JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 124, No. 3, pp. 673–699, March 2005 (© 2005)
DOI: 10.1007/s10957-004-1180-1

Implicit Smoothing and Its Application to Optimization with Piecewise Smooth Equality Constraints[1]

D. RALPH[2] AND H. XU[3]

Communicated by Z. Q. Luo

Abstract. In this paper, we discuss the smoothing of an implicit function defined by a nonsmooth underdetermined system of equations $F(y, z) = 0$. We apply a class of parametrized smoothing methods to smooth F and investigate the limiting behavior of the implicit function solving the smoothed equations. In particular, we discuss the approximation of the Clarke generalized Jacobian of the implicit function when F is piecewise smooth. As an application, we present an analysis of the generalized Karush-Kuhn-Tucker conditions of different forms for a piecewise-smooth equality-constrained minimization problem.

Key Words. Clarke generalized Jacobian, B-subdifferential, smoothing Jacobians, generalized Karush-Kuhn-Tucker conditions, piecewise smooth functions, strong Jacobian consistency, index-consistent functions, essentially index consistent functions, essentially active indices.

1. Introduction

In nonsmooth optimization, we may encounter an underdetermined system of nonsmooth equations in the constraints of the problem

$$F(y, z) = 0, \text{ where } F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \text{ is locally Lipschitz.} \tag{1}$$

[1]This work was supported by the Australian Research Council. It was carried out mainly while the authors were working at the Department of Mathematics and Statistics, University of Melbourne, Melbourne, Australia. The authors appreciate the helpful comments of two anonymous referees.
[2]Reader, Judge Institute of Management Studies, University of Cambridge, Cambridge, UK.
[3]Lecturer, School of Mathematics, University of Southampton, Highfield, Southampton, UK.

For instance, in a mathematical program with equilibrium constraints (MPEC, Refs. 1–2), a variational inequality constraint can be rewritten as a nonsmooth system of the form (1). Similarly, in bilevel programming (BP), the upper level variables are related often to the lower variables through a nonsmooth underdetermined system; see Ref. 3 and the references therein. In MPEC, some smoothing nonlinear complementarity program (NCP) functions are used to deal with the nonsmooth underdetermined constraints; see Ref. 4 and the references therein, which show the potential of smoothing techniques in dealing with nonsmooth equality constraints in MPEC.

In this paper, we discuss the smooth approximations up to first order of the implicit function defined as in (1) by implicitly smoothing F. Our purpose is twofold:

(a) We smooth the implicit solution function defined by solving (1) in y for fixed z by smoothing F. It turns out that, under mild conditions, a smoothing function which approximates the implicit solution can be obtained by solving the smoothed system of equations corresponding to (1) with F replaced by its smoothing. Moreover, we can establish a relationship between the generalized Jacobians (including the Clarke generalized Jacobian, B-subdifferential) of the implicit solution function, the smoothing implicit function, and the smoothing of F, and consequently give some inclusions that bound the Clarke generalized Jacobian of the implicit solution function of (1). The results are strengthened in the circumstance of the piecewise smoothness of F.

(b) Using the established results, we discuss the generalized Karush-Kuhn-Tucker (GKKT) conditions associated with different forms of an optimization problem,

$$\min f(y,z), \quad \text{s.t. } F(y,z)=0, \tag{2}$$

where f is smooth and F is piecewise smooth, which include a smoothing form, a perturbed form, and implicit variants of these. The implicit programming form is of particular interest. To make use of this formulation, we assume that the constraint $F(y,z)=0$ is equivalent, for (y,z) near a give feasible point $(\bar{y},\bar{z})$, to $y=y(z)$ for some Lipschitz function $y(\cdot)$ of z near $\bar{z}$. Then, the optimization problem is equivalent, at least for (y,z) near $(\bar{y},\bar{z})$, to $\min_z f(y(z),z)$.

We note in brief our conclusions regarding optimality conditions based on the Clarke calculus for the different reformulations of the problem (2). On the one hand, the implicit programming formulation $\min_z f(y(z),z)$ apparently has the tightest optimality condition, whereas the other formulations, including (2), have somewhat looser optimality conditions that are roughly equivalent. On the other hand, the optimality condition of the implicit programming approach relies on finding an appropriate element of the Clarke generalized Jacobian of the implicit function $y(\cdot)$ at the given point z, which seems difficult to do in practice; see Ref. 3 for investigation of this question in the context of the parametric nonlinear programming. The optimality condition for (2) uses the generalized Jacobian of F at a given point $(y(\bar{z}), \bar{z})$, which is somewhat easier to calculate than $\partial y(\bar{z})$, though still difficult in general.

Throughout this paper, we use the following notation. D_F denotes the set of points where the function F is Fréchet differentiable and ∇F denotes the usual Jacobian. For sequences, we denote by $\overline{\lim}$ the outer limit of a sequence, that is, the set of all the accumulation points of the sequence. In particular,

$$\partial_B F(x) = \overline{\lim_{\substack{y \in D_F \\ y \to x}}} \nabla F(y)$$

defines the B-subdifferential (Ref. 5) of F at x; the convex hull of $\partial_B F$, denoted by $\partial F(x)$, defines the Clarke generalized Jacobian (Ref. 6) of F at x. For a set-valued mapping $\mathcal{A}: \mathbb{R}^n \times \mathbb{R}^m \to 2^{\mathbb{R}^{n \times (n+m)}}$, we use $\pi_y \mathcal{A}(y, z)$ to denote the set of all $n \times n$ matrices M such that, for some $n \times m$ matrix N, the $n \times (n+m)$ matrix $[MN]$ belongs to $\mathcal{A}(y, z)$. Note that occasionally a matrix $[MN] \in \mathbb{R}^{n \times (n+m)}$, where $M \in \mathbb{R}^{n \times n}$ and $N \in \mathbb{R}^{n \times m}$, will be written as $[M, N]$ to ensure clearity. The 2-norms of a vector and a matrix will be denoted by $\|\cdot\|$. The notation $\epsilon \downarrow 0$ means that the positive scalar ϵ tends 0. Finally, we need some notation for sets: we denote by int Ω, cl Ω, and conv Ω respectively the interior, closure, and convex hull of a set Ω. Also, $B(x, \delta)$ denotes the closed ball of center x and radius $\delta > 0$; $\mathcal{N}(x)$ denotes a more general neighborhood of a point x.

2. Preliminaries

Let $F: \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz function. The basic idea of smoothing F is to find a continuously parametrized differentiable function $F(\cdot, \epsilon)$ which approximates F as the parameter ϵ tends to zero. A variety of smoothing methods (functions) have appeared in the past few years; see Ref. 7 for a list of references. Instead of focusing on a particular smoothing function, here we consider a class of smoothing functions with properties enjoyed by the most known smoothing functions.

Definition 2.1. Let $F:\mathbb{R}^n\to\mathbb{R}^m$ be a locally Lipschitz function. A smoothing of F is any function $\hat{F}:\mathbb{R}^n\times\mathbb{R}\to\mathbb{R}^m$ that satisfies the following conditions:

(a) for every $x\in\mathbb{R}^n$, $\hat{F}(x,0)=F(x)$;
(b) for every $x\in\mathbb{R}^n$, $\hat{F}$ is locally Lipschitz at $(x,0)$;
(c) $\hat{F}$ is continuously differentiable on $\mathbb{R}^n\times(\mathbb{R}\setminus\{0\})$.

We call $\hat{F}$ a strict smoothing if, in addition,

(d) for every $x\in\mathbb{R}^n$, if F is strictly differentiable at x, that is, $\partial F(x)=\{\nabla F(x)\}$, then

$$\pi_x\partial_B\hat{F}(x,0)=\{\nabla F(x)\}.$$

Note that a continuously differentiable function is strictly differentiable; in the general case when F is Lipschitz near x but not strictly differentiable at $x, \pi_x\partial_B\hat{F}(x,0)$ is not a singleton. Consequently, we may relate $\pi_x\partial_B\hat{F}(x,0)$ to $\partial F(x)$ or $\partial_B F(x)$. Moreover, it is easy to observe that

$$\overline{\lim_{\substack{|\epsilon|\downarrow 0\\ x'\to x}}}\nabla_x\hat{F}(x',\epsilon)\subset\pi_x\partial_B\hat{F}(x,0),$$

which implies that $\pi_x\partial_B\hat{F}(x,0)$ contains all the information regarding the limiting behavior of $\nabla_x\hat{F}(x',\epsilon)$ as $x\to x$ and $|\epsilon|\downarrow 0$. Before further discussion in this regard, we present some important smoothing instances.

Example 2.1. Let $F:\mathbb{R}^n\to\mathbb{R}^m$ be a locally Lipschitz function. For every $x\in\mathbb{R}^n$ and $\epsilon\in\mathbb{R}$, let

$$\hat{F}(x,\epsilon)=\int_{\mathbb{R}^n}F(x-\epsilon u)\Phi(u)du, \tag{3}$$

where $\Phi:\mathbb{R}^n\to\mathbb{R}$ is a kernel or probability density function; that is, $\Phi(u)\ge 0$ for all $u\in\mathbb{R}^n$ and

$$\int_{\mathbb{R}^n}\Phi(u)du=1.$$

There are many discussions on $\hat{F}(x,\epsilon)$ with $\epsilon\ge 0$; see for instance Ref. 8. Here, we allow ϵ to take a negative value only for convenience. In particular, if

$$\Phi(u)=\begin{cases}1, & \text{if } \|u\|_\infty\le 1,\\ 0, & \text{otherwise,}\end{cases}$$

where $\|\cdot\|_\infty$ denotes the infinity norm, then

$$\hat{F}(x,\epsilon)=\begin{cases}[1/(2\epsilon)^n]\int_{x_1-\epsilon}^{x_1+\epsilon}\cdots\int_{x_n-\epsilon}^{x_n+\epsilon}F(u)du, & \text{if } \epsilon\neq 0\\ F(x), & \text{if } \epsilon=0.\end{cases} \tag{4}$$

The following is straightforward from Ref. 9, Theorem 3.1.

Remark 2.1. Let $F:\mathbb{R}^n\to\mathbb{R}^m$ be a locally Lipschitz function and let $\hat{F}(x,\epsilon)$ be defined either by (4) or by (3), where Φ is continuously differentiable and the support set $\{u\in\mathbb{R}^n:\Phi(u)>0\}$ is bounded. Then, $\hat{F}(x,\epsilon)$ is a smoothing of F.

Throughout this paper, we consider only the case that Φ is continuously differentiable and the support set $\{u\in\mathbb{R}^n:\Phi(u)>0\}$ is bounded. We call such an integral smoothing method (3) a bounded integral smoothing. In particular, (4) is known as adaptive smoothing (Refs. 10, 11).

Let $\mathcal{A}$ be a subset of $\mathbb{R}^{n\times m}$. The plenary hull of $\mathcal{A}$ is defined as

$$\text{plen } \mathcal{A}=\{A\in\mathbb{R}^{n\times m}:Aa\in\mathcal{A}a,\forall a\in\mathbb{R}^m\},$$

where

$$\mathcal{A}a=\{Aa:A\in\mathcal{A}\}.$$

The notion of plenary hull was introduced by Sweetser (Ref. 12) and was discussed further in Refs. 13, 14, etc. Obviously, if $\mathcal{A}$ is convex, then plen $\mathcal{A}$ is also convex.

Let $F:\mathbb{R}^n\to\mathbb{R}^m$ be a locally Lipschitz function and let $\hat{F}(x,\epsilon)$ be a bounded integral smoothing function (3) or the adaptive smoothing (4). Based on the above motivations and Ref. 9, Theorem 3.1, we can establish a direct relationship between $\partial\hat{F}(x,0)$ and $\partial F(x)$,

$$\pi_x\partial_B\hat{F}(x,0)\subset\text{plen }\partial F(x). \tag{5}$$

Due to the convexity of set plen $\partial F(x)$, (5) implies

$$\pi_x\partial\hat{F}(x,0)\subset\text{plen }\partial F(x).$$

A slightly stronger statement than (5), but one which is not generally valid, is that

$$\pi_x\partial_B\hat{F}(x,0)\subset\partial F(x). \tag{6}$$

We shall see in Remark 4.2 that, if F is a piecewise smooth function and $\hat{F}$ is an adaptive integral smoothing of F, then (6) holds. It can also be easily checked that all smoothing functions in Jiang and Ralph (Ref. 4, Section 7) satisfy (6).

We say a smoothing $\hat{F}(x,\epsilon)$ satisfies the strong Jacobian consistency if (6) holds. The word "strong" is used to distinguish it from a similar notion, called the Jacobian consistency, introduced by Chen, Qi, and Sun (Ref. 15). A smoothing function $\hat{F}(x,\epsilon)$ of $F(x)$ is said to satisfy the Jacobian consistency if, for every $x\in\mathbb{R}^n$,

$$\lim_{|\epsilon|\downarrow 0}\ \min_{V\in\partial_C F(x)} \|\nabla_x\hat{F}(x,\epsilon)-V\|=0,$$

or equivalently,

$$\overline{\lim_{|\epsilon|\downarrow 0}}\ \nabla_x\hat{F}(x,\epsilon)\subset\partial_C F(x),$$

where

$$\partial_C F(x)=\partial F_1(x)\times\cdots\times\partial F_m(x).$$

Actually in Ref. 15, ϵ is restricted to take nonnegative values, but this is only a matter of notation.

We introduce another Jacobian consistency notion. We say that a smoothing $\hat{F}(x,\epsilon)$ of F satisfies the strong B-subdifferential consistency at x if

$$\pi_x\partial_B\hat{F}(x,0)=\partial_B F(x). \tag{7}$$

Obviously, the strong B-subdifferential consistency implies the strong Jacobian consistency.

Since

$$\overline{\lim_{|\epsilon|\downarrow 0}}\ \nabla_x\hat{F}(x,\epsilon)\subset\overline{\lim_{\substack{|\epsilon|\downarrow 0\\ y\to x}}}\ \nabla_x\hat{F}(y,\epsilon)\subset\pi_x\partial_B\hat{F}(x,0),$$

and since $\partial F(x)\subset\partial_C F(x)$, the strong Jacobian consistency implies the Jacobian consistency. When F is strictly differentiable at x, both $\partial_C F(x)$ and $\partial F(x)$ become singletons and strong Jacobian consistency is equivalent to Jacobian consistency at x. Note that the bounded integral smoothing does not generally satisfy (7).

3. Implicit Smoothing of a Locally Lipschitz Function

In this section, we discuss the smoothing of the implicit function defined as in (1) under the condition that F is locally Lipschitz.

Consider the underdetermined system of equations (1). Clarke (Ref. 6) obtained the following (we combine Theorem 7.1.1 and the subsequent corollary of Ref. 6).

Theorem 3.1. Let F be given as in (1) and let $(\bar{y}, \bar{z}) \in \mathbb{R}^n \times \mathbb{R}^m$ be such that $F(\bar{y}, \bar{z}) = 0$. Suppose that $\pi_y \partial F(\bar{y}, \bar{z})$ is nonsingular. Then, there exist neighborhoods Z of $\bar{z}, Y$ of $\bar{y}$, and a locally Lipschitz function $y: Z \to Y$ such that $y(\bar{z}) = \bar{y}$ and, for every $z \in Z$, $y = y(z)$ is the unique solution of the problem $F(y, z) = 0$, $y \in Y$.

In the situation of the previous theorem, up to choice of the neighborhoods Y and Z, $y(\cdot): Z \to Y$ is the unique function satisfying

$$F(z, y(z)) = 0, \qquad \text{for } z \in Z.$$

The function $y(\cdot)$ can be assumed to be (globally) Lipschitz by choosing Z to be closed and bounded, because locally Lipschitz functions are (globally) Lipschitz on compact sets.

In the sequel, where we refer to $(\bar{y}, \bar{z})$, we assume tacitly that $(\bar{y}, \bar{z}) \in \mathbb{R}^n \times \mathbb{R}^m$ satisfies $F(\bar{y}, \bar{z}) = 0$. Here, we shall apply a smoothing method of F defined as in (1) and investigate the limiting behavior of the implicit function determined by the smoothed systems of equations

$$\hat{F}(y, z, \epsilon) = 0. \tag{8}$$

We discuss first the existence of a smoothing of the implicit solution function of (1).

Theorem 3.2. Let $F(y, z)$ be defined as in (1), let $\hat{F}(y, z, \epsilon)$ be a smoothing of $F(y, z)$. Assume that $\pi_y \partial \hat{F}(\bar{y}, \bar{z}, 0)$ in nonsingular. Then:

- (a) there exist neighborhoods Z of $\bar{z}$ and $\mathcal{E}$ of $0 (\in \mathbb{R})$ and a unique Lipschitz function $\hat{y} : Z \times \mathcal{E} \to \mathbb{R}^n$ such that, for every $z \in Z, \epsilon \in \mathcal{E}$, $\hat{F}(\hat{y}(z, \epsilon), z, \epsilon) = 0$;
- (b) for every $z \in Z, \hat{y}(z, 0) = y(z)$, where $y : Z \to \mathbb{R}^n$ is the unique Lipschitz function satisfying $y(\bar{z}) = \bar{y}$ and $F(y(z), z) = 0$;
- (c) $\hat{y}$ is continuously differentiable on $Z \times (\mathcal{E} \backslash \{0\})$.

Proof. Part (a). Since $\pi_y \partial \hat{F}(\bar{y}, \bar{z}, 0)$ is nonsingular, by applying Theorem 3.1 to $\hat{F}(y, x, \epsilon)$ at $(\bar{y}, \bar{z}, 0)$, we know that there exist neighborhoods Z of $\bar{z}$ and $\mathcal{E}$ of $0 (\in \mathbb{R})$ and a unique Lipschitz function $\hat{y} : Z \times \mathcal{E} \to \mathbb{R}^n$ such that, for every $z \in Z, \epsilon \in \mathcal{E}$, $\hat{y}(z, \epsilon)$ satisfies (8).

Part (b). Since $\partial_y F(\bar{y}, \bar{z})$ is clearly contained in $\pi_y \partial \hat{F}(y, x, 0)$, we have the Lipschitz implicit function $y(\cdot)$ as given by Theorem 3.1. By Definition 2.1 (a) and the uniqueness of $y(\cdot)$, we obtain the equation $\hat{y}(z, 0) = y(z)$.

Part (c). By Definition 2.1 (c), $\hat{F}$ is continuously differentiable on $\mathbb{R}^n \times \mathbb{R}^m \times (\mathbb{R} \backslash \{0\})$. The conclusion follows from Part (a) and the classical implicit function theorem. □

Theorem 3.2 shows that a smooth approximation of the implicit function defined as in (1) can be obtained by smoothing F and solving the smoothed system of equations (8). Referring to Definition 2.1, we see that $\hat{y}(\cdot, \cdot)$ is indeed a smoothing of $y(\cdot)$ in a neighborhood of $\bar{z}$. We call $\hat{y}(\cdot, \cdot)$ an implicit smoothing function.

We now investigate the Clarke generalized Jacobian of $\hat{y}(\cdot, \cdot)$ at $(z, 0)$. We intend to establish relationships between such a Jacobian, $\partial y(z)$, and $\partial F(y(z), z)$. For this purpose, we investigate first the Clarke generalized Jacobian of a Lipschitz function where some variables are fixed. The following results establish a relationship between such a Jacobian and the Clarke generalized Jacobian of the function through plenary hulls.

Proposition 3.1. Let $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ be a locally Lipschitz function and let $g(\cdot) = G(\cdot, 0)$. Then, $g : \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitz such that, for every $x \in \mathbb{R}^n$, $\partial g(x) \subset \text{plen}\, \pi_x \partial G(x, 0)$.

Proof. Obviously, g is locally Lipschitz. Let x' be a point in a neighborhood of x at which g is differentiable. It is easy to check that, for every fixed d and v in $\mathbb{R}^n$,

$$v^T \nabla g(x') d \leq (v^T G)^\circ((x', 0); (d, 0)),$$

where $(v^T G)^\circ$ is the Clarke generalized derivative (Ref. 6). Thus,

$$\nabla g(x') d \in \partial G(x', 0)(d^T, 0)^T = [\pi_x \partial G(x', 0)] d.$$

Since d is arbitrary, by the definition of plenary hull, we have

$$\nabla g(x') \in \text{plen}\, \pi_x \partial G(x', 0).$$

By definition, $\partial g(x)$ is the convex hull of all the accumulation matrices of $\nabla g(x')$, where $x' \in D_g$ (the set of points at which g is differentiable) and $x' \to x$. By the upper semicontinuity of plen $\pi_x \partial G(\cdot, 0)$ at x, we obtain the proof. □

As an example, we can apply the above proposition to a smoothing function.

Remark 3.1. Suppose that $\hat{F}(x, \epsilon)$ is a bounded integral smoothing defined by (3) or the adaptive smoothing defined by (4). From (5) and Proposition 3.1, we have

$$\text{plen}\, \pi_x \partial \hat{F}(x, 0) = \text{plen}\ \ \partial F(x). \tag{9}$$

In general, if $\hat{F}(x, \epsilon)$ is a smoothing of F and satisfies the strong B-subdifferential consistency or the strong Jacobian consistency, then (9) holds.

We say that a function $F : \mathbb{R}^n \to \mathbb{R}^n$ is a Lipschitz homeomorphism near x if there exist neighborhoods $\mathcal{N}(x)$ and $\mathcal{N}(F(x))$ of x and $F(x)$ respectively such that F establishes a Lipschitz homeomorphism between the two sets; that is, $F|_{\mathcal{N}(x)} : \mathcal{N}(x) \to \mathcal{N}(F(x))$ is Lipschitz, invertible, and its inverse is also Lipschitz.

Proposition 3.2. Let $F:\mathbb{R}^n \to \mathbb{R}^n$ be a Lipschitz homeomorphism near x. Then, $\partial_B F(x)$ is invertible and

$$\partial_B G(F(x)) = (\partial_B F(x))^{-1}, \tag{10}$$

where G is the inverse function of F near x.

Proof. The details are fairly standard; see Ref. 7. □

Let $F(y, z)$ be defined as in (1) and let $\hat{F}(y, z, \epsilon)$ be a smoothing of $F(y, z)$. In subsequent discussions, we use the following notation: an element $[MNb]$ of the generalized Jacobian $\partial \hat{F}(y, z, 0)$ or the B-subdifferential $\partial_B \hat{F}(y, z, 0)$ and an element $[MN]$ of the B-subdifferential $\partial_B F(y, z)$ or the Clarke generalized Jacobian $\partial F(y, z)$ or its plenary hull plen $\partial F(y, z)$, have $M \in \mathbb{R}^{n \times n}$, $N \in \mathbb{R}^{n \times m}$, and $b \in \mathbb{R}^n$.

Theorem 3.3. Let $F(y, z)$ be defined as in (1) and let $\hat{F}(y, z, \epsilon)$ be a smoothing of $F(y, z)$. Suppose that the conditions of Theorem 3.2 are satisfied. Let $\hat{y} : Z \times \mathcal{E} \to \mathbb{R}^n$ be the unique implicit smoothing function given by Theorem 3.2. Then, for all $z \in Z$,

$$\partial y(z) \subset \text{plen } \pi_z \partial \hat{y}(z, 0), \tag{11}$$

$$\partial \hat{y}(z, 0) \subset \text{plen conv } \{-M^{-1}[N, b] : [MNb] \in \partial_B \hat{F}(y(z), z, 0)\}, \tag{12}$$

$$\pi_z \partial \hat{y}(z, 0) \subset \text{plen conv } \{-M^{-1}N : [MNb] \in \partial_B \hat{F}(y(z), z, 0)\}. \tag{13}$$

Proof. The main idea is to introduce the function

$$G(y, z, \epsilon) = (\hat{F}(y, z, \epsilon), z, \epsilon)^T,$$

where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, and $\epsilon \in \mathbb{R}$, and observe that its generalized Jacobian at $(\bar{y}, \bar{z}, 0)$ is nonsingular; hence, the generalized inverse function theorem (Ref. 6, Theorem 7.1.1) promises that G is a Lipschitzian homeomorphism near $(\bar{y}, \bar{z}, 0)$. The result follows by examining the local inverse of G; see Ref. 7 for details. □

We note that, when $\hat{F}$ is a bounded integral smoothing of F, we have, from (5) and (13),

$$\pi_z \partial \hat{y}(z, 0) \subset \text{plen conv } \{-M^{-1}N : [MN] \in \text{plen } \partial F(y(z), z)\}.$$

From Theorem 3.3, we can deduce some properties of $y(\cdot)$ and $\hat{y}(\cdot, \cdot)$.

Corollary 3.1. Let $\hat{F}(y, z, \epsilon)$ be a smoothing of $F(y, z)$. Suppose that $\pi_y \partial F(\bar{y}, \bar{z})$ is nonsingular. Then:

(a) $\hat{y}(z, \epsilon)$ is Lipschitz near $(\bar{z}, 0)$ of modulus $L_{\hat{y}}$ for any $L_{\hat{y}}$ greater than $\max\{\|M^{-1}[N, b]\| : [MNb] \in \partial_B \hat{F}(\bar{y}, \bar{z}, 0)\}$. Similarly, $y(z)$ is Lipschitz of modulus $L_{\hat{y}}$ near $\bar{z}$.

(b) In addition, if F is strictly differentiable at $(\bar{y}, \bar{z})$ and $\hat{F}$ is a strict smoothing, then $\hat{y}(\cdot, \cdot)$ is strictly differentiable at $(\bar{z}, 0)$ and $\pi_z \hat{y}(\bar{z}, 0) = \nabla y(\bar{z}) = -M^{-1}N$, where $[MN] = \nabla F(\bar{y}, \bar{z})$.

Proof. Part (a) can be obtained easily by using the mean-value theorem (Ref. 6, Proposition 2.6.5) and Theorem 3.3 Part (b) follows from Theorem 3.3 and Part (d) of Definition 2.1. □

The set $\partial_B \hat{F}(y(z), z, 0)$ plays an important role in the estimation of $\partial \hat{y}(z, 0)$. Of particular interest is the case when

$$\partial y(z) = \pi_z \text{conv}\{-M^{-1}[N, b] : [MNb] \in \partial_B \hat{F}(y(z), z, 0)\}, \tag{14}$$

which yields, according to Theorem 3.3,

$$\text{plen } \partial y(z) = \text{plen} \, \pi_z \partial \hat{y}(z, 0).$$

The latter establishes an exact relationship between $\partial y(z)$ and the generalized Jacobian $\hat{y}(z,0)$, which implies that every accumulation matrix of the sequence $\{\nabla y_\epsilon(z)\}$ is contained in plen $\partial y(z)$. Unfortunately, (14) cannot be satisfied easily. Indeed, the relation relies on not only the structure of F, but also the smoothing method. We will discuss this in the next section; see Proposition 4.2 and Theorem 4.2.

4. Implicit Smoothing of a Piecewise Smooth Function

In this section, we will investigate the implicit smoothing of piecewise smooth functions and try to obtain stronger results than Theorem 3.3 for example.

We discuss first the structure of the generalized Jacobians, in particular the B-subdifferential, of the implicit function defined via (1) in the circumstance of piecewise smoothness.

Recall that a function $F:\mathbb{R}^n \to \mathbb{R}^m$ is said to be piecewise smooth (PC^1 for short), on an open set $U \subset \mathbb{R}^n$, if it is continuous and there exists a finite family of C^1 functions $F^i:\mathbb{R}^n \to \mathbb{R}^m, i=1,\ldots,l,$ such that, for every $x \in U, F(x)=F^i(x)$ for at least one index $i \in \{1,\ldots,l\}$. We say that F is a continuous selection of $\{F^i{:}i=1,\ldots,l\}$. We denote by $\mathcal{I}_F(x)$ the sets of indices i such that $F(x)=F^i(x)$. For convenience, when a point is represented by two arguments, say (y,z), we will use $\mathcal{I}_F(y,z)$ rather than $\mathcal{I}_F((y,z))$.

Piecewise smooth functions form a very important class of functions. In nonsmooth equations and nonsmooth optimization, many problems involve only PC^1 functions; see Ref. 16 for analysis and applications.

Proposition 4.1. Let $F:\mathbb{R}^n \to \mathbb{R}^m$ be PC^1. Let $c \in \mathbb{R}^m$ and $\Omega \subset F^{-1}(c)$. Then, there exists a relatively open, dense subset Ω' of Ω with the property that, for every $y \in \Omega'$, there exists a neighborhood $\mathcal{N}_\Omega(y)$ of y relative to Ω such that $\mathcal{I}_F(\cdot)$ is constant on $\mathcal{N}_\Omega(y)$.

Proof. Define Ω' as a set of points $x \in \Omega$ such that, for all x' in a relative neighborhood $\mathcal{N}_\Omega(x), \mathcal{I}_F(x')=\mathcal{I}_F(x)$; clearly, Ω' is open relative to Ω. Take $x \in \Omega\backslash\Omega'$ and any relative neighborhood $\mathcal{N}_\Omega(x)$ such that

$$\mathcal{I}_F(x') \subset \mathcal{I}_F(x), \qquad \text{for all } x' \in \mathcal{N}_\Omega(x).$$

Since $x \notin \Omega'$, there exists a point $x' \in \mathcal{N}_\Omega(x)$ and $i \in \mathcal{I}_F(x)$ such that $\mathcal{I}_F(x') \subset \mathcal{I}_F(x)\backslash\{i\}$. Then, there exists a relative neighborhood $\mathcal{N}_\Omega(x')$ of x' such that $\mathcal{N}_\Omega(x') \subset \mathcal{N}_\Omega(x)$ and, for all $y' \in \mathcal{N}_\Omega(x')$,

$$\mathcal{I}_F(y') \subset \mathcal{I}_F(x)\backslash\{i\}.$$

Since $\mathcal{I}_F(x)$ is finite, by repeating the analysis with the set $\mathcal{I}_F(x')$ and so on, we can obtain finally a point y in $\mathcal{N}_\Omega(x)$ such that $\mathcal{I}_F(\cdot)$ is constant on a relative neighborhood $\mathcal{N}_\Omega(y)$ of y and $\mathcal{N}_\Omega(y) \subset \mathcal{N}_\Omega(x)$. By definition, $y \in \Omega'$. Note that $\mathcal{N}_\Omega(x)$ can be arbitrarily small. This proves $x \in \text{cl } \Omega'$. □

Remark 4.1. Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ be PC^1 and let the hypotheses of Theorem 3.1 hold. In later discussions, we will be interested particularly in taking Ω as the local graph of the implicit function $y:Z \to Y$, that is,

$$\Omega = \{(y(z), z) : z \in Z\}.$$

In this case, Proposition 4.1 guarantees that, for any $z \in Z$, there exists a z' arbitrarily close to z such that $\mathcal{I}_F(\cdot)$ is constant on a relative neighborhood $\mathcal{N}_\Omega(y(z'), z')$. It follows without difficulty that $y(\cdot)$ is differentiable at z' (indeed on a neighborhood of z'); see Proposition 4.2 (b).

Definition 4.1. Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be PC^1. We say that F is index- consistent or i-consistent at a point $x \in \mathbb{R}^n$ with respect to a set $\mathcal{S} \subset \mathbb{R}^n$ containing x if $\mathcal{I}_F(\cdot)$ is constant on $\mathcal{S} \cap B(x, \delta)$ for some $\delta > 0$. Denote by $IC_F(\mathcal{S})$ the set of points in $\mathcal{S}$ at which F is i-consistent with respect to $\mathcal{S}$. We say that F is essentially index-consistent or e-i-consistent at x with respect to a set $\mathcal{S}$ if $\mathcal{I}_F(x) = \bar{\mathcal{I}}_F(x, S)$, where we define

$$\bar{\mathcal{I}}_F(x, S) = \overline{\lim_{\substack{x' \in IC_F(\mathcal{S}) \\ x' \to x}}} \mathcal{I}_F(x').$$

It is clear that i-consistency implies e-i-inconsistency. Consider the case when $F : \mathbb{R}^n \to \mathbb{R}^m$ is i-consistent at a point $x \in \mathbb{R}^n$ with respect to a neighborhood of x. By definition, $\mathcal{I}_F(\cdot)$ is constant near x. This implies that all the functions $F^i, i \in \{1, \ldots, l\}$, coincide with each other and with F near x, hence that F is continuously differentiable in a neighborhood of x. A slightly more general but perhaps more interesting case occurs when F is e-i-consistent at x with respect to $\mathbb{R}^n$. In such a case, we can no longer derive the differentiability of F at x. However, we can see that $\bar{\mathcal{I}}_F(x, \mathbb{R}^n)$ is precisely the essentially active index set (Ref. 17) of F at x, defined by

$$\mathcal{I}_F^e(x) = \{i : x \in \text{cl int } \{x' \in \mathbb{R}^n : i \in \mathcal{I}_F(x')\}\};$$

hence, we are able to use the former set to characterize the B-subdifferential of F at x.

Lemma 4.1. If $F:\mathbb{R}^n \to \mathbb{R}^m$ is PC^1, then for any neighborhood U of x, $\bar{\mathcal{I}}_F(x,U)=\mathcal{I}_F^e(x)$.

Proof. Let $i\in\bar{\mathcal{I}}_F(x,U)$ and let U be a neighborhood of x. By definition, there exists a sequence $x_n \to x$ such that $x_n \in IC_F(U)$ and $i \in \mathcal{I}_F(x_n)$. Hence, F is i-consistent at every x_n with respect to U. Based on our previous comments, this implies that F is continuously differentiable in a neighborhood of x_n. Obviously, such a sequence x_n must lie in the set int $\{x'\in U: i\in\mathcal{I}_F(x')\}$ and hence $i\in\mathcal{I}_F^e(x)$.

Conversely, let $i\in\mathcal{I}_F^e(x)$. Then, there exists a sequence $x_n\to x$ such that

$$x_n \in \operatorname{int}\{x'\in U: i\in\mathcal{I}_F(x')\}.$$

For every x_n, there exists a neighborhood $\mathcal{N}(x_n)$ such that F coincides with F^i and is continuously differentiable on the neighborhood. Clearly, $\mathcal{I}_F(\cdot)$ is index consistent within the neighborhood. Thus, $i\in\bar{\mathcal{I}}_F(x,U)$. □

Corollary 4.1. Let $F:\mathbb{R}^n\to\mathbb{R}^m$ be PC^1. For each $x\in\mathbb{R}^n$,

$$\partial_B F(x)=\{\nabla F^i(x): i\in\bar{\mathcal{I}}_F(x,\mathbb{R}^n)\}.$$

Thus,

$$\partial_B F(x)=\{\nabla F^i(x): i\in\mathcal{I}_F(x)\},$$

if F is e-i-consistent at x with respect to $\mathbb{R}^n$.

Proof. According to Ref. 16, Lemma 2, $\partial_B F(x)$ is exactly the set of Jacobians $\nabla F^i(x)$ for essentially active indices i, giving the first equation. The second equation follows immediately under the e-i-consistency of F at x. □

Example 4.1. Consider the following function:

$$F(y,z)=\begin{cases} f^1(y,z)=y+(1/3)z, & \text{for } z\le 0,\ y\le -(1/3)z,\\ f^2(y,z)=2(y+(1/3)z), & \text{for } z\le 0,\ y\ge -(1/3)z,\\ f^3(y,z)=y+(1/2)z, & \text{for } z\ge 0,\ y\le -(1/2)z,\\ f^4(y,z)=2(y+(1/2)z), & \text{for } z\ge 0,\ y\ge -(1/2)z.\end{cases}$$

Here, $F(y,z)$ is a PC^1 function. Let

$$\mathcal{S}^1 = \{(y,z) : y = -(1/3)z, z \le 0\},$$
$$\mathcal{S}^2 = \{(y,z) : y = -(1/2)z, z \ge 0\},$$
$$\mathcal{S} = F^{-1}(0) = \mathcal{S}^1 \cup \mathcal{S}^2.$$

Then, $F(y,z)$ is i-consistent with respect to $\mathcal{S}$ at any point of $\mathcal{S}\backslash\{(0,0)\}$ and e-i-consistent at (0,0), but not i-consistent.

Example 4.2. Consider the following function:

$$F(y,z) = \begin{cases} f^1(y,z) = y + (1/2)z, & \text{for } y \ngeq |z|, \\ f^2(y,z) = z + (1/2)y, & \text{for } y \ge z \ge 0, \\ f^3(y,z) = (1/2)y, & \text{for } y \ge -z \ge 0. \end{cases}$$

Here, $F(y,z)$ is a PC^1 function. Let

$$\mathcal{S} = F^{-1}(0) = \{(y,z) : y = -(1/2)z\}.$$

Then, $F(y,z)$ is i-consistent with respect to $\mathcal{S}$ at any point of $\mathcal{S}\backslash\{(0,0)\}$. However, F is NOT e-i-consistent at (0,0). Indeed,

$$\overline{\lim_{\substack{(y,z)\in\mathcal{S} \\ (y,z)\to(0,0)}}} \mathcal{I}_F(y,z) = \{1\} \neq \{1,2,3\} = \mathcal{I}_F(0,0).$$

Consider the following nonlinear complementarity problem (NCP):

$$P(y,z) \ge 0, \quad Q(y,z) \ge 0, \quad P(y,z)^T Q(y,z) = 0.$$

where $P, Q : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ are continuously differentiable. Let

$$H(y,z) = \min(P(y,z), Q(y,z))$$

where min is taken componentwise. Suppose that $\pi_y \partial H(\bar{y},\bar{z})$ is nonsingular and that $y(\cdot)$ is the implicit function determined by $H(y,z)=0$ in a neighborhood of $(\bar{y},\bar{z})$. If complementarity is strict or nondegenerate at $(\bar{y},\bar{z})$, that is, if $P_i(\bar{y},\bar{z})=0$ implies $Q_i(\bar{y},\bar{z})>0$, then H is i-consistent at $(\bar{y},\bar{z})$ with respect to set $\{(y(z),z) : z \in Z\}$, where Z is a neighborhood of $\bar{z}$. This is due to the continuous differentiability of H at strictly complementarity points.

Proposition 4.2. Let F be a PC^1 function defined as in (1) and assume that $\pi_y \partial F(\bar{y}, \bar{z})$ is nonsingular. Let $y{:}Z \to Y$ be the associated implicit function as in Theorem 3.3, where we choose Z such that $\pi_y \partial F(y(z), z)$ is nonsingular for each $z \in Z$; let Ω be $\{(y(z), z) : z \in Z\}$; and let $\pi_z IC_F(\Omega)$ be the set of points $z \in Z$ such that F is i-consistent at $(y(z), z)$ with respect to Ω. Suppose that $\bar{z} \notin \pi_z IC_F(\Omega)$. Then:

(a) $\pi_z IC_F(\Omega)$ is open and dense on Z;

(b) $y(\cdot)$ is differentiable on $\pi_z IC_F(\Omega)$ and, for every $z \in \pi_z IC_F(\Omega)$,

$$\nabla y(z) = \nabla y^i(z) = -(M^i)^{-1} N^i, \quad \forall i \in \mathcal{I}_F(y(z), z). \tag{15}$$

where $[M^i N^i] = \nabla F^i(y(z), z)$;

(c) let $\mathcal{J} = \overline{\lim}_{\substack{z \in \pi_z IC_F(\Omega) \\ z \to \bar{z}}} \mathcal{I}_F(y(z), z)$ and $\partial_B^{\mathcal{J}} F(\bar{y}, \bar{z}) = \{\nabla F^j(\bar{y}, \bar{z}) : j \in \mathcal{J}\}$; we have

$$\partial_B y(\bar{z}) = \{-M^{-1} N : [MN] \in \partial_B^{\mathcal{J}} F(\bar{y}, \bar{z})\}; \tag{16}$$

(d) if F is e-i-consistent at $(\bar{y}, \bar{z})$, then

$$\partial_B y(\bar{z}) = \{-M^{-1} N : [MN] \in \partial_B F(\bar{y}, \bar{z})\}.$$

Proof. This requires careful but routine arguments; see Ref. 7. □

Consider Example 4.1. It is easy to see that

$$y(z) = \begin{cases} -(1/3)z, & \text{if } z \le 0, \\ -(1/2)z, & \text{if } z > 0 \end{cases}$$

is the unique implicit function of F at (0, 0). Moreover,

$$\partial y(0) = [-1/2, -1/3].$$

On the other hand,

$$\partial_B F(0,0) = \left\{ \begin{bmatrix} 1 \\ 1/3 \end{bmatrix}, \begin{bmatrix} 1 \\ 1/2 \end{bmatrix}, 2\begin{bmatrix} 1 \\ 1/3 \end{bmatrix}, 2\begin{bmatrix} 1 \\ 1/2 \end{bmatrix} \right\}$$

and

$$\text{conv}\{M^{-1} N : [MN] \in \partial_B F(0,0)\} = \text{conv } \{-1/3, -1/2\} = [-1/2, -1/3],$$

as expected from Proposition 4.2 (d).

Now, consider Example 4.2. The implicit function of F at (0, 0) is $y(z) = -(1/2)z$, which is smooth. $\partial_B^{\mathcal{J}} F(0,0) = \partial_B^1 F(0,0) = (1, 1/2)^T$. This confirms (16). We note that here $\partial y(0)$ has nothing to do with $f^2(y, z)$

and $f^3(y, z)$. In particular $\partial y(0)$ is a strict subset of conv $\{-M^{-1}N : [MN] \in \partial F(0, 0)\}$.

We discuss now the generalized Jacobians of the implicit smoothing $\hat{y}(z, \epsilon)$ of $y(z)$, which is given by Theorem 3.2, at a point $(z, 0)$ under the assumption that F is piecewise smooth. Since the Clarke generalized Jacobian is no more than the convex hull of the B-differential, our results established in terms of the B-subdifferentials can be extended easily to those based on the Clarke generalized Jacobian.

Our first result here is a refinement to the PC^1 case of Proposition 3.1, in which the plenary hull of the projected Clarke operator plen $\pi_x \partial$ is replaced by the projected B-subdifferential operator $\pi_x \partial_B$.

Proposition 4.3. Let $G: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ be PC^1 and $g(\cdot) = G(\cdot, 0)$. Then, $g: \mathbb{R}^n \to \mathbb{R}^n$ is PC^1 and, for every $x \in \mathbb{R}^n$, $\partial_B g(x) \subset \pi_x \partial_B G(x, 0)$.

Proof. The piecewise smoothness of g is obvious. First, we have from Ref. 28, for any $x' \in \mathbb{R}^n$ and $(u, v) \in \mathbb{R}^n \times \mathbb{R}^m$, that

$$G'((x', 0), (u, v)) \in \partial_B G(x', 0)(u, v), \tag{17}$$

where $G'(a, b)$ denotes the directional derivative of G at point a in the direction b; second, $\partial_B G(x', 0)$ coincides with $\{\nabla G^i(x', 0) : i \in \mathcal{I}_G^e(x', 0)\}$, where

$$\mathcal{I}_G^e(x', 0) = \{i : (x', 0) \in \text{cl int}\{(u, v) \in \mathcal{N}(x', 0) : i \in \mathcal{I}_G(u, v)\}\}.$$

Taking $x' \in D_g$ and $v = 0$ in (17) gives

$$\nabla g(x')u = G'((x', 0), (u, 0)) \in [\pi_x \partial G(x', 0)]u,$$

that is,

$$\nabla g(x') \in \text{plen } \pi_x \partial G(x', 0)$$

because u is arbitrary. By Ref. 18, Example 4.09, every countable set of matrices is plenary; hence, $\pi_x \partial_B G(x', 0)$ is a plenary set and consequently

$$\nabla g(x') \in \pi_x \partial_B G(x', 0).$$

Let

$$x' \in D_g \quad \text{and} \quad x' \to x.$$

By the definition of the B-subdifferential and the upper semicontinuity of $\partial_B G(\cdot, 0)$, we know that every accumulation matrix of $\nabla g(x')$ is contained in $\pi_x \partial_B G(x, 0)$. □

Theorem 4.1. Let $F(y,z)$ be defined as in (1) and be PC^1. Let $\hat{F}$ be a smoothing of F that satisfies the conditions of Theorem 3.2 and let $\hat{y}$ be the implicit smoothing function given there. Then, for all z near $\bar{z}$:

(a) we have

$$\partial_B y(z) \subset \pi_z \partial_B \hat{y}(z,0), \tag{18}$$

$$\pi_z \partial_B \hat{y}(z,0) \subset \{-M^{-1}N : [MNb] \in \partial_B \hat{F}(y(z),z,0)\}; \tag{19}$$

(b) if $\hat{F}(y,z,\epsilon)$ satisfies the strong Jacobian consistency at $(y(z),z)$, then

$$\pi_z \partial_B \hat{y}(z,0) \subset \{-M^{-1}N : [MN] \in \partial F(y(z),z)\};$$

(c) if $\hat{F}(y,z,\epsilon)$ satisfies the strong B-subdifferential consistency at $(y(z),z)$, then

$$\pi_z \partial_B \hat{y}(z,0) \subset \{-M^{-1}N : [MN] \in \partial_B F(y(z),z)\}.$$

Proof. Part (a). The relation $\partial_B y(z) \subset \pi_z \partial_B \hat{y}(z,0)$ follows from Proposition 4.3 and (19) follows from a similar proof to that of Theorem 3.3.

Part (b) follows from Part (a) and the strong Jacobian consistency.

Part (c) follows from Part (a) and the strong B-subdifferential consistency. □

We note that Theorem 4.1 is stronger that Theorem 3.3 when F is PC^1. For instance, (18) implies (11), (19) implies (12), and in general the set

$$\text{plen}\{-M^{-1}N : [MNb] \in \partial_B \hat{F}(y,z,0)\}$$

is strictly larger than the set

$$\{-M^{-1}N : [MNb] \in \partial_B \hat{F}(y,z,0)\}.$$

We note also that when F is PC^1, the implicit function $y(\cdot)$ defined as in the situation of Theorem 3.3 is also piecewise smooth. Given the Clarke implicit function theorem, the following is essentially due to Refs. 1,19.

Lemma 4.2. Let $F(y,z)$ be defined as in (1). Suppose that F is PC^1 and $\pi_y \partial F(\bar{y},\bar{z})$ is nonsingular. Then, $y(\cdot)$ is PC^1 and it is a continuous selection of a family of continuously differentiable functions $\{y^i(\cdot) : i \in \mathcal{I}_y^e(\bar{z})\}$, where $\mathcal{I}_y^e(\bar{z}) = \{i : \bar{z} \in \ \text{cl int}\ \{z' \in Z : i \in \mathcal{I}_y(z')\}\}$. Moreover, $y^i(z)$ is associated with a smooth system $F^i(y,z)=0$ in the sense that $y = y^i(z)$ is the unique solution of $F^i(y,z)=0$ for (y,z) close to $(\bar{y},\bar{z})$.

This shows that $\partial y(\bar{z})$ is theoretically computable. However, the set $\mathcal{I}_y^e(\bar{z})$ is often difficult to identify in practice, particularly in the implicit case.

Remark 4.2. Let $F(y,z)$ be defined as in (1) and be PC^1. Suppose that $\hat{F}(y,z,\epsilon)$ is the adaptive integral smoothing of $F(y,z)$. Using Ref. 11, Corollary 2.1, we can prove that $\hat{F}$ satisfies the strong Jacobian consistency at $(\bar{y},\bar{z})$, that is,

$$\pi_{y,z}\partial_B \hat{F}(\bar{y},\bar{z},0) \subset \partial F(\bar{y},\bar{z}).$$

Under the condition of Theorem 4.1, we have that, for all z near $\bar{z}$,

$$\pi_z \partial_B \hat{y}(z,0) \subset \{-M^{-1}N : [MN] \in \partial F(y(z),z)\}.$$

We conjecture that this condition can be extended to the case when $\hat{F}$ is the bounded integral smoothing.

As an application of Proposition 4.2, we want to strengthen Theorem 4.1. It follows directly from Theorem 4.1 (b) and Proposition 4.2 (b) that, for every z such that F is index-consistent at $(y(z),z)$, that is, $z \in \pi_z IC_F(\Omega)$,

$$\lim_{|\epsilon|\downarrow 0} \nabla_z \hat{y}(z,\epsilon) = \nabla y(z).$$

Moreover, by virtue of Theorem 4.1 (a), (c), and Proposition 4.2 (d), we have the following theorem.

Theorem 4.2. Let $F(y,z)$ be defined as in (1) and let $\hat{F}(y,z,\epsilon)$ be a smoothing of $F(y,z)$. Suppose that the conditions of Theorem 3.2 are satisfied and that F is PC^1. Let $\hat{y}: Z \times \mathcal{E} \to \mathbb{R}^n$ be the unique implicit smoothing function defined by Theorem 3.2. Assume $\hat{F}$ satisfies the strong B-subdifferential consistency and that F is e-i-consistent at $(\bar{y},\bar{z})$ with respect to $\pi_z IC_F(\Omega)$. Then, for all $z \in Z$,

$$\partial_B y(z) = \pi_z \partial_B \hat{y}(z,0) = \{-M^{-1}N : [MN] \in \partial_B F(y,z)\}.$$

The above results implies that $\partial_B y(z)$ can be approximated by $\nabla_z \hat{y}(z,\epsilon)$. Unfortunately, strong B-subdifferential consistency does not hold for some interesting smoothing functions. Indeed, $\pi_x \partial_B \hat{F}(x,0)$ is much larger than $\partial_B F(x)$ in general. Note that $\partial_B \hat{F}(x,0)$ is the set of accumulation matrices $\{\nabla \hat{F}(x_k,\epsilon_k)\}$ as $(x_k,\epsilon_k) \to (x,0)$. If we restrict ϵ_k in some way, for example, by taking only a special subsequence, then the set of accumulation matrices is only a subset of $\partial_B \hat{F}(x,0)$. An interesting question is: to what extent can we restrict ϵ_k so that the B-subdifferential consistency holds for a subset of $\partial_B \hat{F}(x,0)$? Let $F:\mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz function, let $\hat{F}(x,\epsilon)$ be a smoothing function of F and $x \in \mathbb{R}^n$. Let $\mathcal{T}$ denote a class of function τ such that, for $x_k \neq x, \tau(x_k) \neq 0$ and $\tau(x_k) \to 0$ as $x_k \to x$. We say $\hat{F}(x,\epsilon)$ satisfies the $\mathcal{T}$-weak B-subdifferential consistency at x if there exists a function $\tau \in \mathcal{T}$ such that $\pi_x \partial_B^\tau \hat{F}(x,0) = \partial_B F(x)$, where

$$\partial_B^\tau \hat{F}(x,0) = \left\{ \lim_{\substack{x_k \neq x \\ x_k \to x}} \nabla \hat{F}(x_k, \tau(x_k)) \right\}.$$

We call $\pi_x \partial_B^\tau \hat{F}(x,0)$ a restricted B-subdifferential with respect to τ. We note that many smoothing NCP functions in Ref. 4 satisfy the $\mathcal{T}$-weak B-subdifferential consistency provided that $\tau(t) = o(t)$, that is, $\tau(t)/t \to 0$ as $t \to 0$. We will not go further in this direction as it is not the main interest of this paper.

In somes cases, it is interesting to consider the following perturbed system of equations of (1):

$$F(y,z) = t, \tag{20}$$

where F is defined as in (1). Suppose that F is PC^1 and $\pi_y \partial F(\bar{y},\bar{z})$ is nonsingular. Then, by Theorem 3.1, there exist neighborhoods Z of $\bar{z}$ and T of $0 \in \mathbb{R}^n$ such that $\tilde{y}: Z \times T \to \mathbb{R}^n$ is the unique solution of (20). Here, we denote the implicit function by $\tilde{y}$ in order to distinguish it from $\hat{y}$, which stands for the implicit smoothing function associated with (8). In the following, we shall investigate $\partial \tilde{y}(\bar{z},0)$.

Proposition 4.4. Suppose that F is PC^1 and $\pi_y \partial F(\bar{y},\bar{z})$ is nonsingular. Then, there exists a function $\tilde{y}(\cdot,\cdot)$, the unique solution of (20) in a neighborhood $Z \times T$ of $(\bar{z},0)$, such that

$$\partial_B \tilde{y}(\bar{z},0) = \{-M^{-1}[N,-I] : [MN] \in \partial_B F(\bar{y},\bar{z})\}, \tag{21}$$

where $M \in \mathbb{R}^{n\times n}$, $N \in \mathbb{R}^{n\times m}$, and $I \in \mathbb{R}^{n\times n}$ is the identity matrix. Also,

$$\partial_B \tilde{F}(\bar{y},\bar{z},0) = [\partial_B F(\bar{y},\bar{z}), -I].$$

Proof. See Ref. 7. □

The significance of this result is that the PC^1 property allows $\partial_B \tilde{y}(\bar{z},0)$ and $\partial_B \tilde{F}(\bar{y},\bar{z},0)$ to be expressed exactly by $\partial_B F(\bar{y},\bar{z})$. Of course, $\partial \tilde{y}(\bar{z},0)$ is the convex hull of the set given by (21).

5. Application to Optimization with Piecewise Smooth Equality Constraints

In this section, we will investigate the generalized Karush-Kuhn-Tucker conditions of several (roughly) equivalent formulations of a nonsmooth equality constrained minimization problem. We are interested mainly in the case when $F(y,z)$ is PC^1.

5.1. Formulation with Nonsmooth Equality Constraints. We study the first-order necessary conditions of the following equality constrained minimization problem:

$$(\mathcal{P}) \quad \min_{y,z} \quad f(y,z), \qquad \text{s.t.} \quad F(y,z)=0,$$

where $f:\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is continuously differentiable and $F:\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is locally Lipschitz. We omit smooth inequality constraints to simplify the discussion. This class of optimization problems is interesting particularly in the study of MPEC and BP.

Here, we utilize the theory developed in the previous sections to investigate the generalized Karush-Kuhn-Tucker (GKKT) conditions of the minimization problem and its variations resulting from the smoothing or perturbation of the constraint functions. See Ref. 20 for details on GKKT points.

Our purpose here lies with the GKKT conditions of different programs related to $(\mathcal{P})$. In particular, we would like to understand the relationship between a smoothed program (implicit or not implicit) or $(\mathcal{P})$ and a perturbed program and hence explain a smoothing method as essentially an inexact method which solves the equality constraints approximately. We would also like to understand that, if $(\mathcal{P})$ is reformulated from MPEC, whether or not numerical methods based on different programs will generate the same GKKT points.

Throughout this section, we assume that $\pi_y \partial F(\bar{y},\bar{z})$ is nonsingular for any $\bar{y} \in \mathbb{R}^n, \bar{z} \in \mathbb{R}^m$ such that $F(\bar{y},\bar{z})=0$; hence there exists a locally unique implicit function $y(\cdot)$ such that $y(\bar{z})=\bar{y}$ and $F(y(z),z)=0$ for every z near $\bar{z}$.

Recall that the GKKT condition of $(\mathcal{P})$ is

$$\nabla f(y,z) + \partial F(y,z)^T \lambda \ni 0,$$
$$F(y,z)=0,$$

where $\lambda \in \mathbb{R}^n$. Note that, by convention, we regard the gradient of a real valved function as a column vector. If a triplet (y,z,λ) satisfies the above equations, (y, z) is called a GKKT stationary point and λ is called a GKKT multiplier.

Remark 5.1. We note that

$$\partial F(y, z)^T \lambda = [\text{plen}\, \partial F(y, z)]^T \lambda.$$

Thus, replacing $\partial F(y, z)$ with plen $\partial F(y, z)$ gives the same GKKT conditions.

Since $\pi_y \partial F(y(z), z)$ is nonsingular, we can obtain the following equivalent GKKT condition of $(\mathcal{P})$:

$$\nabla_z f(y(z), z) \in \Im_F(y(z), z) \nabla_y f(y(z), z), \tag{22}$$

where

$$\Im_F(y, z) = \{N^T M^{-T} : M \in \mathbb{R}^{n\times n}, N \in \mathbb{R}^{n\times m}, [M N] \in \partial F(y, z)\}, \tag{23}$$

where M^{-T} denotes $(M^{-1})^T$.

5.2. Implicit Programming Formulation. In some cases, we may solve $(\mathcal{P})$ determining the vector y from the constraints as a locally unique implicit function of the vector z and substituting it into the objective function. Consequently, $(\mathcal{P})$ is reduced to an unconstrained minimization problem. We write such a program as

$$(\mathcal{P}^I) \quad \min_z f(y(z), z)$$

where $y(z)$ solves $F(y, z) = 0$. Therefore, it is interesting to investigate the relationship between the GKKT conditions of $(\mathcal{P})$ and $(\mathcal{P}^I)$. The GKKT conditions of $(\mathcal{P}^I)$ are given by

$$0 \in (\partial y(z), I)^T \nabla f(y(z), z),$$

which can be rearranged as

$$\nabla_z f(y(z), z) \in -\partial y(z)^T \nabla_y f(y(z), z). \tag{24}$$

When F is PC^1, it follows from Proposition 4.2 that

$$-\partial y(z)^T = \text{conv}\{N^T M^{-T} : [M N] \in \partial_B^{\mathcal{J}} F(y, z)\}; \tag{25}$$

see Proposition 4.2 for details. For simplicity of notation, let

$$\mathcal{G}_F^{\mathcal{J}}(y,z)=\{N^T M^{-T}:[MN]\in\partial_B^{\mathcal{J}} F(y,z)\},$$
$$\Im_F^{\mathcal{J}}(y,z)=\{N^T M^{-T}:[MN]\in\partial^{\mathcal{J}} F(y,z)\},$$

where

$$\partial_F^{\mathcal{J}} F(y,z)=\operatorname{conv}\partial_B^{\mathcal{J}} F(y,z).$$

In general, $\Im_F^{\mathcal{J}}(y,z)$ is not a convex set and hence

$$\operatorname{conv}\mathcal{G}_F^{\mathcal{J}}(y,z)\neq\Im_F^{\mathcal{J}}(y,z).$$

When F is not e-i-consistent at $(y,z), \mathcal{I}_F(y,z)$ is strictly larger than $\mathcal{J}$, hence $\Im_F(y,z)$ may be larger than $\Im_F^{\mathcal{J}}(y,z)$, which is roughly equivalent to $-\partial y(z)$ (Ref. 7) in that there is a bijection between these sets that preserves the generating elements $M^{-1}N$ of the latter; c.f. (25). As such, we regard the GKKT conditions of $(\mathcal{P}^I)$ as being sharper than those of $(\mathcal{P})$.

Example 5.1. Let F be defined as in Example 4.2. Consider a minimization program with the objective function

$$f(y,z)=-0.5y-z+y^2+z^2$$

and equality constraint

$$F(y,z)=0.$$

At the point (0, 0),

$$\nabla f(0,0)=(-0.5,-1)^T,$$
$$\partial F(0,0)=\left\{\begin{bmatrix}1\\0.5\end{bmatrix},\begin{bmatrix}0.5\\1\end{bmatrix},\begin{bmatrix}0.5\\0\end{bmatrix}\right\}.$$

Then, there exists $\lambda=1$ such that

$$0\in\nabla f(0,0)+\lambda\partial F(0,0).$$

Therefore, (0,0) is a generalized stationary point of $(\mathcal{P})$.

Now, we look at $(\mathcal{P}^I)$. The implicit function defined by $F(y,z)=0$ is

$$y=-0.5z.$$

Substituting it into f, we have

$$\hat{f}\equiv f(-0.5z,z)=-0.5(-0.5z)-z=-0.75z+1.25z^2.$$

Obviously, $z=0$ is not a stationary point of $(\mathcal{P}^I)$.

5.3. Formulation with Perturbed Nonsmooth Constraints. We investigate now the perturbed case of $(\mathcal{P})$ which is defined as

$$(\mathcal{P}_t) \quad \min_{y,z,t} \quad f(y,z),$$
$$\text{s.t.} \quad F(y,z)=t,$$
$$t=0.$$

A point $(y,z,0)$ is a GKKT stationary point of $(\mathcal{P}_t)$ if there exist $\lambda \in \mathbb{R}^n$, $\mu \in \mathbb{R}^n$ such that

$$\begin{pmatrix} \nabla f(y,z) \\ 0 \end{pmatrix} + \begin{pmatrix} \partial F(y,z)^T \\ -I \end{pmatrix} \lambda + \begin{pmatrix} 0 \\ I \end{pmatrix} \mu \ni 0,$$
$$F(y,z)-t=0,$$
$$t=0.$$

Equivalently, there exist $M \in \mathbb{R}^{n\times n}$ such that $[M,N] \in \partial F(y,z)$ and

$$-M^{-T}\nabla_y f(y,z)=\lambda,$$
$$\nabla_z f(y,z)-N^T M^{-T}\nabla_y f(y,z)=0,$$
$$-\lambda+\mu=0,$$
$$F(y,z)-t=0,$$
$$t=0.$$

When F is PC^1, the above equations can be reduced to

$$\nabla_z f(y(z),z) \in \Im_F(y(z),z)\nabla_y f(y(z),z), \tag{26}$$

where $\Im_F(y(z),z)$ is defined by (23).

Proposition 5.1. Let $F:\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ be a PC^1. Then, the GKKT condition of $(\mathcal{P})$ is equivalent to that $(\mathcal{P}_t)$; that is, if (y,z) is a GKKT point of $(\mathcal{P})$, then $(y,z,0)$ (here $0 \in \mathbb{R}^n$) is a GKKT point of $(\mathcal{P}_t)$; conversely, if (y,z,t) is a GKKT point of $(\mathcal{P}_t)$, then (y,z) is a GKKT point of $(\mathcal{P})$.

5.4. Smoothing Approach. When the equality constraints in $(\mathcal{P})$ are smoothed by a smoothing method (see Ref. 4 and references therein), we have the following smoothed program:

$$(\mathcal{P}_\epsilon) \quad \min_{y,z,\epsilon} \quad f(y,z),$$
$$\text{s.t.} \quad \hat{F}(y,z,\epsilon)=0,$$
$$\epsilon=0,$$

where $\hat{F}(y, z, \epsilon)$ is a smoothing of F. We will discuss the application of nonlinear programming techniques to this formulation later. The GKKT condition of $(\mathcal{P}_\epsilon)$ can be given as follows: there exists $M \in \mathbb{R}^{n\times n}$, $N \in \mathbb{R}^{n\times m}, b \in \mathbb{R}^n$ such that

$$[M, N, b] \in \partial \hat{F}(y, z, \epsilon),$$

with

$$\begin{aligned}
&-M^{-T}\nabla_y f(y, z) = \lambda,\\
&\nabla_z f(y, z) - N^T M^{-T}\nabla_y f(y, z) = 0,\\
&-\lambda^T b + \mu = 0,\\
&\hat{F}(y, z, \epsilon) = 0,\\
&\epsilon = 0.
\end{aligned}$$

Thus, the GKKT condition is equivalent to the following:

$$\nabla_z f(y(z), z) \in \mathcal{B}_{\mathcal{P}}(y(z), z)\nabla_y f(y(z), z),$$

where

$$\mathcal{B}_{\mathcal{P}}(y(z), z) = \{N^T M^{-T} : [MN] \in \pi_{(y,z)}\partial \hat{F}(y(z), z, 0)\}.$$

If $\hat{F}$ satisfies (9), then

$$\text{plen } \pi_{(y,z)}\partial \hat{F}(y(z), z, 0) = \text{plen } \partial F(y(z), z);$$

consequently,

$$\begin{aligned}
&\{N^T M^{-T} : [MN] \in \text{plen } \pi_{(y,z)}\partial \hat{F}(y(z), z, 0)\}\\
&= \{N^T M^{-T} : [MN] \in \text{plen } \partial F(y(z), z)\}.
\end{aligned}$$

Here, we use the fact that the plenary hull does not change the nonsingularity; see Ref. 27. Following Remark 5.1, we know that the GKKT conditions of $(\mathcal{P}_\epsilon)$ and $(\mathcal{P})$ are equivalent. Based on this observation and Proposition 5.1, we have the following proposition.

Proposition 5.2. Suppose that $\hat{F}$ is a smoothing of F and satisfies (9). Then, the GKKT condition of $(\mathcal{P}_\epsilon)$ is equivalent to that of $(\mathcal{P})$ and $(\mathcal{P}_t)$ in a sense similar to Proposition 5.1.

The equivalence of the GKKT conditions between $(\mathcal{P}_\epsilon)$ and $(\mathcal{P}_t)$ has some interesting implications. Practically, because of nonsmoothness of F, we may use some smoothing method to find a smoothing function $\hat{F}(y,z,\epsilon)$ to replace $F(y,z)$ and get a smooth program. At step $k, \epsilon=\epsilon_k$, we solve exactly the smooth program $(\mathcal{P}_\epsilon)$ by some minimization method and obtain a stationary point $(\hat{y}(z_k,\epsilon_k), z_k, \epsilon_k)$. As we discussed, if $(\hat{y}(z_k,\epsilon_k), z_k, \epsilon_k) \to (y,z,0)$ as $|\epsilon_k| \downarrow 0$, then, $(y,z,0)$ is a stationary point of $(\mathcal{P}_\epsilon)$ and, by Proposition 5.2, $(y,z,0)$ (here $0 \in \mathbb{R}^n$) is a stationary point of $(\mathcal{P}_t)$. Also, since $\epsilon_k \neq 0, (\hat{y}(z_k,\epsilon_k), z_k)$ is not a solution of $F(y,z)=0$, but of $F(y,z)=t_k$, where $t_k = F(\hat{y}(z_k,\epsilon_k), z_k)$. Thus $(\hat{y}(z_k,\epsilon_k), z_k, t_k)$ is a feasible point of the following program:

$$\min_{y,z,t} \quad f(y,z), \tag{27a}$$

$$\text{s.t.} \quad F(y,z)=t, \tag{27b}$$

$$t=t_k. \tag{27c}$$

Note that $(\hat{y}(z_k,\epsilon_k), z_k, t_k)$ is not necessarily a stationary point of (27). However, any stationary point (y'_k, z'_k, t_k) of (27) must be a feasible point and therefore satisfy $F(y'_k, z'_k)=t$ and $t=t_k$. Moreover, $t_k \to 0$ as $|\epsilon_k| \downarrow 0$. It is easy to check that every accumulation point $(y', z', 0)$ of (y'_k, z'_k, t_k) is a stationary of $(\mathcal{P}_t)$. By Proposition 5.2 $(y', z', 0)$ is also a stationary point of $(\mathcal{P}_\epsilon)$. Furthermore, if $(y', z', 0)$ is the unique GKKT point of $(\mathcal{P}_t)$, then $(\mathcal{P}_\epsilon)$ has the unique stationary point $(y', z', 0)$. We can conclude that smoothing the constraint functions can be interpreted as a (proper) perturbation of the constraints.

5.5. Further Implicit Programming Approaches. We may consider the implicit form of the perturbed program $(\mathcal{P}_t)$,

$$(\mathcal{P}_t^I) \quad \min_{z,t} \quad f(\tilde{y}(z,t), z), \qquad \text{s.t.} \quad t=0, \tag{28}$$

where $\tilde{y}(z,t)$ solves

$$\tilde{F}(y,z,t) = F(y,z) - t = 0.$$

In the PC^1 case it can be shown (Ref. 7) that the GKKT conditions of $(\mathcal{P}_t^I)$ are roughly equivalent to those of $(\mathcal{P})$.

The implicit form of $(\mathcal{P}_\epsilon)$ is defined as

$$(\mathcal{P}_\epsilon^I) \quad \min_{z,\epsilon} \quad f(\tilde{y}(z,\epsilon), z), \qquad \text{s.t.} \quad \epsilon=0, \tag{29}$$

where $\hat{y}(z,\epsilon)$ solves

$$\hat{F}(y,z,\epsilon)=0.$$

As discussed in Ref. 7, the GKKT conditions of $(\mathcal{P}_\epsilon^I)$ are generally weaker than those of $(\mathcal{P}^I)$, but they are roughly equivalent when $\hat{F}$ satisfies the strong B-subdifferential consistency and F is e-i-consistent.

5.6. Overview of GKKT Conditions. Based on the preceding discussions, we compare the stationary conditions of various formulations:

(i) $(\mathcal{P}^I)$ generally has the sharpset GKKT conditions;
(ii) the GKKT conditions of $(\mathcal{P})$ and $(\mathcal{P}_t)$ are equivalent; when $\hat{F}$ satisfies (9), they are equivalent to those of $(\mathcal{P}_\epsilon)$;
(iii) when $\hat{F}$ satisfies the strong B-subdifferentials consistency and F is e-i-consistent, the GKKT conditions of $(\mathcal{P}_\epsilon^I), (\mathcal{P}^I), (\mathcal{P}_\epsilon), (\mathcal{P}), (\mathcal{P}_t), (\mathcal{P}_t^I)$ are roughly equivalent.

The implication for numerical methods solving $(\mathcal{P})$ is as follows. Apart from smoothing methods, we may deal with the nonsmoothness of the equality constraints with other methods which solve the equality approximately. For instance, we may find an approximate solution (near the exact solution) at which F is continuously differentiable. Under the proper control of the residual of the equality constraints, such approximate methods may generate the same GKKT points as a smoothing method. Therefore, in dealing with equality constraints, a smoothing method is essentially equivalent to an approximation method.

References

1. Luo, Z. Q., Pang, J. S., and Ralph, D., *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, UK, 1996.
2. Outrata, J., Kocvara, M., and Zowe, J., *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications, and Numerical Constraints*, Kluwer Academic Publishers, Dordrecht, Holland, 1998.
3. Dempe, S., *An Implicit Function Approach to Bilevel Programming Problems*, Multilevel Optimization Algorithm and Applications, Edited by A. Migdalas, P. M. Pardalos and P. Varbrand, Kluwer Academic Publishers, Dordrecht, Holland, pp. 273–294, 1998.
4. Jiang, H., and Ralph, D., *Smooth SQP Methods for Mathematical Programs with Nonlinear Complementarity Constraints*, SIAM Journal on Optimization, Vol. 10, pp. 779–808, 2002.
5. Qi, L., *Convergence Analysis of Some Algorithm for Solving Nonsmooth Equations*, Mathematics of Operations Research, Vol. 18, pp. 227–244, 1993.

6. Clarke F. H., *Optimization and Nonsmooth Analysis*, Wiley, New York, NY, 1983.
7. Ralph, D., and Xu, H., *Implicit Smoothing and Its Application to Optimization with Piecewise Smooth Equality Constraints*, Preprint, Department of Mathematics and Statistics, University of Melbourne, Melbourne, Australia, 1999.
8. Qi, L., and Chen, X., *A Globally Convergent Successive Approximation Method for Severely Nonsmooth Equations*, SIAM Journal on Control and Optimization, Vol. 33, pp. 402–418, 1995.
9. Sun, D., *Smoothing-Nonsmooth Reformulations of Variational Inequality Problems*, Preprint, School of Mathematics, University of New South Wales, Sydney, Australia, 1998.
10. Mayne, D. Q., and Polak, E., *Nondifferentiable Optimization Adaptive Smoothing*, Journal of Optimization Theory and Applications, Vol. 43, pp. 601–613, 1984.
11. Xu, H., *Adaptive Smoothing Methods, Deterministically Differentiable Jacobians, and Newton's Methods*, Journal of Optimization Theory and Applications, Vol. 109, pp. 215–224, 2001.
12. Sweetser, T. H., *A Minimal Set-Valued Strong Derivative for Vector-Valued Lipschitz Functions*, Journal of Optimization Theory and Applications, Vol. 23, pp. 549–562, 1977.
13. Ralph, D., *Rank-1 Support Functionals and the Generalized Jacobian, Piecewise Linear Homeomorphisms*, PhD Thesis, University of Wisconsin, Madison, Wisconsin, 1990.
14. Xu, H., *Set-Valued Approximation and Newton's Methods*, Mathematical Programming, Vol. 84, pp. 401–420, 1999.
15. Chen, X., Qi, L., and Sun, D., *Global and Superlinear Convergence of the Smoothing Newton's Method and Its Application to General Box Constrained Variational Inequalities*, Mathematics of Computation, Vol. 67, pp. 519–540, 1998.
16. Pang, J. S., and Ralph, D., *Piecewise Smoothness, Local Invertibility, and Parametric Analysis of Normal Maps*, Mathematics of Operations Research, Vol. 21, pp. 401–426, 1996.
17. Kummer, B., *Newton's Methods for Nondifferentiable Functions*, Advances in Mathematical Optimization, Edited by J. Guddat et al., Akademie Verlag, Berlin, Germany, Vol. 45, pp. 114–125, 1988.
18. Sweetser, T. H., *A Set-Valued Strong Derivative in Infinite-Dimensional Spaces, with Applications in Hilbert Spaces*, PhD Thesis, University of California, San Diego, California, 1979.
19. Ralph, D., and Scholtes, S., *Sensitivity Analysis of Composite Piecewise Smooth Equations*, Mathematical Programming, Vol. 76, pp. 593–612, 1997.
20. Hiriart-Urruty, J. B., *Refinements of Necessary Optimality Conditions in Nondifferentiable Programming, I*, Applied Mathematics and Optimization, Vol. 5, pp. 63–82, 1979.