given any two Lipschitz maps that fix z,

$$S : \mathcal{Y} \cap W \to \mathcal{M} \quad \text{and} \quad Q : \mathcal{M} \cap W \to \mathcal{X},$$

if Q is a right inverse for P, meaning

$$P(Q(w)) = w \quad \text{for all } w \in \mathcal{M} \cap W,$$

then starting from any point $x \in \mathcal{X}$ near z, the iteration

$$x \leftarrow Q\big(S\big(y(x)\big)\big)$$

converges quadratically to z. Examples of such maps S and Q are the projections

$$S = \mathrm{Proj}_\mathcal{M} : \mathcal{Y} \cap W \to \mathcal{M} \quad \text{and} \quad Q(w) = \mathrm{Proj}_{P^{-1}(w)}(w) \quad \text{for all } w \in \mathcal{M} \cap W,$$

respectively.

Proof. The manifold property of the image $\mathcal{M}$ follows from the Constant Rank Theorem. The projection example for the map S follows from the well-known fact that the projection onto a $C^{(2)}$-manifold around a point z is $C^{(1)}$-smooth around z. The remainder of the result follows from Corollary 1, with the exception of the projection example for the map Q: we just need to show that this map is in fact $C^{(1)}$-smooth around z.

By the Constant Rank Theorem, there exist local coordinate systems for the manifolds $\mathcal{X}$ and $\mathcal{Y}$ around the point z with respect to which the map P is linear and in fact, maps the point with coordinate vector $(u, v) \in \mathbf{U} \times \mathbf{V}$ to the point with coordinate vector $(u, 0) \in \mathbf{U} \times \mathbf{W}$, for suitable Euclidean spaces $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{W}$. To be more precise, there exist $C^{(2)}$-smooth maps $G : \mathbf{U} \times \mathbf{V} \to \mathbf{E}$ and $H : \mathbf{U} \times \mathbf{W} \to \mathbf{E}$ (the coordinate systems for $\mathcal{X}$ and $\mathcal{Y}$, respectively), satisfying $G(0, 0) = z = H(0, 0)$ and with injective derivatives $DG(0, 0)$ and $DH(0, 0)$, and such that

$$P(G(u, v)) = H(u, 0) \quad \text{for all small } u \in \mathbf{U}, \ v \in \mathbf{V}.$$

Around z, the manifold $\mathcal{M}$ then just consists of points of the form $H(u, 0)$ for small vectors $u \in \mathbf{U}$. An obvious example of a right inverse map Q with the desired properties is given by

$$Q(H(u, 0)) = G(u, 0) \quad \text{for all small } u \in \mathbf{U}$$

because then we have

$$P(Q(H(u, 0))) = P(G(u, 0)) = H(u, 0) \quad \text{for all small } u \in \mathbf{U}.$$

For any small $\delta > 0$, the coordinate map G is a diffeomorphism from the compact set

$$\{(u, v) \in \mathbf{U} \times \mathbf{V} : |u| \leq \delta, \ |v| \leq \delta\}$$

onto a compact neighborhood $\hat{\mathcal{X}}$ of the point z in the manifold $\mathcal{X}$, and the map $u \mapsto H(u, 0)$ is a diffeomorphism from the compact set $\{u \in \mathbf{U} : |u| \le \delta\}$ onto a compact neighborhood of z in the manifold $\mathcal{M}$. Furthermore, for all $u \in \mathbf{U}$ satisfying $|u| \le \delta$, we have

$$P^{-1}(H(u, 0)) \cap \hat{\mathcal{X}} = \{G(u, v) : v \in \mathbf{V}, \ |v| \le \delta\}.$$

Thus, for all small u, projecting the point $w = H(u, 0)$ onto the set $P^{-1}(w)$ amounts to solving the optimization problem

$$\inf_{v \in \mathbf{V}} \{|G(u, v) - H(u, 0)|^2 : |v| \le \delta\}.$$

The infimum is attained, by compactness. By an easy continuity argument, if u is sufficiently small, any optimal v satisfies $|v| < \delta$ and hence, also the first-order necessary condition

$$D_v G(u, v)^*(G(u, v) - H(u, 0)) = (0, 0). \tag{8}$$

The left-hand side of this equation is $C^{(1)}$-smooth. When $u = 0$, the equation reduces to

$$D_v G(0, v)^*(G(0, v)) = (0, 0).$$

The derivative at $v = 0$ of the map on the left-hand side is $D_v G(0, 0)^* D_v G(0, 0)$, which is invertible because $D_v G(0, 0)$ is injective. By the implicit function theorem, for all small $u \in \mathbf{U}$, Equation (8) has a unique small solution $v(u)$, depending smoothly on u, and clearly $v(0) = 0$. Because we have shown, for such u, the property

$$\operatorname*{Proj}_{P^{-1}(H(u,0))} (H(u, 0)) = G(u, v(u)),$$

the result now follows. □

Returning to the setting of generalized equations $0 \in \Phi(u)$, for a variable u in a Euclidean space $\mathbf{U}$ and a set-valued mapping $\Phi : \mathbf{U} \rightrightarrows \mathbf{U}$, we next introduce the notion of partial smoothness developed in Lewis and Liang [25].

Definition 1 (Partly Smooth Mappings). For any degree $r = 1, 2, 3, \ldots$, a set-valued mapping $\Phi : \mathbf{U} \rightrightarrows \mathbf{U}$ is called $C^{(r)}$-*partly smooth* at a point $\bar{u} \in \mathbf{U}$ *for a value* $\bar{v} \in \Phi(\bar{u})$ when its graph is a $C^{(r)}$-smooth manifold around the point $(\bar{u}, \bar{v})$ and the map defined by $(u, v) \mapsto u$ for $(u, v) \in \operatorname{gph} \Phi$ is constant rank nearby.

By definition, the constant rank condition means that the coderivative,

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

is a subspace of constant dimension for points (u, v) near $(\bar{u}, \bar{v})$. The following result is Lewis and Liang [25, proposition 3.2].

Proposition 1 (Active Manifold). *Assuming Definition 1, there exists a set $\mathcal{M}$ in the space $\mathbf{U}$ such that any point (u, v) sufficiently near the point $(\bar{u}, \bar{v})$ in the graph of the mapping Φ satisfies $u \in \mathcal{M}$. Any such set $\mathcal{M}$ is a $C^{(r)}$-smooth manifold around $\bar{u}$ and is locally unique: that is, any two such sets are identical around $\bar{u}$.*

We call the set $\mathcal{M}$ in this result the *active manifold at $\bar{u}$ for $\bar{v}$.*

With this terminology, we can specialize Theorem 2 to generalized equations.

Theorem 3 (Active Set Method for Generalized Equations). *Consider an isolated zero $\bar{u} \in \mathbf{U}$ for a set-valued mapping $\Phi : \mathbf{U} \rightrightarrows \mathbf{U}$. Suppose that Φ is $C^{(2)}$-partly smooth at $\bar{u}$ for 0, with active manifold $\mathcal{M} \subset \mathbf{U}$, and assume the transversality condition*

$$0 \in D^*\Phi(\bar{u}|0)(y) \quad \Rightarrow \quad y = 0.$$

Then, given any sufficiently small neighborhood W of $\bar{u}$, for all points $(u, v) \in \operatorname{gph} \Phi$ near $(\bar{u}, 0)$, the linearized equation

$$-v \in D\Phi(u|v)(w - u), \quad w \in W$$

has a unique solution $w(u,v)$. Furthermore, given any Lipschitz map $S : W \to \mathcal{M}$ that fixes the point $\bar{u}$, and any Lipschitz map $Q : \mathcal{M} \cap W \to U$ satisfying $Q(\bar{u}) = 0$ and

$$Q(u) \in \Phi(u) \quad \text{for all } u \in \mathcal{M} \cap W,$$

starting from any point $(u,v) \in \text{gph } \Phi$ near $(\bar{u}, 0)$, the iteration

$$u \leftarrow S(w(u,v)), \quad v \leftarrow Q(u)$$

converges quadratically to $(\bar{u}, 0)$. Examples of such maps S and Q are the projections

$$S = \underset{\mathcal{M}}{\text{Proj}} : W \to \mathcal{M} \quad \text{and} \quad Q(u) = \underset{\Phi(u)}{\text{Proj}}(0) \quad \text{for all } u \in \mathcal{M} \cap W,$$

respectively.

Proof. We simply apply Theorem 2 in the framework of Section 4:

$$\mathcal{X} = \text{gph } \Phi, \, \mathcal{Y} = U \times \{0\}$$
$$P(u,v) = (u, 0) \text{ for } (u,v) \in \text{gph } \Phi.$$

The result follows immediately. □

Theorem 3 describes a Newton-type method for generalized equations of the form $0 \in \Phi(u)$, under a transversality condition equivalent (as we observed after Assumption 1) to metric regularity of the mapping Φ at the solution. An extensive literature treats nonsmooth Newton methods under metric regularity assumptions: for variational inequalities and nonlinear programs in particular, one standard reference is Bonnans [2]. Such methods typically linearize a "semismooth" single-valued ingredient while leaving a set-valued ingredient unchanged, and the underlying assumption is "strong metric subregularity." Contemporary monographs treating Newton-type methods in this context include Klatte and Kummer [19], Facchinei and Pang [10], Dontchev and Rockafellar [6], and Izmailov and Solodov [17], and the recent survey by Izmailov and Solodov [18] is instructive. Algorithms, like that described in the recent manuscript by Gfrerer and Outrata [15], that are based on linearizing the underlying set-valued mapping itself are more rare: our approach is just such a technique.

6. Partly Smooth Variational Inequalities

Consider an isolated solution $\bar{u}$ of the variational inequality $VI(K, F)$:

$$0 \in F(u) + N_K(u),$$

for a $C^{(2)}$-smooth map $F : U \to U$ and a closed (possibly nonconvex) set $K \subset U$. (In particular, if $F = \nabla f$ for some $C^{(3)}$-smooth function $f : U \to \mathbf{R}$, then the variational inequality reduces to the first-order condition for the optimization problem $\inf_K f$.) Our aim is to apply Theorem 3 (active set method for generalized equations).

To this end, we suppose throughout this section that the set K is prox-regular at $-F(\bar{u})$ and that $\bar{u}$ is in fact (following standard terminology of Facchinei and Pang [10]) a *nondegenerate* solution:

$$-F(\bar{u}) \in \text{ri } N_K(\bar{u}).$$

The practically important but much more complex case where nondegeneracy fails is analyzed in Fadili et al. [12]. Earlier discussions for nonlinear programming and complementarity problems include Facchinei et al. [11] and Izmailov and Solodov [17, chapter 3.4].

We assume furthermore that K is $C^{(3)}$-*partly smooth* at $\bar{u}$ for $-F(\bar{u})$ relative to some $C^{(3)}$-manifold $\mathcal{M} \subset K$ around $\bar{u}$, by which we mean that the normal cone mapping N_K is inner semicontinuous at $\bar{u}$ relative to $\mathcal{M}$ and satisfies the *sharpness* condition

$$\text{span } N_K(\bar{u}) = N_{\mathcal{M}}(\bar{u}).$$

In this case, we have the local property (Drusvyatskiy and Lewis [8])

$$\text{gph } N_K = \text{gph } N_{\mathcal{M}} \quad \text{around } (\bar{u}, -F(\bar{u})). \tag{9}$$

(The graph of $N_{\mathcal{M}}$ is sometimes called the *normal bundle* for the manifold $\mathcal{M}$.) It is precisely this property that underlies active set approaches to the original problem.

6.1. Example: The Basic Projection Algorithm

To illustrate, suppose for this example that the set K is convex. For any constant $a > 0$, by defining a continuous map $y_a : \mathbf{U} \to Q$ by

$$y_a(u) = \mathrm{Proj}_K\left(u - \frac{1}{a}F(u)\right)$$

for points $u \in \mathbf{U}$, following a standard route (Facchinei and Pang [10, section 12.1]) we can convert the variational inequality into a fixed point problem: u solves VI(K, F) if and only if $y_a(u) = u$. Indeed, the map is just a special case of the map (7) that we defined for more general maximal monotone operators. In the example of convex optimization, where F is the gradient of a smooth convex function, the iteration $u \leftarrow y_a(u)$ is just the projected gradient method. Again specializing our earlier discussion, by definition we have

$$a(u - y_a(u)) - F(u) \in N_K(y_a(u))$$

for all points $u \in U$. Under certain general conditions (Facchinei and Pang [10, theorem 12.1.2]), any sequence of iterates $u_k \in \mathbf{U}$ (for $k = 0, 1, 2, \ldots$) must converge to the solution $\bar{u}$. By continuity, we have

$$y_k = y_a(u_k) \to \bar{u},$$

and also,

$$v_k = a(u_k - y_k) - F(u_k) \to -F(\bar{u}),$$

with

$$v_k \in N_K(y_k).$$

Hence, by the reduction property (9), we deduce $y_k \in \mathcal{M}$ eventually: in other words, the sequence y_k *identifies* the active manifold $\mathcal{M}$. In the case of the projected gradient method for convex optimization, this is exactly the result of Wright [45], based on earlier work in Dunn [9]. The monograph (Facchinei and Pang [10]) includes a wide-ranging discussion of general solution algorithms for variational inequalities: the identification behavior holds for quite broad classes. Global techniques often aim to drive down the natural *residual* $\|u - y_a(u)\|$ indirectly by driving down a merit function such as the *regularized gap*

$$\theta_a(u) = \sup_{y \in K}\left\{\langle F(u), u - y\rangle - \frac{a}{2}\|u - y\|^2\right\}$$

or the *D-gap* given by the function $\theta_a - \theta_b$, for some constant b in the interval $(0, a)$ (see Facchinei and Pang [10, propositions 10.3.7 and 10.3.8]). Such approaches also lead to identification: if the iterates u_k converge to some limit $\bar{u}$, and the corresponding residuals converge to zero, then continuity ensures that $\bar{u}$ is a fixed point of the map y_a and hence, a solution of the variational inequality; so, under the partial smoothness assumptions, the nearby sequence $y_a(u_k)$ eventually lies in the manifold $\mathcal{M}$.

Returning to the general case, property (9) shows that, locally, around the solution $\bar{u}$, the original variational inequality VI(K, F) is equivalent to the variational inequality VI($\mathcal{M}, F$):

$$0 \in F(u) + N_{\mathcal{M}}(u).$$

As we remarked, this reduction captures the essence of the active set philosophy: we can henceforth focus on applying Theorem 3 for the mapping $\Phi = F + N_{\mathcal{M}}$.

Following Lewis and Liang [25], the normal space mapping $N_{\mathcal{M}}$ is $C^{(2)}$-partly smooth at $\bar{u}$ for $-F(\bar{u})$, with active manifold $\mathcal{M}$, so by an easy calculus rule by Lewis and Liang [25], the mapping $F + N_{\mathcal{M}}$ is $C^{(2)}$-partly smooth at $\bar{u}$ for zero, as we need to apply Theorem 3. The theorem involves two more key ingredients. The first is the transversality condition to check, which we can write

$$-(DF(\bar{u}))^*y \in D^*N_{\mathcal{M}}(\bar{u}| - F(\bar{u}))(y) \quad \Rightarrow \quad y = 0,$$

using routine coderivative calculus (Rockafellar and Wets [41, example 10.43]). The second is the linearized equation: given points $u \in \mathbf{U}$ near $\bar{u}$ and small $v \in F(u) + N_{\mathcal{M}}(u)$, we solve

$$-v \in DF(u)(w - u) + DN_{\mathcal{M}}(u|v - F(u))(w - u)$$

for the new iterate w. Here, once again we have applied some standard calculus, this time for derivatives (Rockafellar and Wets [41, example 10.43]).

In this light, we therefore next discuss how to compute the derivative and coderivative maps

$$DN_{\mathcal{M}}(u|z) \quad \text{and} \quad D^* N_{\mathcal{M}}(u|z) \tag{10}$$

for $u \in \mathcal{M}$ and $z \in N_{\mathcal{M}}(u)$. These are second-order concepts because $N_{\mathcal{M}} = \partial \delta_{\mathcal{M}}$. Indeed, using the notation of the Mordukhovich *generalized Hessian* (Mordukhovich [38]), we have

$$D^* N_{\mathcal{M}} = \partial^2 \delta_{\mathcal{M}}.$$

The Hessian of a $C^{(2)}$-smooth function $f : \mathbf{U} \to \mathbf{R}$ has an important symmetry property that we can express as the relationship

$$D(\nabla f) = D^*(\nabla f).$$

By contrast, for general sets $\mathcal{M}$, there is no relationship between the derivative and coderivative mappings (10). On the other hand, for fully amenable sets (and hence, in particular, for manifolds), we always have the *derivative-coderivative inclusion*

$$DN_{\mathcal{M}}(u|z) \subset D^* N_{\mathcal{M}}(u|z);$$

see Rockafellar and Wets [41, theorem 13.57]. Our next result shows that, for manifolds, this inclusion in fact holds with equality: informally, second derivatives of manifold indicator functions are symmetric.

Theorem 4 (Derivative-Coderivative Equality for Manifolds). *Consider a $C^{(2)}$-smooth manifold $\mathcal{M} \subset \mathbf{U}$ around a point $u \in \mathcal{M}$ and a normal vector $v \in N_{\mathcal{M}}(u)$. Then, the normal mapping $N_{\mathcal{M}} : \mathbf{U} \overset{\rightarrow}{\rightarrow} \mathbf{U}$ satisfies the derivative-coderivative equality*

$$DN_{\mathcal{M}}(u|v) = D^* N_{\mathcal{M}}(u|v).$$

Proof. By definition, we can suppose the manifold $\mathcal{M}$ to be defined as the set of common zeros near the point u of $C^{(2)}$-smooth functions $h_i : \mathbf{U} \to \mathbf{R}$ (for $i = 1, 2, 3, \ldots, m$), where the gradients $\nabla h_i(u)$ are linearly independent, and that $v = \sum_i \lambda_i \nabla h_i(u)$ for some vector $\lambda \in \mathbf{R}^m$. Define a corresponding self-adjoint linear map $H : \mathbf{U} \to \mathbf{U}$ by $H = \sum_i \lambda_i \nabla^2 h_i(u)$. By definition, for vectors $u, w \in \mathbf{U}$ we have

$$z \in DN_{\mathcal{M}}(u|v)(w)$$

if and only if

$$(w, z) \in T_{\mathrm{gph}\, N_{\mathcal{M}}}(u, v),$$

which is equivalent (by Lewis and Zhang [27, theorem 2.8]) to

$$w \in T_{\mathcal{M}}(u) \quad \text{and} \quad z - Hw \in N_{\mathcal{M}}(u). \tag{11}$$

On the other hand, we have

$$z \in D^* N_{\mathcal{M}}(u|v)(w)$$

if and only if

$$(z, -w) \in N_{\mathrm{gph}\, N_{\mathcal{M}}}(u, v),$$

which is also equivalent (by Lewis and Zhang [27, theorem 2.8]) to condition (11). $\square$

As an aside, we note an elegant consequence of this result: again informally speaking, partly smooth *functions* have symmetric second derivatives. To be precise, we gather together in the next result various properties of functions $f : \mathbf{U} \to \overline{\mathbf{R}}$ that are partly smooth in the sense of Drusvyatskiy and Lewis [7], the functional version of the definition for sets in Drusvyatskiy and Lewis [8]. This idea is a slight variant of

the original notion from Lewis [24]: it corresponds closely to our current notion of partial smoothness applied to the subdifferential mapping ∂f.

Corollary 2. *Consider a $C^{(r)}$-smooth manifold $\mathcal{M} \subset \mathbf{U}$ around a point $\bar{u} \in \mathcal{M}$, for some degree $r > 1$. Suppose a subgradient v of a function $f : \mathbf{U} \to \overline{\mathbf{R}}$ at $\bar{u}$ satisfies the nondegeneracy condition*

$$\bar{v} \in \mathrm{ri}\,\partial f(\bar{u}),$$

as well as the following conditions:
- *f is prox-regular and subdifferentially continuous at $\bar{u}$ for $\bar{v}$;*
- *f agrees with some $C^{(r)}$-smooth function $\bar{f} : \mathbf{U} \to \mathbf{R}$ on $\mathcal{M}$ around $\bar{u}$;*
- *the affine span of 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 $\mathcal{M}$.*

Then, the subdifferential mapping ∂f is $C^{(r-1)}$-partly smooth at $\bar{u}$ for $\bar{v}$, with active manifold $\mathcal{M}$, and locally, satisfies

$$\mathrm{gph}\,\partial f = \{(u, \nabla \bar{f}(u) + v) : u \in \mathcal{M},\ v \in N_{\mathcal{M}}(u)\} \quad \text{around } (\bar{u}, \bar{v}).$$

Furthermore, the following derivative-coderivative equality holds:

$$D(\partial f)(\bar{u}|\bar{v}) = D^*(\partial f)(\bar{u}|\bar{v}) = \partial^2 f(\bar{u}|\bar{v}) = \nabla^2 \bar{f}(\bar{u}) + D(N_{\mathcal{M}})(\bar{u}|\bar{v} - \nabla \bar{f}(\bar{u})).$$

Note that the bulleted properties (without the assumption of subdifferential continuity) together constitute the definition of *partial smoothness* of f at $\bar{u}$ for $\bar{v}$.

Proof. All except the final claim are simply a restatement from Lewis and Liang [25]. The final claim follows from the fact that the graphs of the mappings ∂f and $\nabla \bar{f} + N_{\mathcal{M}}$ agree around the point $(\bar{u}, \bar{v})$, and hence, these mappings have the same derivative and coderivative mappings. The result now follows from derivative and coderivative calculus and Theorem 4. □

Having calculated the relevant derivatives and coderivatives in Theorem 4, we can now interpret the active set method described in Theorem 3 for the variational inequality VI(K, F). Under the partial smoothness assumptions at the beginning of this section, the problem reduces locally around the solution $\bar{u}$ to the variational inequality

$$0 \in F(u) + N_{\mathcal{M}}(u)$$

or equivalently, using the notation of the proof of Theorem 4, to the system

$$F(u) + \sum_{i=1}^{m} \lambda_i \nabla h_i(u) = 0 \tag{12}$$

$$h_i(u) = 0 \quad (i = 1, 2, 3, \ldots, m). \tag{13}$$

in the variables $u \in \mathbf{U}$ and $\lambda \in \mathbf{R}^m$. By assumption, there exists a unique vector $\bar{\lambda} \in \mathbf{R}^m$ such that the function $G = F + \sum_i \bar{\lambda}_i \nabla h_i$ is zero at $\bar{u}$.

As we have seen, the transversality condition is

$$-(DF(\bar{u}))^* y \in D^* N_{\mathcal{M}}(\bar{u}| - F(\bar{u}))(y) \quad \Rightarrow \quad y = 0,$$

or equivalently, by property (11),

$$y \in T_{\mathcal{M}}(\bar{u}) \quad \text{and} \quad DG(\bar{u})^* y \in N_{\mathcal{M}}(\bar{u}) \quad \Rightarrow \quad y = 0.$$

Not surprisingly, this simply amounts to the invertibility of the system (12) and (13) linearized around the solution $(\bar{u}, \bar{\lambda})$. In particular, under these conditions the fact that $\bar{u}$ is an isolated solution is an automatic consequence. In the case of optimization (where the map F is a gradient), the condition reduces to non-singularity of the Hessian of the Lagrangian function, projected onto the tangent space $T_{\mathcal{M}}(\bar{u})$.

In other terminology, as we observed earlier, the transversality condition is also equivalent to metric regularity of the mapping $F + N_{\mathcal{M}}$ at the point $\bar{u}$ for the value zero. Invertibility of the system (12) and (13) in fact guarantees *strong metric regularity*: localized around $\bar{u}$, the mapping has a single-valued Lipschitz inverse

around zero. This fact is analogous to the equivalence of metric regularity and strong metric regularity for Karush-Kuhn-Tucker mappings (Dontchev and Rockafellar [6, theorem 4I.1]).

Turning to the linearized equation, we consider a current point $u \in \mathbf{U}$ near $\bar{u}$ and small vector $v \in F(u) + N_{\mathcal{M}}(u)$, which we can write as $L(u)$, where

$$L = F + \sum_i \lambda_i \nabla h_i$$

is the *Lagrangian*, for some unique multiplier vector $\lambda \in \mathbf{R}^m$. We must solve

$$-v \in DF(u)s + DN_{\mathcal{M}}\left(u \Big| \sum_i \lambda_i \nabla h_i(u)\right)s$$

for the Newton step s and then update $u \leftarrow u + s$. We can write this equation equivalently, by property (11), as the linear system

$$s \in T_{\mathcal{M}}(u) \quad \text{and} \quad L(u) + DL(u)s \in N_{\mathcal{M}}(u),$$

so this step amounts to seeking a critical point of the Lagrangian restricted to the tangent space, a familiar operation in sequential quadratic programming approaches.

We can now interpret Theorem 3, after using this active set approach to reduce to the concrete system (12) and (13). We linearize this system around the current iterate u and current Lagrange multiplier estimate $\lambda \in \mathbf{R}^m$ and solve for the new pair (u', λ'). We then resort to exact feasibility restoration—often unrealistic in practice but nonetheless useful conceptually. To be precise, we map the first component u' back onto a new iterate u on the active manifold $\mathcal{M}$ defined by Equation (13): conceptually, we could simply consider the nearest-point projection, especially if the original set K is convex, because then the projections Proj_K and $\text{Proj}_{\mathcal{M}}$ agree near the solution. Finally, we estimate a new multiplier vector $\lambda \in \mathbf{R}^m$: for example, by minimizing $\|F(u) + \sum_i \lambda_i \nabla h_i(u)\|$. Repeating this process generates iterates (u, λ) converging quadratically to the solution pair $(\bar{u}, \bar{\lambda})$.

Acknowledgments

References

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

[2] Bonnans J (1994) Local analysis of Newton-type methods for variational inequalities and nonlinear programming. *Appl. Math. Optim.* 29:161–186.

[3] Burke J (1990) On the identification of active constraints. II. The nonconvex case. *SIAM J. Numerical Anal.* 27:1081–1103.

[4] Burke J, Moré J (1988) On the identification of active constraints. *SIAM J. Numerical Anal.* 25:1197–1211.

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

[6] Dontchev A, Rockafellar R (2009) *Implicit Functions and Solution Mappings* (Springer-Verlag, Berlin).

[7] Drusvyatskiy D, Lewis A (2012) Optimality, identifiability, and sensitivity. Preprint, submitted July 27, https://arxiv.org/abs/1207.6628.

[8] Drusvyatskiy D, Lewis A (2014) Optimality, identifiability, and sensitivity. *Math. Programming* 147:467–498.

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

[10] Facchinei F, Pang JS (2003) *Finite Dimensional Variational Inequalities and Complementarity Problems* (Springer-Verlag, New York).

[11] Facchinei F, Fischer A, Kanzow C (1999) On the accurate identification of active constraints. *SIAM J. Optim.* 9:14–32.

[12] Fadili J, Malick J, Peyré G (2018) Sensitivity analysis for mirror-stratifiable convex functions. *SIAM J. Optim.* 28:2975–3000.

[13] Ferris M (1991) Finite termination of the proximal point algorithm. *Math. Programming Series A* 50(3):359–366.

[14] Flåm S (1992) On finite convergence and constraint identification of subgradient projection methods. *Math. Programming* 57:427–437.

[15] Gfrerer H, Outrata J (2019) On a semismooth* Newton method for solving generalized equations. Preprint, submitted April 19, https://arxiv.org/abs/1904.09167.

[16] Hare W, Sagastizábal C (2009) Computing proximal points of nonconvex functions. *Math. Programming* 116:221–258.

[17] Izmailov A, Solodov M (2014) *Newton-Type Methods for Optimization and Variational Problems* (Springer, Cham, Switzerland).

[18] Izmailov A, Solodov M (2015) Newton-type methods: A broader view. *J. Optim. Theory Appl.* 164:577–620.

[19] Klatte D, Kummer B (2002) *Nonsmooth Equations in Optimization* (Kluwer, Dordrecht, Netherlands).

[20] Lee J (2003) *Introduction to Smooth Manifolds* (Springer, New York).

[21] Lee S, Wright S (2012) Manifold identification in dual averaging for regularized stochastic online learning. *J. Machine Learn. Res.* 13:1705–1744.

[22] Lemaréchal C, Sagastizábal C (1997) Practical aspects of the Moreau-Yosida regularization: Theoretical preliminaries. *SIAM J. Optim.* 7:367–385.

[23] Lemaréchal C, Oustry F, Sagastizábal C (2000) The U-Lagrangian of a convex function. *Trans. Amer. Math. Soc.* 352:711–729.

[24] Lewis A (2003) Active sets, nonsmoothness, and sensitivity. *SIAM J. Optim.* 13:702–725.

[25] Lewis A, Liang J (2018) Partial smoothness and constant rank. Preprint, submitted July 9, https://arxiv.org/pdf/1807.03134.pdf.

[26] Lewis A, Wright S (2016) A proximal method for composite minimization. *Math. Programming* 158:501–546.

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

[28] Liang J (2016) Convergence rates of first-order operator splitting methods. PhD thesis, University of Caen Normandie, Caen, France.

[29] Liang J, Fadili J, Peyré G (2014) Local linear convergence of forward–backward under partial smoothness. Preprint, submitted July 21, https://arxiv.org/abs/1407.5611v6.

[30] Liang J, Fadili J, Peyré G (2017) Activity identification and local linear convergence of forward-backward-type methods. *SIAM J. Optim.* 27:408–437.

[31] Liang J, Fadili J, Peyré G (2018) Local linear convergence analysis of primal-dual splitting methods. *Optimization* 67:821–853.

[32] Liang J, Fadili J, Peyré G, Luke R (2015) Activity identification and local linear convergence of Douglas-Rachford/ADMM under partial smoothness. Aujol JF, Nikolova M, Papadakis N, eds. *Scale Space and Variational Methods in Computer Vision*, vol. 9087 (Lecture Notes in Computer Science, Springer International Publishing, Cham, Switzerland), 642–653.

[33] Mifflin R, Sagastizábal C (2002) Proximal points are on the fast track. *J. Convex Anal.* 9:563–579.

[34] Mifflin R, Sagastizábal C (2003) Primal-dual gradient structured functions: Second-order results; links to epi-derivatives and partly smooth functions. *SIAM J. Optim.* 13:1174–1194.

[35] Mifflin R, Sagastizábal C (2004) VU-smoothness and proximal point results for some nonconvex functions. *Optim. Methods Software* 19:463–478.

[36] Mifflin R, Sagastizábal C (2005) A VU-algorithm for convex minimization. *Math. Programming* 104:583–608.

[37] Miller S, Malick J (2005) Newton methods for nonsmooth convex minimization: Connections among U-Lagrangian, Riemannian Newton and SQP methods. *Math. Programming* 104:609–633.

[38] Mordukhovich B (1992) *Sensitivity Analysis in Nonsmooth Optimization. Theoretical Aspects of Industrial Design (Wright-Patterson Air Force Base, OH, 1990)* (SIAM, Philadelphia), 32–46.

[39] Mordukhovich B (2006) *Variational Analysis and Generalized Differentiation I. Basic Theory*, vol. 330 (Grundlehren der mathematischen Wissenschaften, Springer, Berlin).

[40] Nocedal J, Wright S (2006) *Numerical Optimization*, 2nd ed. (Springer, New York).

[41] Rockafellar R, Wets RB (1998) *Variational Analysis*, vol. 317 (Grundlehren der mathematischen Wissenschaften, Springer, Berlin).

[42] Sagastizábal C (2018) A VU-point of view of nonsmooth optimization. Sirakov B, Ney de Souza P, Viana M, eds. *Proc. Internat. Congress Mathematicians, Rio de Janeiro, Brazil* (World Scientific, Singapore), 3785–3806.

[43] Vaiter S, Peyré G, Fadili J (2018) Model consistency of partly smooth regularizers. *IEEE Trans. Inform. Theory* 64:1725–1737.

[44] Vaiter S, Deledalle C, Fadili J, Peyré G, Dossal C (2017) The degrees of freedom of partly smooth regularizers. *Ann. Inst. Statist. Math.* 69:791–832.

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