Math. Program. 84: 401–420 (1999) / DOI 10.1007/s10107980028a

H. Xu

Set-valued approximations and Newton's methods ⋆

Received October 2, 1995 / Revised version received May 5, 1998
Published online October 9, 1998

Abstract. We introduce a point-based set-valued approximation for a mapping from R^n to R^m. Under the assumption of semi-smoothness of the mapping, we prove that the approximation can be obtained through the Clarke generalized Jacobian, Ioffe-Ralph generalized Jacobian, B-subdifferential and their approximations. As an application, we propose a generalized Newton's method based on the point-based set-valued approximation for solving nonsmooth equations. We show that the proposed method converges locally superlinearly without the assumption of semi-smoothness. Finally we include some well-known generalized Newton's methods in our method and consolidate the convergence results of these methods.

Key words. *PBSVA* – plenary sets – Ioffe-Ralph generalized Jacobian – Newton's methods

1. Introduction

Let F be a mapping from R^n to R^m. It is well-known that, at every point $x \in R^n$, $F(x)$ can be approximated through a variety of derivatives of F at x when they exist. In this paper, we propose a new approximation to $F(x)$, more precisely, we introduce a set-valued mapping $\mathcal{A}F : R^n \to 2^{R^{m\times n}}$ by which the set-valued term $F(y) + \mathcal{A}F(y)(y - x)$ approximates $F(x)$ in some sense for y sufficiently close to x. Clearly, such an approximation is closely related to the Newton's method which has been playing an essential role in solving nonlinear equations. The classic Newton's method has been widely used to solve mathematical programming, nonlinear variational inequality and nonlinear complementarity problems. The book of Ortega and Rheinbodlt [18] gives an excellent treatment of the classic Newton's method and references.

As it is well known, the essence of the classic Newton's method is to replace, at the current iterate, the mapping F whose zero is sought by an approximate linear mapping that can be solved more easily. A zero of this linear approximation mapping is then found to replace the current iterate and the process is restarted. All this depends on the fact that F is differentiable.

More recently, stimulated by its important applications in treating mathematical and equilibrium programming, a nonsmooth version of Newton's method has appeared and grown rapidly. See for instance [2,3,5–7,14,15,19,20,22,23,25,27,28,30,29,31,33–37].

H. Xu: School of Information Technology and Mathematical Sciences, University of Ballarat, Victoria, VIC 3353, Australia

Mathematics Subject Classification (1991): 90C30, 90C33

⋆ An earlier version of this paper was presented to the 16th Biennial Conference on Numerical Analysis, Dundee University, Scotland, June 1995.

When F is not differentiable, Robinson [30] sought to define a point-based approximation in order to get around the difficulties resulting from the lack of the differentiability of F. The approximation is single-valued and generally nonlinear. This work was further extended by Kummer [14], Pang [21], Gabriel and Pang [8], Ralph [27] and Dirkse and Ferris [5,6].

In this paper, we introduce a point-based set-valued approximation to $F(x)$ by using Robinson's idea. The existence of such an approximation is proved for a mapping F which is not necessarily locally Lipschitz. We prove that the plenary hull of a point-based set-valued mapping is also a point-based set-valued mapping, by which we prove further that if a mapping F is semi-smooth then not only the Clarke generalized Jacobian but also its plenary hull, the Ioffe-Ralph generalized Jacobian, is a point-based set-valued mapping. We also show that some approximations of the Clarke generalized Jacobian and the Ioffe-Ralph generalized Jacobian are also point-based set-valued approximations. As an application, a generalized Newton's method is proposed based on the point-based set-valued approximation. The convergence results of the proposed method are obtained without assumption of semi-smoothness. We consolidate the convergence results of some generalized Newton's methods without adding more conditions. Finally, we note that Qi [24] introduced a notion of C-differential operator which is closely related to our point-based set-valued approximation. We compare our results with Qi's results in [24]. Further discussion on the development in this direction was made in [38].

The remainder of this paper are organized as follows: In Section 2, we discuss the Clarke generalized Jacobian and its plenary hull. An approximation to the latter is introduced. In Section 3, we introduce a point-based set-valued approximation and discuss some important properties of such an approximation. Further, under the assumption of semi-smoothness, many examples of the point-based set-valued approximation are presented. In Section 4 we propose a generalized Newton's method based on the point-based set-valued approximation for solving nonsmooth equations and discuss the convergence of the proposed method. Finally, using the theory of Section 3, we strengthen the convergence results of some generalized Newton's methods in Section 5.

2. Generalized Jacobian and weak approximations

2.1. Basic definition and notion

Throughout this paper we will use the following notation. R^n will denote the n-dimensional Euclidean space with the usual inner product $\langle\cdot;\cdot\rangle$, and $R^{n\times m}$ will be the space of $n\times m$ real matrices. $\|x\|$, for $x \in R^n$, will represent the 2-norm of a vector x, and $\|A\|$, for a matrix $A \in R^{n\times m}$, will be the norm defined by $\|A\| = \{\max \|Au\| : u \in R^n, \|u\| = 1\}$. We will use B to denote the unit ball both in R^n and in $R^{n\times m}$. For a constant $\delta > 0$, B_δ will denote δB. More specifically, a closed ball in R^n with center x and radius δ will be represented by $B(x, \delta)$. For a mapping $F : R^n \to R^m$, we will use D_F to denote the set of points at which F is differentiable.

We will also use frequently the following notion.

A subset of matrices $\mathcal{A} \subset R^{n\times n}$ is said to be nonsingular if every $A \in \mathcal{A}$ is nonsingular.

A set-valued mapping $\mathcal{A}: R^n \longrightarrow 2^{R^{n\times m}}$ is said to be

(a) closed if for $x_k \to x$, $\xi_k \in \mathcal{A}(x_k)$, $\xi_k \to \xi$, then $\xi \in \mathcal{A}(x)$;
(b) compact if $\mathcal{A}(x)$ is compact for every $x \in R^n$;
(c) Hausdorff upper semi-continuous if, for every $x \in R^n$, $\epsilon > 0$ there exists a $\delta > 0$ such that for all $y \in B(x, \delta)$, $\mathcal{A}(y) \subset \mathcal{A}(x) + \epsilon B$.

Let $F : R^n \to R^m$ be a locally Lipschitz mapping. The Rademacher theorem guarantees that F is differentiable almost everywhere in R^n. Clarke [1] introduced the *generalized Jacobian* of F at a point $x \in R^n$ by

$$\partial F(x) = \text{conv}\{\lim_{x_i\in D_F;\, x_i\longrightarrow x} \nabla F(x_i)\},$$

where 'conv' denotes the closed convex hull. Obviously $\partial F(x)$ can be represented as the convex hull of $\partial_B F(x)$, where

$$\partial_B F(x) = \{\lim_{x_i\in D_F;\, x_i\longrightarrow x} \nabla F(x_i)\}.$$

$\partial_B F(x)$ was introduced by Qi in [23] and called B-subdifferential.

The following is well-known.

Proposition 1. *Let $F : R^n \to R^m$ be a locally Lipschitz mapping and $\partial F(x + B_\delta) = \bigcup_{x'\in B(x,\delta)} \partial F(x')$. Then*

$$\lim_{\delta\to 0} \partial F(x + B_\delta) = \bigcap_{\delta>0} \partial F(x + B_\delta) = \partial F(x).$$

An analytic expression of the support function of $\partial F(x)$ was given by Hiriart-Urruty [11] and further discussed by Ralph [26] in Banach space.

Proposition 2. *([11, Theorem 2.1]) Let $F : R^n \to R^m$ be a locally Lipschitz mapping, and $x \in R^n$. Then for every $a, b \in R^n$,*

$$\max_{V\in\partial F(x)} \langle Va, b\rangle = F^o(x; a, b),$$

where

$$F^o(x; a, b) = \lim_{y\longrightarrow x;\, t\longrightarrow 0} \sup\langle F(y + ta) - F(y), b\rangle/t.$$

From the above proposition, it follows that for every $V \in \partial F(x)$, and vector $a, b \in R^n$,

$$\langle Va, b\rangle \le F^o(x; a, b). \tag{1}$$

An interesting question is: if there exists a matrix V satisfying (1) for all $a, b \in R^n$, does $V \in \partial F(x)$ hold? To answer this question, we need the notion of 'plenary set' which was introduced by Sweetser [32].

2.2. Plenary sets

A subset of matrices $\mathcal{A} \subset R^{m\times n}$ is said to be plenary if and only if it includes every $A \in R^{m\times n}$ such that

$$Ab \in \mathcal{A}b, \text{ for all } b \in R^n.$$

It follows immediately from the definition that the intersection of plenary sets is also plenary. Let $\mathcal{A} \subset R^{m\times n}$ be a subset of matrices. Define

$$plen\mathcal{A} = \{A \in R^{m\times n} : Ab \in \mathcal{A}b, \forall b \in R^n\}. \tag{2}$$

From the definition, it follows that if $\mathcal{A}$ is a nonempty set, then $plen\mathcal{A}$ is also a nonempty set. As expected, $plen\mathcal{A}$ contains in general more matrices than $\mathcal{A}$. See an example by Sweetser in [32]. Moreover, the operator *plen* has some interesting properties.

Proposition 3. *Let $\mathcal{A}, \mathcal{B}$ be two subset of matrices of $R^{m\times n}$, and α be a scalar. Then*

(a) if $\mathcal{A} \subset \mathcal{B}$, then $plen\mathcal{A} \subset plen\mathcal{B}$;
(b) $plen(\alpha\mathcal{A}) = \alpha\ plen\mathcal{A}$;
(c) if $\mathcal{A}$ is bounded, then $plen\mathcal{A}$ is also bounded;
(d) if $\mathcal{A}$ is closed, then $plen\mathcal{A}$ is also closed;
(e) if $m = n$ and $\mathcal{A}$ is nonsingular, then $plen\mathcal{A}$ is also nonsingular.

Proof. Some of the results are obvious or well-known, we include a proof here for completeness.

(a) Let $A \in plen\mathcal{A}$. By definition, for every $a \in R^n$, $Aa \in \mathcal{A}a$. Since $\mathcal{A} \subset \mathcal{B}$, then $Aa \in \mathcal{B}a$, for all $a \in R^n$, which implies $A \in plen\mathcal{B}$.
(b) Let $A \in plen(\alpha\mathcal{A})$. The equality holds for $\alpha = 0$. Assume now $\alpha \neq 0$. By definition, for every $a \in R^n$, $Aa \in \alpha\mathcal{A}a$, or equivalently $\frac{1}{\alpha}Aa \in \mathcal{A}a$, hence $\frac{1}{\alpha}A \in plen\mathcal{A}$, which implies $A \in \alpha\mathcal{A}$.
(c) Let S be a bounded set of R^n, since $\mathcal{A}$ is bounded, then $\mathcal{A}S$ is bounded. By definition, $plen\mathcal{A}S \subset \mathcal{A}S$, which implies that $plen\mathcal{A}$ is bounded. Thus, $plen\mathcal{A}$ is bounded.
(d) Let $A_k \in plen\mathcal{A}_k$, and $A_k \to A$. We need to show that $A \in plen\mathcal{A}$. By definition, for each $a \in R^n$, $A_k a \in \mathcal{A}a$. Since $\mathcal{A}$ is closed, then $\mathcal{A}a$ is also closed. Hence $Aa \in \mathcal{A}a$, which implies that $A \in plen\mathcal{A}$.
(e) By contradiction, assume that $plen\mathcal{A}$ is singular. Then there exists $A_0 \in plen\mathcal{A}$, and $a_0 \in R^n$, $a_0 \neq 0$ such that $A_0 a_0 = 0$. By definition, $A_0 a_0 \in \mathcal{A}a_0$, which contradicts the fact $\mathcal{A}$ is nonsingular.

The proof is complete. □

Proposition 4. *Let $\{\mathcal{A}_\epsilon\}_{\epsilon>0} \subset R^{m\times n}$ be compact and monotonic increasing. Suppose that $\mathcal{A}_\epsilon \to \mathcal{A}$ as $\epsilon \to 0$. Then $plen\mathcal{A}_\epsilon \to plen\mathcal{A}$, as $\epsilon \to 0$.*

Proof. Let $\mathcal{A}_\epsilon \to \mathcal{A}$ as $\epsilon \to 0$. Then $\bigcap_{\epsilon>0} \mathcal{A}_\epsilon = \mathcal{A}$. We first prove that for every $a \in R^n$,

$$\bigcap_{\epsilon>0} \mathcal{A}_\epsilon a = (\bigcap_{\epsilon>0} \mathcal{A}_\epsilon)a. \tag{3}$$

Let $u \in (\bigcap_{\epsilon>0} \mathcal{A}_\epsilon)a$. Then there exists $A \in \bigcap_{\epsilon>0} \mathcal{A}_\epsilon$ such that $u = Aa$. Note that $A \in \mathcal{A}_\epsilon$, for every $\epsilon > 0$. Then $u = Aa \in \mathcal{A}_\epsilon a$, for every $\epsilon > 0$. Thus $u \in \bigcap_{\epsilon>0} \mathcal{A}_\epsilon a$. Conversely, let $u \in \bigcap_{\epsilon>0} \mathcal{A}_\epsilon a$. Then $u \in \mathcal{A}_\epsilon a$, for every $\epsilon > 0$. Let $A_\epsilon \in \mathcal{A}_\epsilon$ be such that $u = A_\epsilon a$, for $\epsilon > 0$. Since $\mathcal{A}_\epsilon$ is compact and monotonic decreasing as $\epsilon \to 0$, then $\mathcal{A}$ is nonempty, compact and every accumulation matrix A of $\mathcal{A}_\epsilon$ belongs to $\mathcal{A}$. Obviously $u = Aa \in \mathcal{A}a = (\bigcap_{\epsilon>0} \mathcal{A}_\epsilon)a$. This proves (3). Now let $a \in R^n$ be an arbitrary vector. Note that $plen\mathcal{A}_\epsilon a = \mathcal{A}_\epsilon a$ and $\bigcap_{\epsilon>0} plen\mathcal{A}_\epsilon a = \bigcap_{\epsilon>0} \mathcal{A}_\epsilon a$. Since $\mathcal{A}_\epsilon$ is compact, from Proposition 3 (c) and (d), it follows that $plen\mathcal{A}_\epsilon$ is also compact. Thus (3) holds for $\{plen\mathcal{A}_\epsilon\}$. Consequently $(\bigcap_{\epsilon>0} plen\mathcal{A}_\epsilon)a = (\bigcap_{\epsilon>0} \mathcal{A}_\epsilon)a = \mathcal{A}a$. Note that $\bigcap_{\epsilon>0} plen\mathcal{A}_\epsilon$ is a plenary set. Then $\bigcap_{\epsilon>0} plen\mathcal{A}_\epsilon = plen\mathcal{A}$. Hence $\lim_{\epsilon\to 0} plen\mathcal{A}_\epsilon = plen\mathcal{A}$. The proof is complete.

□

Proposition 5. *Suppose that $\mathcal{A} : R^n \to 2^{R^{m\times n}}$ is a closed set-valued mapping. Then $plen\mathcal{A}$ is also closed.*

Proof. Let $x_k \to x$, $A_k \in plen\mathcal{A}(x_k)$, and $A_k \to A$. It suffices to show that $A \in plen\mathcal{A}(x)$, or equivalently, $Aa \in \mathcal{A}(x)a$, for all $a \in R^n$. Note that for each a, $A_k a \in \mathcal{A}(x_k)a$, $A_k a \to Aa$, since $\mathcal{A}a$ is closed, then $Aa \in A(x)a$. The proof is complete.

□

Corollary 1. *Suppose that $\mathcal{A} : R^n \to 2^{R^{m\times n}}$ is Hausdorff upper semi-continuous at x and $\mathcal{A}(B(x, \delta))$ is bounded for some $\delta > 0$. Then $plen\mathcal{A}$ is also Hausdorff upper semi-continuous at x.*

It is an open question whether Hausdorff continuity is retained under operation *plen*.

2.3. Ioffe-Ralph generalized Jacobian and weak approximations

We are now ready to discuss (1). Hiriart-Urruty [11] obtained a relation between the matrices satisfying (1) and the Clarke generalized Jacobian.

Proposition 6. *Let $F : R^n \to R^m$ be a locally Lipschitz mapping and $\mathcal{C}F(x) = \{V : \langle Va, b\rangle \le F^o(x; a, b)\}$. Then*

(a) the following relation holds:

$$\mathcal{C}F(x) = plen\partial F(x); \tag{4}$$

(b) $\mathcal{C}F$ is Hausdorff upper semi-continuous.

Proof. Part (a) was proved in [11]. Part (b) is a well-known result. See for example [9, 26]. However, we note that the result also follows directly from Corollary 1. The proof is complete.

□

The existence and property of $\mathcal{C}F(\cdot)$ were extensively discussed in Banach space, see for example [12, 26]. We call $\mathcal{C}F(x)$ *Ioffe-Ralph generalized Jacobian* as it was independently derived by Ioffe [12] and Ralph [26]. In what follows, we discuss approximations to the Ioffe-Ralph generalized Jacobian.

Definition 1. *Let $F : R^n \to R^m$ be a locally Lipschitz mapping and X be a compact subset of R^n. We say that $\{\mathcal{A}_\epsilon F\}_{\epsilon>0}$ is a weak γ, μ-approximation to $\partial F(x)$ if, for every $\gamma > 0, \mu > 0$, there exists an $\epsilon > 0$ such that, for all $x \in X$,*

$$\partial F(x)a \subset \mathcal{A}_\epsilon F(x)a \subset \partial F(x + B_\gamma)a + B_\mu a, \forall a \in R^n. \tag{5}$$

Furthermore, we say that $\{\mathcal{A}_\epsilon F\}_{\epsilon>0}$ is monotonic if for two positive constants ϵ_1, ϵ_2, $\epsilon_1 < \epsilon_2$,

$$\mathcal{A}_{\epsilon_1} F(x) \subset \mathcal{A}_{\epsilon_2} F(x), \forall x \in X.$$

With this definition, we are able to find an approximation to the Ioffe-Ralph generalized Jacobian.

Theorem 1. *Let $F : R^n \to R^m$ be a locally Lipschitz mapping and X be a compact subset of R^n. Suppose that $\{\mathcal{A}_\epsilon F\}_{\epsilon>0}$ is a weak γ, μ-approximation to ∂F on X. Then*

(a) for every $\gamma > 0, \mu > 0$, there exists an $\epsilon > 0$ such that

$$\mathcal{C}F(x) \subset plen\mathcal{A}_\epsilon F(x) \subset plen(\partial F(x + B_\gamma) + B_\mu); \forall x \in X; \tag{6}$$

(b) if, in addition, $\{\mathcal{A}_\epsilon F\}_{\epsilon>0}$ is monotonic, then

$$\lim_{\epsilon\to 0} plen\mathcal{A}_\epsilon F(x) = \mathcal{C}F(x), \forall x \in X.$$

Proof. Note that part (a) follows directly from (2) and (5). We prove part (b). By Proposition 1 and Proposition 4, it follows that

$$\lim_{\gamma\to 0,\mu\to 0} plen(\partial F(x + B_\gamma) + B_\mu) = plen\partial F(x), \forall x \in X.$$

By part (a), for each $\gamma > 0, \mu > 0$, there exists a positive constant ϵ depending on γ and μ such that (6) holds. Note that $\{\mathcal{A}_\epsilon F\}_{\epsilon>0}$ is monotonic. Then

$$\lim_{\epsilon\to 0} plen\mathcal{A}_\epsilon F(x) = \bigcap_{\epsilon>0} plen\mathcal{A}_\epsilon F(x) = \mathcal{C}F(x), \forall x \in X.$$

The proof is complete.

□

Corollary 2. *Let $F : R^n \to R^m$ be a locally Lipschitz mapping and X be a compact subset of R^n, let $x \in X$. Suppose that $\{\mathcal{A}_\epsilon F\}_{\epsilon>0}$ is a weak monotonic γ, μ-approximation to ∂F on X and $\partial F(x)$ is nonsingular. Then for ϵ sufficiently small, $plen\mathcal{A}_\epsilon F(x)$ is also nonsingular.*

Practically, we need to know how to find an approximation to $\mathcal{C}F(x)$ or equivalently a weak γ, μ-approximation to ∂F. In [37], Xu, Rubinov and Glover introduced a kind of strong continuous approximations to ∂F which will be useful here.

Let $F : R^n \to R^m$ be a locally Lipschitz mapping, and X be a compact set in R^n. $\{\mathcal{A}_\epsilon F\}_{\epsilon>0}$ is said to be a *strong continuous approximation* to ∂F on X if

(a) for given $\epsilon > 0, \sigma > 0$, there exists $\tau > 0$ such that

$$\partial F(x + B_\tau) \subset \mathcal{A}_\epsilon F(x) + B_\sigma, \forall x \in X;$$

(b) for each $x \in X$, and all $0 < \epsilon_1 < \epsilon_2$,

$$\mathcal{A}_{\epsilon_1} F(x) \subset \mathcal{A}_{\epsilon_2} F(x);$$

(c) for $\epsilon > 0$, $\mathcal{A}_\epsilon F(x)$ is Hausdorff continuous with respect to x in X, that is, for every $\sigma > 0$, there exists a $\tau > 0$ such that for $x, y \in X$, $\|x - y\| \le \tau$,

$$\mathcal{A}_\epsilon F(y) \subset \mathcal{A}_\epsilon F(x) + B_\sigma$$

and

$$\mathcal{A}_\epsilon F(x) \subset \mathcal{A}_\epsilon F(y) + B_\sigma;$$

(d) for every $\gamma > 0$, $\mu > 0$, there exists an $\epsilon > 0$, such that, for all $x \in X$

$$\begin{aligned}\partial F(x) &\subset \mathcal{A}_\epsilon F(x)\\ &\subset \partial F(x + B_\gamma) + B_\mu.\end{aligned}$$

In [37], strong continuous approximations have been proposed for some important locally Lipschitz mappings. For details, see [37].

Remark 1. A strong continuous approximation is a weak monotonic γ, μ-approximation.

Definition 2. *Let $F : R^n \to R^m$ be a locally Lipschitz mapping and X be a compact subset of R^n. We say that a weak γ, μ-approximation $\{\mathcal{A}_\epsilon F\}_{\epsilon>0}$ is pertinent on X if there exists a function $p : R^2_+ \to R_+$ such that for every $\gamma > 0, \mu > 0, x \in X$,*

$$\begin{aligned}\partial F(x)a &\subset \mathcal{A}_{p(\gamma,\mu)} F(x)a\\ &\subset \partial F(x + B_\gamma)a + B_\mu a, \forall a \in R^n,\end{aligned} \tag{7}$$

where p is called pertinence function.

In what follows, we present a simple way for constructing a pertinence function.

We say that $\bar{p} : R^2_+ \to R_+$ is *strictly increasing* if

$$\bar{p}(t_1, u_1) < \bar{p}(t_2, u_2),$$

when $0 \le t_1 < t_2, 0 \le u_1 < u_2$, and $\bar{p}(0, u)) = \bar{p}(t, 0) = 0$, for $t \ge 0, u \ge 0$.

Let $\gamma_0, \mu_0 \in (0, 1), \mathcal{N}$ denote the set of natural numbers.

$$\mathcal{P} = \{\bar{p}(\gamma_0^k, \mu_0^l) : k, l \in \mathcal{N} \bigcup \{0\}\},$$

$$\begin{aligned}p_{F,\mathcal{A},X}(\gamma, \mu) = \max\{\bar{p} \in \mathcal{P} : \partial F(x)a &\subset \mathcal{A}_{\bar{p}} F(x)a\\ &\subset \partial F(x + B_\gamma)a + B_\mu a, \forall a \in R^n, x \in X\}.\end{aligned} \tag{8}$$

We have the following remark.

Remark 2. If $\{\mathcal{A}_\epsilon F\}_{\epsilon>0}$ is a weak monotonic γ, μ-approximation to ∂F on X, then for every $x \in X$, p defined by (8) is a pertinence function and $\{\mathcal{A}_p F\}$ is pertinent. More specifically, if $\{\mathcal{A}_\epsilon F\}_{\epsilon>0}$ is a strong continuous approximation to ∂F, then for every $x \in X$, p defined by (8) is a pertinence function which is also strictly increasing.

3. Point-based set-valued approximation

We now turn to discuss main approximation of this paper. Let $V \in R^{m\times n}$ be a matrix and $h \in R^n$ be a nonzero vector. For convenience, let

$$\Gamma_F(x, h, V) = (F(x+h) - F(x) - Vh)/\|h\|.$$

3.1. Basic definition and properties

Definition 3. *Let D be an open subset of R^n. A mapping $F : D \subset R^n \to R^m$ is said to have a Point-Based Set-Valued Approximation (PBSVA for short) at a point $x^* \in D$, if there exists a set-valued mapping $\mathcal{A}F : R^n \to 2^{R^{m\times n}}$, such that for every $\epsilon > 0$, there exists a $\delta > 0$ such that*

$$\sup_{A(x)\in\mathcal{A}F(x)} \|\Gamma_F(x^*, x - x^*, A(x))\| \le \epsilon, \text{ for all } x \in S(x^*, \delta). \tag{9}$$

The collection of all set-valued mapping $\mathcal{A}F$ satisfying (9) is denoted by $\mathcal{M}(F, x^*)$.

Robinson first considered a point-based approximation in [30], his approximation is single-valued and usually nonlinear if F is not smooth. The originality of the definition of point-based approximation lies in the fact that $A(x)$ depends on x. Kummer [14] further considered point-based approximations which are derivative-related and set-valued. Gabriel and Pang [8] used a compact and Hausdorff upper semicontinuous set-valued mapping $\mathcal{A}F \in \mathcal{M}(F, x^*)$ in the study of trust region method for constrained nonsmooth equations. Recently, Qi [24] treated such a mapping as a differential operator, namely *C-differential operator*.

A natural question is: under which conditions is $\mathcal{M}(F, x^*)$ nonempty?

Example 1. [1] Let $\mathcal{A}_N F = \{A(\cdot) : A(x) = A_N F(x), x \in R^n\}$, where

$$A_N F(x) = \begin{cases} (F(x) - F(x^*))\frac{(x-x^*)^T}{\|x-x^*\|^2}, & \text{if } x \ne x^*; \\ 0, & \text{if } x = x^*. \end{cases}$$

Then $\mathcal{A}_N F \in \mathcal{M}(F, x^*)$.

This example shows that every mapping F (not necessarily locally Lipschitz) has a *PBSVA*. Unfortunately, $\mathcal{A}_N$ seems to be trivial in multi-dimensional case as it is of rank one. We will discuss later on some useful *PBSVA*s.

[1] This example was given by an anonymous referee during an earlier submission of this paper.

Proposition 7. *Let $\mathcal{A}F \in \mathcal{M}(F, x^*)$. Then*

$$plen\mathcal{A}F \in \mathcal{M}(F, x^*),$$

where plen is defined in (2).

The proof is straightforward.

Let $\bar{r}$ denote $\{1, \cdots, r\}$, for natural number r.

Proposition 8. *Let $F_i : R^n \to R^m$, $i \in \bar{r}$, be mappings and c_i, $i \in \bar{r}$ be constants. Let $F = \Sigma_{i\in\bar{r}} c_i F_i$, and $\mathcal{A}F = \Sigma_{i\in\bar{r}} c_i \mathcal{A}F_i$. If $\mathcal{A}F_i \in \mathcal{M}(F_i, x^*)$, then*

$$plen\mathcal{A}F \in \mathcal{M}(F, x^*). \tag{10}$$

Proof. It is obvious that if $\mathcal{A}F_i \in \mathcal{M}(F_i, x^*)$, then $c_i\mathcal{A}F_i(\cdot) \in \mathcal{M}(c_i F_i, x^*)$, for every constant c_i. Note also that r is finite. Then it suffices to show that the conclusion holds for $F = F_1 + F_2$. Let $\mathcal{A}F = \mathcal{A}F_1 + \mathcal{A}F_2$. Then

$$\begin{aligned}
&\max_{A(x)\in\mathcal{A}F(x)} \|\Gamma_F(x^*, x - x^*, A(x))\| \\
&\le \max_{A_1(x)\in\mathcal{A}F_1(x), A_2(x)\in\mathcal{A}F_2(x)} \|\Gamma_{F_1+F_2}(x^*, x - x^*, A_1(x) + A_2(x))\| \\
&\le \max_{A_1(x)\in\mathcal{A}F_1(x)} \|\Gamma_{F_1}(x^*, x - x^*, A_1(x))\| + \max_{A_2(x)\in\mathcal{A}F_2(x)} \|\Gamma_{F_2}(x^*, x - x^*, A_2(x))\| \\
&= 2\epsilon, \text{ for all } x \in S(x^*, \delta),
\end{aligned}$$

where δ is some positive number. The proof is complete.

□

Proposition 9. *Let $F : R^n \to R^m$, $R^m = R^r \times R^s$ and $F = (F_r^T, F_s^T)^T$. If $\mathcal{A}F_i \in \mathcal{M}(F_i, x^*)$, for $i = r, s$, then*

$$\mathcal{A}F_r \times \mathcal{A}F_s \in \mathcal{M}(F, x^*).$$

We omit the proof as it is straightforward.

We now consider some specific mappings.

3.2. Compositions involving locally Lipschitz mappings

First we consider the following inner smooth composite mapping:

$$F = G \circ H, \tag{11}$$

where $G : R^n \to R^m$ is locally Lipschitz, and $H : R^s \to R^n$ is continuously differentiable. (11) was considered in [37] and it includes many important mappings in practical instance such as piecewise C^r mappings considered by Ralph and Scholtes [28].

Theorem 2. *Let F be defined by (11), $x^* \in R^s$, and $\mathcal{A}G \in \mathcal{M}(G, H(x^*))$. If $\mathcal{A}G$ is compact, then $\mathcal{A}G(H)\nabla H \in \mathcal{M}(F, x^*)$.*

Proof. Let $B \in \mathcal{A}G(H)\nabla H$. Then there exists $A_G \in \mathcal{A}G(H)$ such that $B = A_G(H)\nabla H$. Let $\delta > 0$ be sufficiently small. Then, for $x \in B(x^*, \delta)$,

$$\|\Gamma_F(x^*, x-x^*, B(x))\| \leq \|\Gamma_G(H(x^*), H(x)-H(x^*), A_G(H(x)))\| \|H(x)-H(x^*\|/\|x-x^*\| + \|A_G(H(x))\Gamma_H(x^*, x-x^*, \nabla H(x))\|.$$

Note that $\mathcal{A}G$ is a compact set-valued mapping and $B(x^*, \delta)$ is a compact set. Then $\mathcal{A}G(H(S(X^*, \delta)))$ is bounded. Note also that H is continuously differentiable. Thus,

$$\|A_G(H(x))\Gamma_H(x^*, x-x^*, \nabla H(x))\| = o(\|x-x^*\|).$$

The rest is straightforward.

□

We now consider an outer smooth composite mapping:

$$F = Q \circ P, \tag{12}$$

where $Q : R^m \to R^l$ is continuously differentiable, and $P : R^n \to R^m$ is locally Lipschitz. (12) was also considered in [37] and it includes some important mappings in practical instances such as the norm of normal mappings considered by Ferris and Ralph [7].

Theorem 3. *Let F be defined by (12), $x^* \in R^n$, and $\mathcal{A}P \in \mathcal{M}(P, x^*)$. Then $\nabla Q(P)\mathcal{A}P \in \mathcal{M}(F, x^*)$.*

Proof. Let $D \in \nabla Q(P)\mathcal{A}P$. Then there exists $A_P \in \mathcal{A}P$ such that $D = \nabla Q(P)A_P$. Let $\delta > 0$ be sufficiently small. Then, for $x \in B(x^*, \delta)$, we have

$$\|\Gamma_F(x^*, x-x^*, D(x))\| \leq \|\Gamma_Q(P(x^*), P(x)-P(x^*), \nabla Q(P(x)))\| \|P(x)-P(x^*\|/\|x-x^*\| + \|\nabla Q(P(x))\Gamma_H(x^*, x-x^*, A_P(x))\|.$$

Since P is continuous and Q is continuously differentiable, then $\nabla Q(P(S(x^*, \delta)))$ is bounded and

$$\|\nabla Q(P(x))\Gamma_P(x^*, x-x^*, A_P(x))\| = o(\|x-x^*\|).$$

The rest is straightforward.

□

Qi [24] discussed the calculus of C-differential operators. Note that a C-differential operator is a compact and semi-continuous *PBSVA*, while a *PBSVA* is not necessarily a C-differential operator. For instance, $\mathcal{A}_N F$ in Example 1 is generally not semi-continuous. It is yet not clear whether a general continuous mapping F admits a C-differential operator.

It seems that the chain rules developed here could play very limited role in Newton's method. Consider the composite mapping (11). Suppose we know that, at a solution point x^* of $F(x) = 0$, $\partial F(x^*)$ is nonsingular. It is expected that a *PBSVA* of F at x^* is also nonsingular in a neighborhood of x^*. However, generally this is incorrect for the *PBSVA* $\mathcal{A}G(H)\nabla H$ in Theorem 2. The reason is that $\mathcal{M}(F, x^*)$ is too large. This comment also applies to C-differential operators.

We now discuss some more specific locally Lipschitz mappings.

3.3. Semi-smooth locally Lipschitz mappings

Let $F : R^n \to R^m$ be a locally Lipschitz mapping. F is said to be semi-smooth at $x \in R^n$ if for every $h \in R^n$, the following limit

$$\lim_{V \in \partial F(x+th');\, h' \longrightarrow h;\, t \longrightarrow 0} Vh'$$

exists. See [25].

Lemma 1. *See [16, Proposition 2]. Let $F : R^n \to R^m$ be a locally Lipschitz mapping, and $x^* \in R^n$. If F is semi-smooth at x^*, then*

$$\lim_{x \to x^*} \|\Gamma_F(x^*, x - x^*, V)\| = 0,$$

for every $V \in \partial F(x)$.

Theorem 4. *Let $F : R^n \to R^m$ be a locally Lipschitz mapping, and $x^* \in R^n$. Suppose that F is semi-smooth at x^*. Then the following hold.*

(a) $\partial F \in \mathcal{M}(F, x^*)$*;*
(b) $\partial_B F \in \mathcal{M}(F, x^*)$.

Proof. Part (a). Since F is semi-smooth at x^*, by Lemma 1, we have

$$\lim_{x \to x^*,\, V \in \partial F(x)} \|\Gamma_F(x^*, x - x^*, V)\| = 0.$$

Thus $\partial F \in \mathcal{M}(F, x^*)$.

Part (b) is obvious since $\partial_B F(x)$ is only a subset of $\partial F(x)$. The proof is complete.

□

The importance of Theorem 4 lies in the fact that under the assumption of semi-smoothness a *PBSVA* of a locally Lipschitz mapping can be obtained through the Clarke generalized Jacobian or B-subdifferential of the mapping. The latter have been intensively investigated and found useful in the solution of nonsmooth equations. With Theorem 4 and Proposition 7, we can easily obtain the following.

Corollary 3. *Let $F : R^n \to R^m$ be a locally Lipschitz mapping, and $\mathcal{C}F(x)$ be given by (4), let $x^* \in R^n$. If F is semi-smooth at $x \in R^n$, then the following hold:*

(a) $\mathcal{C}F \in \mathcal{M}(F, x^*)$;
(b) $plen\partial_B F \in \mathcal{M}(F, x^*)$.

Thus we have proved that the Ioffe-Ralph generalized Jacobian and the plenary hull of B-subdifferential are *PBSVA*s.

Example 2. See for example [37]. We consider the following mappings:

$$F(x) = \min_{j \in \bar{r}} H_j(x), \tag{13}$$

where $H_j : R^n \to R^m$, $j \in \bar{r}$, are continuously differentiable, and 'min' is taken componentwise. Obviously (13) can be regarded as a composite mapping in the form of (11) with

$$G(Y) = \min_{j \in \bar{r}} y_j,$$

$Y = (y_1^T, \cdots, y_r^T)^T$, and $H = (H_1^T, \cdots, H_r^T)^T$. Note that $G(Y)$ is piecewise linear. Then $G(Y)$ is semi-smooth in R^{mr}. Let $Y^* \in R^{mr}$ be a fixed vector. Then by Theorem 4 (a), there exists a δ-neighborhood of Y^* such that $\partial G \in \mathcal{M}(G, Y^*)$, where

$$\partial G(Y) = \text{conv}\{[\{e_{ij}, j \in J_i\}, i \in \bar{r}]\},$$

$J_i = \{j \in \bar{r}, y_{ij} = \min_{j \in \bar{r}} y_{ij}\}$, $e_{ij} \in R^{mr}$ with the $((j-1)m + i)$-th component one, and the others zero. If $H(x^*) = Y^*$, then by Theorem 2, we have

$$\partial G(H) \nabla H \in \mathcal{M}(F, x^*).$$

Example 3. We now consider another mapping:

$$F(x) = Q(x_+) + x - x_+, \tag{14}$$

where $Q : R^n \to R^m$ is continuously differentiable, x_+ is the Euclidean projection of x onto R^n_+. (14) defines a normal map. See [5–7, 10, 27–29]. Clearly x_+ is locally Lipschitz semi-smooth in R^n. Let $F_1(x) = Q(x_+)$, $F_2 = x - x_+$, $x^* \in R^n$. By virtue of Theorem 3, we have

$$\nabla Q(x_+) \partial x_+ \in \mathcal{M}(F_1, x^*),$$

where

$$\partial x_+ = (\partial (x_i)_+, i \in \bar{m}),$$

$$\partial (x_i)_+ = \begin{cases} e_i, & \text{if } x_i > 0; \\ 0, & \text{if } x_i < 0; \\ \text{conv}\{0, e_i\}, & \text{if } x_i = 0. \end{cases}$$

Here e_i is an n-dimensional vector with the i-th component one and the rest zero. Likewise, we can calculate ∂F_2 and using Proposition 8, we can easily obtain a *PBSVA* of F at x^*.

3.4. First order approximations

The *PBSVA*s presented in the previous subsection are based on generalized Jacobians. However, in many practical instances, it is more convenient to consider the approximations of the generalized Jacobians. In what follows, we will prove under mild conditions that approximations to Ioffe-Ralph generalized Jacobian introduced in Section 2 can be used to construct a *PBSVA*.

Let $\mathcal{E}$ be the set of real-valued functions $\{e : R_+ \to R_+, e(0) = 0, \lim_{t\to 0} e(t)/t = 0\}$. Let $\epsilon > 0$ be a constant. Recall that the ϵ-generalized Jacobian of F at $x \in R^n$ is defined as:

$$\partial_\epsilon F(x) = \text{conv} \bigcup_{x' \in B(x,\epsilon)} \partial F(x'), \tag{15}$$

See for example [35]. The following was also proved in [35].

Lemma 2. *See [35, Lemma 3.2]. Let $F : R^n \to R^m$ be a locally Lipschitz mapping and e be a real-valued function from set $\mathcal{E}$, let $x^* \in R^n$. If F is semi-smooth at x^*, then for every matrix $U \in \partial_{e(\|x-x^*\|)} F(x)$, there exists a matrix $V \in \partial F(x)$ such that*

$$\lim_{x\to x^*} (U - V)(x - x^*)/\|x - x^*\| = 0. \tag{16}$$

Theorem 5. *Let $F : R^n \to R^m$ be a locally Lipschitz mapping and e be a real-valued function from set $\mathcal{E}$, let $x^* \in R^n$ and*

$$\mathcal{A}_e F = \{A(\cdot) : A(x) \in \partial_{e(\|x-x^*\|)} F(x)\}. \tag{17}$$

If F is semi-smooth at x^*, then $\mathcal{A}_e F \in \mathcal{M}(F, x^*)$.

Proof. Since F is semi-smooth at x^*, by Lemma 1,

$$\lim_{x\to x^*} \max_{V\in\partial F(x)} \|\Gamma_F(x^*, x - x^*, V)\| = 0. \tag{18}$$

Let $U \in \partial_{e(\|x-x^*\|)} F(x)$. It follows from Lemma 2 that there exists a matrix $V \in \partial F(x)$ such that (16) holds. Combining (16) with (18), we have

$$\begin{aligned}\lim_{x\to x^*} \|\Gamma_F(x^*, x - x^*, U)\| &\le \lim_{x\to x^*} \|\Gamma_F(x^*, x - x^*, V)\| \\ &+ \lim_{x\to x^*} \|(U - V)(x - x^*)\|/\|x - x^*\| \\ &= 0.\end{aligned}$$

This completes the proof. □

Theorem 6. *Let $F : R^n \to R^m$ be a locally Lipschitz mapping and e be a real-valued function from set $\mathcal{E}$, let X be a compact subset of R^n and $x^* \in X$. Suppose that*

$$\mathcal{A}_W F = \mathcal{A}_{p(e(\|\cdot-x^*\|), e(\|\cdot-x^*\|))} F(\cdot) \tag{19}$$

is a weak pertinent approximation to ∂F on X with a pertinence function p. If F is semi-smooth at x^*, then plen$\mathcal{A}_W F \in \mathcal{M}(F, x^*)$.

Proof. From Definition 2, it follows that

$$\begin{aligned}\partial F(x)(x-x^*) &\subset \mathcal{A}_{p(e(\|x-x^*\|),e(\|x-x^*\|))}F(x)(x-x^*)\\ &\subset (\partial F(x+B_{e(\|x-x^*\|)})+B_{e(\|x-x^*\|)})(x-x^*)\\ &= (\partial_{e(\|x-x^*\|)}F(x)+e(\|x-x^*\|)B)(x-x^*), \forall x \in R^n, \end{aligned} \quad (20)$$

Therefore

$$\begin{aligned}\mathcal{C}F(x) &\subset plen\mathcal{A}_W(x)\\ &\subset plen(\partial_{e(\|x-x^*\|)}F(x)+e(\|x-x^*\|)B), \forall x \in R^n,\end{aligned}$$

By Theorem 5, we know that $\partial_{e(\|\cdot-x^*\|)}F(\cdot) \in \mathcal{M}(F, x^*)$. It can also be easily checked that $e(\|\cdot-x^*\|)B \in \mathcal{M}(F, x^*)$. The rest follows directly from Proposition 8.

□

4. Newton's methods

In this section we shall propose a generalized Newton's method based on a point-based set-valued approximation for solving the following nonsmooth equations:

$$F(x) = 0, \quad (21)$$

where $F : R^n \to R^n$ is locally Lipschitz but it is not necessarily differentiable.

In the past few years there has been an increasing discussion on (21). Two main factors have stimulated the increase. One is that nonsmooth equations provide a unified framework for the study of many important problems in mathematical and equilibrium programming, the other is that on the basis of their nonsmooth equation framework, some new methods can be developed for solving optimization and equilibrium problems, these methods are not only highly efficient but also resolve the lack of robustness in many previous solution approaches. See [22] for details.

As far as numerical methods are concerned, there have been mainly two kinds of methods proposed: Newton's methods, see for instance [3, 14, 19, 22, 23, 25, 27, 30, 35, 36], and quasi-Newton methods, see [2, 13, 15].

Newton's methods have been developed in many ways. One way based on B-derivatives is to solve (21) by solving iteratively

$$F'(x_k, d_k) = 0,$$

at each iterate. In general, $F'(x_k, d_k)$ is nonsmooth and nonlinear in d_k. See [19, 20].

Another way based on the Clarke generalized Jacobian is to find a Newton step by

$$x_{k+1} = x_k - V_k^{-1}F(x_k), \quad (22)$$

where matrix $V_k \in \partial F(x_k)$ is supposed nonsingular. Local superlinear convergence was shown under that condition that F is semi-smooth at solution points. See [14, 25].

The third way introduced by Robinson [30] is to solve

$$A(x_k, d_k) = 0,$$

where $A(\cdot,\cdot)$ is Robinson's point-based approximation. This work was further extended in [8, 5, 6, 27].

Recently, Demyanov [3] also developed a Newton's method for nonsmooth equations with codifferentials.

In what follows, we try to derive Newton's method from the point-based set-valued approximation (9). Let x^* be a solution point of (21). Suppose that $\partial F(x^*)$ is nonsingular. Clearly the sequence generated by (22) in a neighbourhood of x^* converges superlinearly to x^* if and only if

$$\|F(x_k) - F(x^*) - V_k(x_k - x^*)\| = o(\|x_k - x^*\|),$$

or equivalently $\partial F \in \mathcal{M}(F, x^*)$. Replacing ∂F with a *PBSVA* $\mathcal{A}F \in \mathcal{M}(F, x^*)$, we can propose a more general Newton's method.

Theorem 7. *Let x^* be a solution point of (21). Suppose that there exists a $\mathcal{A}F \in \mathcal{M}(F, x^*)$ such that $\mathcal{A}F$ is Hausdorff upper semi-continuous at x^* and $\mathcal{A}(x^*)$ is compact and nonsingular. Then iteration*

$$x_{k+1} = x_k - A_k^{-1}F(x_k),\ \ A_k \in \mathcal{A}F(x_k) \tag{23}$$

is well defined and converges superlinearly to x^* in a neighbourhood of x^*.

We will not present a proof for this theorem since it can be regarded as a corollary of the following theorem.

Theorem 8. *Let x^* be a solution point of (21). Suppose that there exists a $\mathcal{A}F \in \mathcal{M}(F, x^*)$ such that $\mathcal{A}F$ is Hausdorff upper semi-continuous at x^* and $\mathcal{A}F(x^*)$ is compact and nonsingular. Then iteration*

$$x_{k+1} = x_k - A_k^{-1}F(x_k),\ \ A_k \in plen(\mathcal{A}F(x_k)) \tag{24}$$

is well defined and converges superlinearly to x^* in a neighbourhood of x^*.

Proof. By assumption, $\mathcal{A}F(x^*)$ is compact and nonsingular. Then there exists an $\epsilon_0 > 0$, such that $\mathcal{A}F(x^*) + \epsilon_0 B$ is also compact and nonsingular. From Proposition 3 (c-e), it follows that $plen(\mathcal{A}F(x^*) + \epsilon_0 B)$ is compact and nonsingular. Hence there exists a constant $C > 0$, such that

$$\max_{V \in plen(\mathcal{A}F(x^*)+\epsilon_0 B)} \|V^{-1}\| \le C. \tag{25}$$

Note that $\mathcal{A}F$ is Hausdorff upper semi-continuous at x^*. Then for given $\epsilon_0 > 0$, there exists a $\delta_0 > 0$ such that $\mathcal{A}F(x) \subset \mathcal{A}F(x^*) + \epsilon_0 B$ for all $x \in B(x^*, \delta_0)$. By virtue of Proposition 3 (a) and (25), we have, for all $x \in B(x^*, \delta_0)$,

$$\begin{aligned}\max_{V \in plen\mathcal{A}(x)} \|V^{-1}\| &\le \max_{V \in plen(\mathcal{A}(x^*)+\epsilon_0 B)} \|V^{-1}\| \\ &\le C.\end{aligned}$$

On the other hand, since $\mathcal{A}F \in \mathcal{M}(F, x^*)$, it follows from Proposition 7 that $plen\mathcal{A}F \in \mathcal{M}(F, x^*)$. The rest is straightforward. The proof is complete.

□

A similar result to Theorem 7 was obtained by Gabriel and Pang [8] in the study of trust region method for constrained nonsmooth equations. Qi [24] also obtained a similar result to Theorem 7 based on C-differential operators. The difference is that we assume here $\mathcal{A}F(x^*)$ is compact while $\mathcal{A}F \in \mathcal{M}(F, x^*)$ is not necessarily compact.

Note also that in Theorem 8 we only assume that $\mathcal{A}F(x)$ instead of $plen\mathcal{A}F(x)$ is nonsingular at $x = x^*$. In most cases, the former is strictly smaller than the latter. Thus, theoretically Theorem 8 is stronger than Theorem 7. In computation, note that, if $plen\mathcal{A}F(x) = plen\mathcal{B}(x)$, then we can use $\mathcal{B}(x)$ to replace $\mathcal{A}F(x)$ whenever the former can be more easily calculated. This may lead to some convenience in computation. See further comments in Section 5.

The following example illustrates that Theorem 7 and Theorem 8 are stronger than [25, Theorem 3.2].

Example 4. Consider the following function:

$$P(x) = \begin{cases} x + x^2 \sin\frac{1}{x}, & x \neq 0; \\ 0, & x = 0. \end{cases}$$

$x^* = 0$ is a solution of $P(x) = 0$. Let $\mathcal{A}P(x) = \{1 + x\sin\frac{1}{x}, x \neq 0\}$, $\mathcal{A}P(0) = \{1\}$. Clearly $P'(x) \notin \mathcal{A}P(x)$, for $x \neq 0$ and $P(x)$ is not semi-smooth at 0. However all conditions of Theorem 7 are satisfied for this problem.

5. Case studies

In this section, we will include some of generalized Newton's methods developed in the past few years in (23) and consolidate the convergence results of these methods with Theorem 8.

5.1. Ioffe-Ralph generalized Jacobian based Newton's method

Qi and Sun [25] proposed a generalized Newton iterative process (22) for finding a solution of (21). They obtained the following results.

Theorem 9. *([25, Theorem 3.2]) Let x^* be a solution of (21). Suppose that F is semi-smooth at x^* and $\partial F(x^*)$ is nonsingular. Then iteration (22) is well-defined and superlinearly convergent to x^* in a neighbourhood of x^*.*

By Theorem 4, it follows that Theorem 7 subsumes Theorem 9.

By referring to Theorem 8, we can easily extend Theorem 9 to the following.

Theorem 10. *Let x^* be a solution of (21) and $\mathcal{C}F(x)$ be given by (4). Suppose that F is semi-smooth at x^* and $\partial F(x^*)$ is nonsingular. Then iteration*

$$x_{k+1} = x_k - U_k^{-1}F(x_k), U_k \in \mathcal{C}F(x_k)$$

is well-defined and converges superlinearly to x^* in a neighbourhood of x^*.

Proof. The conclusion follows directly from Corollary 3 (a) and Theorem 8.

□

The significance of the above theorem is that, on one hand, it may result in some convenience in computation as we can now take an element of the plenary hull of Clarke generalized Jacobian in Newton's iteration; on the other hand, it shows theoretically that Newton's method can be carried out with Ioffe-Ralph generalized Jacobian under the same condition as that of [25, Theorem 3.2]. Note also that Qi [24] explicitly assumed that the image of a C-differential operator is nonsingular at a solution point x^* of (21). In this setting, this is equivalent to assuming that $\mathcal{C}F(x^*)$ is nonsingular. We have already pointed out previously that in general $\mathcal{C}F(x)$ is larger than $\partial F(x)$. Thus, we show that Theorem 8 is stronger than Theorem 7 and [24, Theorem 3.1].

5.2. B-subdifferential based Newton's method

In [23], a revised version of (22) was suggested by Qi:

$$x_{k+1} = x_k - W_k^{-1} F(x_k),\ W_k \in \partial_B F(x_k). \tag{26}$$

(26) is intended to reduce the assumptions for convergence of generalized Newton iteration (22) because $\partial_B F(x)$ is only a strict subset of $\partial F(x)$.

Theorem 11. *([23, Theorem 2]) Let x^* be a solution of (21). Suppose that F is semi-smooth at x^* and $\partial_B F(x^*)$ is nonsingular. Then iteration (26) is well-defined and superlinearly convergent to x^* in a neighbourhood of x^**

It was proved by Xu and Glover [36] that $\partial_B F$ is Hausdorff upper semi-continuous. By Theorem 4, $\partial_B F \in \mathcal{M}(F, x^*)$. Hence, Theorem 7 subsumes Theorem 11. Furthermore, it is an easy exercise to extend Theorem 11 into the plenary case by replacing $\partial_B F(x_k)$ in (26) with $plen\partial_B F(x_k)$.

5.3. ϵ-generalized Jacobian based Newton's method

Consider generalized Newton iteration (22). When one hopes to weaken the conditions for the convergence of (22), $\partial F(x_k)$ is replaced by $\partial_B F(x_k)$. However, when the importance of convenience in computing V_k of (22) is emphasized, a larger set than $\partial F(x_k)$ in (22) is sought so that more alternatives for updating matrices are available. This led Xu and Chang [35] to employ another kind of Newton iteration:

$$x_{k+1} = x_k - G_k^{-1} F(x_k),\ G_k \in \partial_{\epsilon_k} F(x_k),$$

where $\partial_\epsilon F(x)$ is defined by (15).

Xu and Chang [35] obtained the following results:

Theorem 12. *([35, Theorem 3.2]) Let x^* be a solution of (21) and $e \in \mathcal{E}$. Suppose that F is semi-smooth at x^*, and $\partial F(x^*)$ is nonsingular. Then the iteration:*

$$x_{k+1} = x_k - A_k^{-1} F(x_k),\ \ A_k \in \partial_{e(\|x_k - x^*\|)} F(x_k)$$

is well defined and converges superlinearly to x^* in a neighbourhood of x^*.

Theorem 12 is included in Theorem 7. Similarly, we can extend Theorem 12 to the following.

Theorem 13. *Let x^* be a solution of (21) and $\mathcal{A}_e F$ be defined by (17). Assume that F is semi-smooth at x^* and $\partial F(x^*)$ is nonsingular. Then the iteration:*

$$x_{k+1} = x_k - A_k^{-1} F(x_k), \;\; A_k \in plen \mathcal{A}_e F(x_k), \tag{27}$$

is well defined and converges superlinearly to x^* in a neighbourhood of x^*.

For a proof, we only need to verify the conditions of Theorem 8 in this setting.

Lemma 3. *Let $F : R^n \to R^m$ be a locally Lipschitz mapping and $\mathcal{A}_e F$ be defined by (17). Then $\mathcal{A}_e F$ is compact and Hausdorff upper semicontinuous at x^*.*

Proof. Obviously $\mathcal{A}_e F$ is compact. It suffices to show that for every $\epsilon > 0$, there exists $\delta > 0$ such that

$$\partial_{e(\|x-x^*\|)} F(x) \subset \partial F(x^*) + \epsilon B. \tag{28}$$

By definition,

$$\begin{aligned} \partial_{e(\|x-x^*\|)} F(x) &= \mathrm{conv}_{x' \in B(x, e(\|x-x^*\|))} \partial F(x') \\ &\subset \mathrm{conv}_{x' \in B(x^*, \|x-x^*\| + e(\|x-x^*\|))} \partial F(x') \end{aligned}$$

By Proposition 1, for $\epsilon > 0$, there exists a $\delta > 0$ such that

$$\partial F(x) \subset \partial F(x^*) + \epsilon B,$$

for all $x \in B(x^*, \delta)$. Note that $e(\|x - x^*\|) \to 0$ as $x \to x^*$. Letting $\|x - x^*\| + e(\|x - x^*\|) < \delta$, we have

$$\partial_{e(\|x-x^*\|)} F(x) \subset \mathrm{conv}(\partial F(x^*) + \epsilon B). \tag{29}$$

(29) implies (28) as $\partial F(x^*) + \epsilon B$ is convex. The proof is complete.

□

With Theorem 5 and Lemma 3, the proof of Theorem 13 is straightforward.

5.4. Weak approximation based Newton's methods

We now consider another kind of Newton's iterations based on (19).

Theorem 14. *Let x^* be a solution of (21) and $\mathcal{A}_W F$ is defined by (19). Suppose that F is semi-smooth at x^* and $\partial F(x^*)$ is nonsingular. If $\{\mathcal{A}_\epsilon F\}_{\epsilon > 0}$ is a family of weak pertinent approximation to ∂F on a compact set X containing a neighborhood of x^* with pertinence function p, then the iteration:*

$$x_{k+1} = x_k - A_k^{-1} F(x_k), \;\; A_k \in plen \mathcal{A}_W F(x_k), \tag{30}$$

is well defined and converges superlinearly to x^* in a neighbourhood of x^*.

Before presenting a proof, we give the following remarks.

Remark 3. We cannot prove this theorem by directly checking the conditions of Theorem 8 since $\mathcal{A}_W F$ is not necessarily Hausdorff upper semicontinuous at x^*. However as we show in the proof below, Theorem 8 can be applied to the iteration (30) with $\mathcal{G}$ replacing $\mathcal{A}_W F$, where

$$\mathcal{G}(x) = \partial F(x + B_{e(\|x-x^*\|)}) + B_{e(\|x-x^*\|)}. \tag{31}$$

Proof (Theorem 14). Let $\mathcal{G}(x)$ be defined by (31). Then it is easy to check that $\mathcal{G}$ is compact and Hausdorff upper semicontinuous at x^* with $\mathcal{G}(x^*) = \partial F(x^*)$. Further, similar to the proof of Theorem 5, we can show that $\mathcal{G} \in \mathcal{M}(F, x^*)$. Thus, by Theorem 8, iteration (30) with $\mathcal{G}$ replacing $\mathcal{A}_W F$ is well defined and converges superlinearly to x^* in a neighbourhood of x^*. On the other hand, it follows from (7) that $plen\mathcal{A}_W F(x) \subset plen\mathcal{G}(x)$ for all x in a neighborhood of x^*. Thus iteration (30) is well defined and converges superlinearly to x^* in the neighbourhood of x^*.

□

Remark 4. By Remarks 1-2 and Theorem 14, one can find a range of Newton's methods based on the strong continuous approximations to ∂F.

Acknowledgements. The author is grateful to Professor R. Fletcher and Dr. S. Leyffer for helpful discussions. He is also thankful to Professor S. M. Robinson for providing helpful references. Finally he would like to thank two anonymous referees for insightful comments which have improved the quality of this paper.

References

1. Clarke, F.H. (1983): Optimization and Nonsmooth Analysis. Wiley, New York
2. Chen, X., Qi, L. (1994): A parameterized Newton method and a quasi-Newton method for nonsmooth equations. Comput. Optim. Appl. **3**, 157–179
3. Demyanov, V.F. (1995): Fixed pointed theorem in nonsmooth analysis and its applications. Numer. Funct. Anal. Optim. **16**, 53–109
4. Dennis, J.E., More, J.J. (1977): Quasi-Newton methods, motivation and theory. SIAM Rev. **19**, 46–89
5. Dirkse, S.P., Ferris, M.C. (1994): "A pathsearch damped Newton method for computing general equibria". Technical Report 94-03. Department of Computer Sciences, University of Wisconsin, Madison, WI
6. Dirkse, S.P., Ferris, M.C. (1995): The PATH solver: A non-monotone stabilization scheme for mixed complementarity problems. Optim. Method Software **5**, 123–156
7. Ferris, M.C., Ralph, D. (1995): "Projected gradient methods for nonlinear complementarity problems via normal maps". In: Du, D.Z., Qi, L., Womersley, R.S., eds., Recent Advances in Nonsmooth Optimization, pp. 57–87. World Scientific, Singapore
8. Gabriel, S.A., Pang, J.S. (1994): "A trust region method for constrained nonsmooth equations". In: Hager, W.W., Hearn, D.W., Pardalos, P.M., eds., Large-Scale Optimization: State of Art, pp. 159–186. Kluwer Academic, Boston
9. Glover, B.M., Ralph, D. (1994): First order approximations to nonsmooth mappings with application to metric regularity. Numer. Funct. Anal. Optim. **15**, 599–620
10. Harker, P.T., Xiao, B. (1990): Newton's method for nonlinear complementarity problem: A B-differentiable equation approach. Math. Program. **48**, 339–357
11. Hiriart-Urruty, J.-B. (1982): Characterizations of the plenary hull of the generalized Jacobian matrix. Math. Program. Study **17**, 1–12
12. Ioffe, A.D. (1984): Approximate subdifferentials and applications I: The finite dimensional theory. Trans. Am. Math. Soc. **281**, 389–416
13. Kojima, M., Shindo, S. (1986): Extensions of Newton and quasi-Newton methods to systems of PC^1 equations. J. Oper. Res. Soc. Jpn. **29**, 352–374

14. Kummer, B. (1992): "Newton's method based on generalized derivatives for nonsmooth functions: Convergence analysis". In: Oettli, W., Pallaschke, D., eds., Advance in Optimization, pp. 171–194. Springer, Berlin
15. Ip, C.M., Kyparisis, J. (1992): Local convergence of quasi-Newton methods for B-differentiable equations. Math. Program. **58**, 71–89
16. Martinez, J.M., Qi, L. (1995): Inexact Newton methods for solving nonsmooth equations. J. Comput. Appl. Math. **60**, 127–145
17. Miffilin, R. (1977): Semismooth and semiconvex functions in constrained optimization. SIAM J. Control Optim. **15**, 957–972
18. Ortega, J.M., Rheinbodlt, W.C. (1970): Iterative Solution of Nonlinear Equations in Several Variables. Academic, New York
19. Pang, J.S. (1990): Newton's method for B-differentiable equations. Math. Oper. Res. **15**, 311–341
20. Pang, J.S. (1991): A B-differentiable equation-based, globally, and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems. Math. Program. **51**, 101–131
21. Pang, J.S. (1993): A degree-theoretic approach to parametric nonsmooth equations with motivated perturbed solution sets. Math. Program. **62**, 359–383
22. Pang, J.S., Qi, L. (1993): Nonsmooth equations: motivation and algorithms. SIAM J. Optim. **3**, 443–465
23. Qi, L. (1993): Convergence analysis of some algorithms for solving nonsmooth equations. Math. Oper. Res. **18**, 227–244
24. Qi, L. (1996): "C-Differential operators, C-Differentiability and generalized Newton methods". Applied Mathematics Report, 96-5. School of Mathematics, the University of New South Wales, Sydney, Australia
25. Qi, L., Sun, J. (1993): A nonsmooth version of Newton's method. Math. Program. **58**, 353–367
26. Ralph, D. (1990): "Rank-1 Support Functionals and the Generalized Jacobian, Piece-wise Linear Homeomorphisms". Ph.D Thesis, University of Wisconsin, Madison, WI
27. Ralph, D. (1994): Global convergence of damped Newton's method for nonsmooth equations, via the path search. Math. Oper. Res. **19**, 352–389
28. Ralph, D., Scholtes, S. (1995): "Sensitivity analysis and Newton's method for composite piece-wise smooth equations". Department of Mathematics, University of Melbourne, Melbourne, Australia
29. Robinson, S.M. (1987): Local structure of feasible sets in nonlinear programming, part III: stability and sensitivity. Math. Program. Study **30**, 45–66
30. Robinson, S.M. (1994): Newton's method for a class of nonsmooth functions. Set-Valued Anal. **2**, 291–305
31. Sun, J., Qi, L. (1992): An interior point algorithm of $\circ(\sqrt{m} \mid \ln\epsilon \mid)$ iterations for C^1-convex programming. Math. Program. **57**, 239–257
32. Sweetser, T.H. (1977): A minimal set-valued strong derivative for vector-valued Lipschitz functions. J. Optim. Theory Appl. **23**, 549–562
33. Xiao, B., Harker, P.T. (1994): A nonsmooth Newton method for variational inequalities, I: theory. Math. Program. **65**, 151–194
34. Xu, H. (1995): A pointed-based set-valued approximation and Newton-type methods. Technical Report, 160. Department of Mathematics and Computer Science, University of Dundee, Dundee, Scotland
35. Xu, H., Chang, X. (1997): Approximate Newton methods for nonsmooth equations. J. Optim. Theory Appl. **93**, 373–394
36. Xu, H., Glover, B.M. (1997): A new version of Newton's method for nonsmooth equations. J. Optim. Theory Appl. **93**, 395–415
37. Xu, H., Rubinov, A.M., Glover, B.M. (1996): "Continuous approximations to generalized Jacobian with applications to the nonsmooth least-squares minimization". Technical Report, 96-17. School of Information Technology and of Mathematical Sciences, University of Ballarat, Victoria, Australia
38. Xu, H. (1998): "Point-based set-valued, *C*-differential operator and application". Technical Report, 98-5. School of Information Technology and of Mathematical Sciences, University of Ballarat, Victoria, Australia