Approximate Newton Methods for Nonsmooth Equations[1]

H. Xu[2] and X.-W. Chang[3]

Communicated by D. Q. Mayne

Abstract. We develop general approximate Newton methods for solving Lipschitz continuous equations by replacing the iteration matrix with a consistently approximated Jacobian thereby reducing the computation in the generalized Newton method. Locally superlinear convergence results are presented under moderate assumptions. To construct a consistently approximated Jacobian, we introduce two main methods: the classic difference approximation method and the ϵ-generalized Jacobian method. The former can be applied to problems with specific structures while the latter is expected to work well for general problems. Numerical tests show the two methods are efficient. Finally, a norm-reducing technique for the global convergence of the generalized Newton method is briefly discussed.

Key words. Lipschitz continuous, semismooth, consistently approximated Jacobian, difference approximation, ϵ-generalized approximate Jacobian.

---

[1]The authors are grateful to the referee and D. Q. Mayne for their insightful comments and constructive suggestions. Thanks go to L. Qi and M. Kojima for providing some helpful references and S. Leyffer for many helpful discussions.

[2]Lecturer, Department of Mathematics, Ningbo University, P. R. China; and Visiting Scholar, Department of Mathematics and Computer Sciences, University of Dundee, Scotland.

[3]Graduate student, School of Computer Science, McGill University, Montreal, Canada.

1 Introduction

We study the following system of nonsmooth equations,

$$F(x) = 0, \tag{1}$$

where $F : R^n \to R^n$ is locally Lipschitz continuous.

Since the introduction of B-derivatives by Robinson (Ref. 1), a lot of efforts have been made to generalize Newton's method and quasi-Newton methods to the nondifferentiable case. Among them are Refs. 2-10, most of which are intended to solve some specific nondifferentiable equations such as those transformed from nonlinear complementarity problems and mathematical programming problems. Kojima and Shindo (Ref. 11) first discussed a Newton-like method for the systems of piece-wise continuously differentiable equations. Ip and Kyparisis (Ref. 12) extended the classic quasi-Newton methods especially the Broyden method to B-differentiable equations. Locally superlinear convergence results were obtained under the condition that $F(x)$ has Gateaux derivatives at a solution point of (1). However, the extension seems to be limited, since a bound on the deterioration of the updating matrix can not be maintained if F is not differentiable at a solution point.

Qi and Sun (Ref. 13) proposed a generalized Newton method for (1). They employed the following iteration:

$$x_{k+1} = x_k - V_k^{-1} F(x_k), \quad V_k \in \partial F(x_k) \tag{2}$$

where $\partial F(x_k)$ is the generalized Jacobian of F at x_k defined by Clarke (Ref. 14), and V_k is arbitrarily taken from $\partial F(x_k)$. The iterates produced by (2) were proved locally superlinearly convergent under mild conditions. However, difficulties may occur when (2) is applied to solving real-life problems. The obvious one is the calculation of V_k when F is not differentiable at x_k. A simple idea that we are familiar with is to replace V_k by the Jacobian of F at a point near x_k, but the computation of a Jacobian is also a formidable task even in the smooth case. Another difficulty is that iteration (2) may not work well if there is no significant reduction of $\|F(x)\|$ when the initial iterate is far from a solution point.

Pang and Qi (Ref. 15) developed a new kind of Gauss-Newton method for solving a certain class of nonsmooth equations. They also extended the classical superlinear convergence results of Dennis and More for smooth equations and those of Ip and Kyparisis for B-differentiable equations. There are also other new surveys on the topic, see Refs. 16-18, for instance.

In this paper, we develop general approximate Newton methods for (1) in order to avoid the complicated computation of V_k in (2). In Section 2 we establish a general approximate Newton iterative scheme by introducing a concept, namely "consistently approximated Jacobian " (CAJ in brief). Local convergence results are presented under some mild conditions. In Section 3 we propose two practical methods: the classic difference approximation method and ϵ-generalized Jacobian method. The former can apparently be applied to a few problems with specific structures, while the latter, ϵ-generalized Jacobian method, is expected to work well for more general problems. In Section 4 we introduce very briefly a norm reducing technique to force the generalized Newton algorithms to converge when the initial iterate point is far from a solution point of (1). Finally we give numerical experiments in Section 5.

2 General Approximate Newton Iteration

2.1 Notation

Throughout this paper we use the following notation. R^n denotes the vector space of n-tuples with 2-norm $\|\cdot\|$ and $L(R^n)$ the matrix space of $n \times n$ real matrices with the induced norm $\|\cdot\|$. $S(x, \delta)$ denotes the open ball in R^n with center x and radius δ, and B an open unit ball in $L(R^n)$. The closure of a set D is specified by $\bar{D}$. If $\alpha(h)$ is a vector-valued function (in this context, we simply call it a function for brevity) or a matrix function of $h \in R^n$, we use $\alpha(h) = o(\|h\|)$ to denote the case that $\|\alpha(h)\|/\|h\| \to 0$, as $\|h\| \to 0$, and $\alpha(h) = O(\|h\|)$ if there exist constants $C_1, C_2 > 0$ such that $C_1\|h\| \leq \alpha(h) \leq C_2\|h\|$, as $\|h\| \to 0$. The set of positive integers $[1, \cdots, n]$ is denoted by $\bar{n}$. The set of points of R^n at which F is differentiable is denoted by D_F.

2.2 Basics

We assume throughout this paper that F is locally Lipschitz continuous in R^n in the sense that for every x, there exist $L > 0$, and $\delta > 0$, such that

$$\|F(y) - F(z)\| \leq L\|y - z\|$$

for all y, $z \in S(x, \delta)$. Here L is called the Lipschitz constant of F at x.

According to the Rademacher theorem, F is differentiable almost everywhere in R^n. The generalized Jacobian of F at x is defined by Clarke in Ref. 14 as follows:

$$\partial F(x) = \text{conv}\,[\lim_{x_i \to x, x_i \in D_F} \nabla F(x_i)].$$

Here, and later on, $\nabla F(x)$ denotes the Jacobian of F at $x \in D_F$, and conv denotes the convex hull.

Proposition 2.1. (Proposition 2.6.2, Ref. 14) $\partial F(x)$ is compact and upper semi-continuous in the sense that for every $\epsilon > 0$ there exists a $\delta > 0$ such that

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

for all $y \in S(x, \delta)$.

For the sake of convenience in the consequent discussion we summarize some preliminary concepts and results given by Qi and Sun in Ref. 13.

F is said to be semismooth at x if for every $h \in R^n$

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

exists.

Lemma 2.1. Suppose that F is semismooth at x. Then for every $h \in R^n$

$$\lim_{h' \to h, t \to 0} [F(x + th') - F(x)]/t = \lim_{V \in \partial F(x+th'), h' \to h, t \to 0} Vh' = F'(x, h).$$

Particularly, when $h' \equiv h$,

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

and

$$F(x + h) - F(x) - F'(x, h) = o(\|h\|). \qquad (3)$$

Proof. It follows directly from Proposition 2.1 and (2.17) of Ref. 13 and the definition of semismoothness. $\square$

If there exists $0 \le p < 1$ such that for any $V \in \partial F(x+h)$, $h \to 0$,

$$Vh - F'(x,h) = O(\|h\|^{1+p}),$$

then F is said to be p-order semismooth at x. In Ref. 13 Qi and Sun proved that if F is semismooth at x, then

$$F(x+h) - F(x) - F'(x,h) = O(\|h\|^{1+p}).$$

Lemma 2.2. (Lemma 2.2, Ref. 13) Suppose that $F'(x,h)$ exists at x for any h. Then for every h, there exists a $V \in \partial F(x)$ such that

$$F'(x,h) = Vh.$$

Lemma 2.3. (Theorem 2.3, Ref. 13) The following statements are equivalent:

(a) F is semismooth at x;

(b) For any $V \in \partial F(x+h)$, $h \to 0$, $\quad Vh - F'(x,h) = o(\|h\|)$;

(c) $\lim\limits_{h \to 0} [F'(x+h,h) - F'(x,h)]/\|h\| = 0$.

$\partial F(x)$ is said to be nonsingular if all $V \in \partial F(x)$ are nonsingular.

Lemma 2.4. (Proposition 3.1, Ref. 13) Suppose that $\partial F(x)$ is nonsingular. Then there exist $\delta > 0$, $C_1 > 0$, such that for each $y \in S(x,\delta)$, $\partial F(y)$ is nonsingular and

$$\|V^{-1}\| \le C_1, \quad \text{for all } \; V \in \partial F(y),\ y \in S(x,\delta).$$

2.3 Approximate Newton Method

We now introduce an approximate Newton iteration. Let x_k be the current iterate, and let x_{k+1} be the next iterate. Then x_{k+1} is computed by

$$x_{k+1} = x_k - J(x_k, s_k)^{-1} F(x_k), \tag{4}$$

where $J(x_k, s_k) \in L(R^n)$ is an approximation of some $V_k \in \partial F(x_k)$, and s_k is an $l-$dimensional parameter vector.

The iteration (4) is a generalization of approximate Newton methods in the smooth case, which were extensively discussed by Ortega and Rheinboldt in Ref. 19. The main difficulty here is the construction of a $J(x, s)$. For the sake of convenience, we present a strict theoretical definition of $J(x, s)$.

Let $N(x, \delta)$ denote the closed ball with center x and radius δ in R^l in order to distinguish it from $S(x, \delta)$.

Definition 2.1. Let $J : D \times D_l \subset R^n \times R^l \to L(R^n)$. If $0 \in R^l$ is a limiting point of D_l, and

$$\lim_{s \to 0} \operatorname{dist}\,[J(x, s); \partial F(x)] = 0, \tag{5}$$

for all $x \in D$, then $J(x, s)$ is called a *consistently approximated Jacobian* of F in D (We call it CAJ for brevity). Here

$$\operatorname{dist}\,[J(x, s); \partial F(x)] \equiv \min_{V \in \partial F(x)} \|J(x, s) - V\|. \tag{6}$$

Lemma 2.5. Let $x^* \in R^n$. Suppose that F is semismooth at x^* and that $J(x, s)$ is a CAJ of $F(x)$ in a neighbourhood of x^*. Then

$$\|F'(x^* + h, h) - J(x^* + h, s)h\| \leq w(h, s)\|h\|, \tag{7}$$

where

$$\lim_{h\to 0; s\to 0} w(h,s) = 0. \tag{8}$$

Proof. Let $\delta > 0$ be sufficiently small so that $J(x,s)$ is a CAJ of F in $S(x^*,\delta)$. By Definition 2.1,

$$\lim_{s\to 0} \text{ dist } [J(x^*+h,s); \partial F(x^*+h)] = 0$$

uniformly with respect to $\|h\| \le \delta$. Note that $\partial F(x^*+h)$ is closed. Then for each s, there exists a $V(x^*+h,s) \in \partial F(x^*+h)$ such that

$$\|V(x^*+h,s) - J(x^*+h,s)\| = \text{ dist } [J(x^*+h,s); \partial F(x^*+h)] \to 0 \tag{9}$$

uniformly with respect to $\|h\| \le \delta$, as $s \to 0$. Since F is semismooth at x^*, from Lemma 2.3 (b),

$$V(x^*+h,s)h - F'(x^*,h) = o(\|h\|) \tag{10}$$

uniformly with respect to s. Using Lemma 2.3 (c), we get

$$F'(x^*+h,h) - F'(x^*,h) = o(\|h\|). \tag{11}$$

Write

$$\begin{aligned} w(h;s) \quad &\equiv \quad \|V(x^*+h) - J(x^*+h,s)\| + [\|F'(x^*+h,h) - F'(x^*,h)\| \\ &\qquad + \|V(x^*+h,s)h - F'(x^*,h)\|]/\|h\|. \end{aligned}$$

Then we see (8) holds from (9)–(11), and

$$\begin{aligned} & \|F'(x^*+h,h) - J(x^*+h,s)h\| \\ \le \quad & \|F'(x^*+h,h) - F'(x^*,h)\| + \|V(x^*+h,s)h - F'(x^*,h)\| \\ & + \|V(x^*+h,s) - J(x^*+h,s)\|\|h\| \\ = \quad & w(h,s)\|h\|. \end{aligned}$$

The proof is complete. $\square$

Now we can state our main results.

Theorem 2.1. Let $x^* \in D$ be a solution of (1). Suppose F is semismooth at x^*, and $\partial F(x^*)$ is nonsingular. Let $J : D \times D_l \subset R^n \times R^l \to L(R^n)$ is a CAJ of F in D, and assume $0 \in D_l$. Then there exist $\delta > 0$ and $\gamma > 0$ such that $J(x, s)$ is nonsingular for $x \in S(x^*, \delta)$, $s \in N(0, \gamma)$. Furthermore, the function

$$G(x, s) \equiv x - J(x, s)^{-1} F(x) \tag{12}$$

is contractive in the sense that

$$\|x^* - G(x, s)\| \leq u(x, s)\|x - x^*\|, \tag{13}$$

for all $x \in D \bigcap S(x^*, \delta)$; $s \in D_l \bigcap N(0, \gamma)$, where

$$\lim_{x \to x^*; s \to 0} u(x, s) = 0. \tag{14}$$

Proof. First, we show that there exist $\delta > 0$ and $\gamma > 0$, such that $J(x, s)$ is nonsingular for any $x \in D \bigcap S(x^*, \delta)$ and any $s \in D_l \bigcap N(0, \gamma)$. It follows from Definition 2.1 that for each $\epsilon > 0$, there exists a constant $\gamma > 0$ such that

$$\text{dist}\,[J(x, s); \partial F(x)] \leq \epsilon \quad \text{for all } x \in D,\ s \in D_l \bigcap N(0, \gamma). \tag{15}$$

Since $\partial F(x^*)$ is nonsingular, by Lemma 2.4, there exist a constant $C_1 > 0$ and a $\delta > 0$, such that $\partial F(x)$ is nonsingular and $\|V(x)^{-1}\| \leq C_1$ for all $x \in D \bigcap S(x^*, \delta)$. Let $\hat{V}(x) \equiv \arg\min_{V \in \partial F(x)} \|J(x, s) - V(x)\|$. Then from (15) and (6) we have

$$\|J(x, s) - \hat{V}(x)\| \leq \epsilon. \tag{16}$$

Suppose without loss of generality that $\epsilon C_1 < 1$. Then by the Banach perturbation theorem and (16), $J(x,s)$ is nonsingular and satisfies

$$\|J(x,s)^{-1}\| \leq \|\hat{V}(x)^{-1}\|/(1-\|\hat{V}(x)^{-1}\|\|J(x,s)-\hat{V}(x)\|) \leq C_1/(1-\epsilon C_1). \quad (17)$$

Now we prove (13). In fact

$$\begin{aligned}
& \|x^* - G(x,s)\| \\
= \;& \|J(x,s)^{-1}[F(x) - J(x,s)(x-x^*)]\| \\
\leq \;& \|J(x,s)^{-1}\| \, [\, \|F(x) - F(x^*) - F'(x^*, x-x^*)\| \\
& + \|F'(x, x-x^*) - F'(x^*, x-x^*)\| + \|F'(x, x-x^*) - J(x,s)(x-x^*)\| \,] \\
\leq \;& C_1/(1-\epsilon C_1)[o(\|x-x^*\|) + o(\|x-x^*\|) + w(x-x^*, s)\|x-x^*\|] \\
\equiv \;& u(x,s)O(\|x-x^*\|) \quad \text{(say)}.
\end{aligned}$$

The last inequality is due to Lemma 2.3, Lemma 2.5, (3) and (17). Clearly (14) holds. The proof is complete. □

Corollary 2.1. Suppose that the assumptions in Theorem 2.1 are satisfied. Let $G(x,s)$ be defined by (12), and let $g : R^n \to R^s$, be a continuous function with $g(x^*) = 0$. If $s = g(x)$, then there exists a $\delta > 0$ such that $G(x, g(x))$ is contractive in $S(x^*, \delta)$. Thereby, the sequence produced by (4) converges to x^* superlinearly for a sufficiently good starting point.

Proof. From (14), it follows that for any given $\epsilon > 0$, there exist $\delta_1 > 0$ and $\sigma > 0$ such that $u(x,s) \leq \epsilon$ for all $x \in S(x^*, \delta_1)$, $\|s\| \leq \sigma$. Since $s = g(x)$ and g is continuous at x^* with $g(x^*) = 0$, there exists a $\delta < \delta_1$ such that $\|s\| = \|g(x)\| \leq \sigma$ for $x \in S(x^*, \delta)$, thereby $u(x, g(x)) \leq \epsilon$ for all $x \in S(x^*, \delta)$. The conclusion is obvious as ϵ can be arbitrarily small. □

3 Two Practical Approximation Methods

In Section 2, we established a unified framework of superlinearly convergent approximate Newton methods based on CAJ. The results may be regarded as theoretical since there are no practical methods presented for the construction of a CAJ. In this section, we discuss how to construct a CAJ practically. Obviously, the existence of a CAJ depends on the local property of $F(x)$. It seems unlikely to provide a unified approach for general problems. In what follows, we propose two methods: the finite difference approximation method and the ϵ-generalized Jacobian approximation method.

3.1 Finite Difference Approximation

We first introduce some useful notation. If a_j, $j \in \bar{n}$ are n vectors of R^n, we will denote by $[a_1, \cdots, a_n]$ the matrix whose j-th column is a_j. Similarly if $A_j, j \in \bar{n}$ are subsets of R^n, then $[A_1, \cdots, A_n]$ will be $[\,[a_1, \cdots, a_n] : a_j \in A_j$ for all $j \in \bar{n}\,]$.

Now we suppose that F is semismooth in R^n. By Lemma 2.1, for each $x \in R^n$

$$F'(x, e_j) \in \partial F(x) e_j, \quad \text{for } j \in \bar{n},$$

where e_j is the j-th unit vector. On the other hand,

$$\partial F(x) e_j e_j^T = [0, \cdots, \partial_j F, \cdots, 0], \quad \text{for } j \in \bar{n},$$

where $\partial_j F = \text{conv}\,[\lim_{x_i \in D_F; x_i \to x} \nabla_j F(x_i)]$, and $\partial_j F$ denotes the derivatives of F with respect to the j-th component of the variable x at a point $x_i \in D_F$.

Hence,

$$F'(x, e_j)e_j^T \in [0, \cdots, \partial_j F, \cdots, 0], \quad \text{for } j \in \bar{n},$$

and consequently

$$\sum_{j\in\bar{n}} F'(x, e_j)e_j^T \in [\partial_1 F, \cdots, \partial_n F].$$

In general,

$$[\partial_1 F, \cdots, \partial_n F] \neq \partial F,$$

but equality holds at least in two particular cases: first, when all functions f_j except possibly one , say f_1, are strictly differentiable at x, in such a case, $\partial F = [\partial f_1, \nabla f_2, \cdots, \nabla f_n]^T$; another equality case is when the functions are nondifferentiable but only with respect to *nonrelated* variables.

Example 3.1. Let $F : R^2 \to R^2$ be defined by $F(x) = (|x_1|+x_2, x_1+|x_2|)^T$. Then,

$$\partial F = \left\{ \begin{bmatrix} \alpha & 1 \\ 1 & \beta \end{bmatrix}, \alpha \in [-1, 1], \beta \in [-1, 1] \right\} = [\partial F_1, \partial_2 F].$$

So far, we are able to reach the following conclusions:

Lemma 3.1. Suppose F is a function whose components are either nondifferentiable only with respect to '*nonrelated*' variables or strictly differentiable except for one, and F is semismooth in R^n. Then

$$J(x, s) \equiv \sum_{j\in\bar{n}} ([F(x + se_j) - F(x)]/s)e_j^T \in \partial F(x) + o(s). \tag{18}$$

Proof. Since $F(x)$ is semismooth, by Lemma 2.1,

$$[F(x + se_j) - F(x)]/s - F'(x, e_j) = o(s), \quad \text{for } j \in \bar{n}.$$

Hence

$$
\begin{aligned}
J(x,s) &= \Big(\sum_{j\in\bar{n}}[F(x+se_j)-F(x)]/s\Big)e_j^T \\
&= \sum_{j\in\bar{n}}F'(x,e_j)e_j^T + o(s) \in [\partial_1 F,\cdots,\partial_n F] + o(s).
\end{aligned}
$$

By assumption, we get

$$
[\partial_1 F,\cdots,\partial_n F] = \partial F(x).
$$

This completes the proof. □

Theorem 3.1. Let x^* be a solution of (1). Assume that F is semismooth in a neighbourhood of x^*, and that $\partial F(x^*)$ is nonsingular. Suppose the assumptions of Lemma 3.1 are satisfied, and $J(x,s)$ is defined by (19), which holds uniformly with respect to x in a neighbourhood of x^*. Then for $s = O(\|F(x)\|)$, there exists a $\delta > 0$, such that if starting point $x_0 \in S(x^*,\delta)$, the sequence $\{x_k\}$ generated by (4) is well defined and converges to x^* superlinearly.

Proof. It follows from Lemma 3.1 that $J(x,s)$ is a CAJ of F in some neighbourhood of x^* under the assumption that (18) holds uniformly with respect to x in a neighbourhood of x^*. The rest follows directly from Corollary 2.1. □

Remark 3.1. Theorem 3.1 is a nonsmooth version of the classic difference approximate Newton method. It seems that the nonsmooth finite approximation can only fit in a small class of problems. Therefore we are led to develop other techniques which are likely to deal with more general problems.

3.2 ϵ-Sub-Jacobian Approximation

In nonsmooth optimization, when the steepest subgradient method failed to solve Wolfe's counter example, people realized that the subdifferential of a nonsmooth function contains too less information for finding a convergent search direction. A larger set called ϵ-subdifferential was then introduced and the existing difficulties were consequently overcome. In this context, using the same method as for defining the ϵ-subdifferential, we try to define an ϵ-generalized Jacobian for a vector-valued function. However, rather than out of the convergence consideration, the definition here is intended to provide more practical substitutes of the generalized Jacobian which are usually difficult to compute.

Definition 3.2. Let $F(x)$ be locally Lipschitz continuous. Then we say $\partial_\epsilon F(x)$ is the ϵ-generalized Jacobian of F at x if

$$\partial_\epsilon F(x) = \text{conv } [\partial F(x') : x' \in \overline{S(x;\epsilon)}]. \tag{19}$$

The ϵ- generalized Jacobian has many interesting properties.

Proposition 3.1. Let $\partial_\epsilon F(x)$ be defined by (19). Then

(a) for every $x \in R^n$, $\partial_0 F(x) = \partial F(x)$;

(b) for every $x \in R^n$, if $\epsilon < \epsilon'$, then $\partial_\epsilon F(x) \subset \partial_{\epsilon'} F(x)$;

(c) for any $\epsilon \geq 0$, $\partial_\epsilon F(x)$ is convex and bounded on a bounded set;

(d) for every $\mu > 0$, there exists a $\delta > 0$ such that $\partial_\epsilon F(y) \subset \partial F(x) + \mu B$, for all $y \in S(x,\delta)$, $\epsilon < \delta$.

Proof. We only prove (d). By Proposition 2.1, for every $\mu > 0$, there exists a $\delta > 0$ such that $\partial F(x') \subset \partial F(x) + \mu B$, for all $x' \in S(x,\delta)$. From

Definition 3.2, we have, for $\epsilon < \delta/2$, $y \in S(x, \delta/2)$,

$$\begin{aligned} \partial_\epsilon F(y) &= \operatorname{conv}[\partial F(x') : x' \in \overline{S(y,\epsilon)}] \\ &\subset \operatorname{conv}[\partial F(x') : x' \in S(x,\delta)] \subset \partial F(x) + \mu B. \end{aligned}$$

□

Proposition 3.2. For each $V_\epsilon \in \partial_\epsilon F(x)$,

$$\lim_{\epsilon \to 0} V_\epsilon \in \partial F(x). \tag{20}$$

Proof. Note that for $V_\epsilon \in \partial_\epsilon F(x)$, $\epsilon_0 > 0$, $[V_\epsilon]$ is bounded for $\epsilon \in [0, \epsilon_0]$. Assume that there is a subsequence $\{V_{\epsilon_j}\}$, $V_{\epsilon_j} \to V' \notin \partial F(x)$ as $\epsilon_j \to 0$ and seek a contradiction.

Since

$$\partial F(x) = \bigcap_{0 \le \epsilon \le \epsilon_0} \partial_\epsilon F(x),$$

we have

$$V' \notin \bigcap_{0 \le \epsilon \le \epsilon_0} \partial_\epsilon F(x),$$

or equivalently

$$V' \in (\bigcap_{0 \le \epsilon \le \epsilon_0} \partial_\epsilon F(x))^c = \bigcup_{0 \le \epsilon \le \epsilon_0} (\partial_\epsilon F(x))^c,$$

where A^c denotes the complementary set of A. Thus there exists $0 \le \epsilon' \le \epsilon_0$ such that

$$V' \in (\partial_{\epsilon'} F(x))^c. \tag{21}$$

On the other hand, let j be so large that $\epsilon_j < \epsilon'$. Then by Proposition 3.1 (b),

$$V_{\epsilon_j} \in \partial_{\epsilon_j} F(x) \subset \partial_{\epsilon'} F(x).$$

Since $\partial_{\epsilon'} F(x)$ is compact, we have

$$V' \in \partial_{\epsilon'} F(x),$$

which leads to a contradiction to (21). $\square$

In general, it is very restrictive if we require that (20) hold uniformly with respect to x in some open set D. Therefore V_ϵ is not necessarily a CAJ in general cases. This leads us to consider a weaker condition for the convergence of the approximate Newton methods.

Lemma 3.2. Suppose F is semismooth at x^*, and $E : R_+ \to R_+$ is a real function with $E(0) = 0$ and $\lim_{t \to 0} E(t)/t = 0$. Then for every $U \in \partial_{E(\|x-x^*\|)} F(x)$, there exists a $V \in \partial F(x)$ such that

$$\lim_{x \to x^*} (U - V)(x - x^*)/\|x - x^*\| = 0. \tag{22}$$

Proof. Let $t \equiv \|x - x^*\|$ and let $h \equiv (x - x^*)/\|x - x^*\|$. Then, for each $U \in \partial_{E(\|x-x^*\|)} F(x)$, we have by Definition 3.2 and Carathedory's theorem that

$$U = \sum_{i=1}^{l(x)} \alpha_i(x) U_i(x),$$

where $l(x)$ is an upper bounded positive number, $\alpha_i(x)$, $i = 1, \ldots, l(x)$, are nonnegative scalars with $\sum_{i=1}^{l(x)} \alpha_i(x) = 1$, and $U_i(x) \in \partial F(x^* + th + \theta_i(x) E(t) s_i(x))$ with $s_i(x) \in R^n$, $\|s_i(x)\| = 1$, and $\theta_i(x) \in [0, 1]$ for $i = 1, \ldots, l(x)$. Let $h_i(t) \equiv h + \theta_i(x) E(t)/t$. Suppose without loss of generality that $h \to \hat{h}$ as $t \to 0$. Since F is semismooth at x^*, it follows from Lemma 2.1 that for every $V \in \partial F(x)$

$$\lim_{t \to 0} U_i(x) h_i(t) = \lim_{t \to 0} U_i(x) h = \lim_{t \to 0} V h = F'(x^*, \hat{h}),$$

which implies (22). $\square$

Remark 3.2. The conclusion of Lemma 3.2 still holds when $E(t)$ is replaced by $kE(t)$ for some positive constant k.

Theorem 3.2. Let x^* be a solution point of (1). Suppose that F is semismooth at x^*, and that $\partial F(x^*)$ is nonsingular. Let $E : R_+ \to R_+$ be a real function with $E(0) = 0$ and $\lim_{t\to 0} E(t)/t = 0$. Then the iteration

$$x_{k+1} = x_k - U_k^{-1}F(x_k), \quad U_k \in \partial_{E(\|x_k - x^*\|)}F(x_k), \tag{23}$$

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

Proof. Since $\partial F(x^*)$ is nonsingular, by Proposition 3.1 (d), there exists a $\delta > 0$ such that $\partial_\epsilon F(x)$ is also nonsingular and $\|U^{-1}\|$ is bounded for every $U \in \partial_\epsilon F(x)$ when $\epsilon < \delta$, $x \in S(x^*, \delta)$. Hence

$$\begin{aligned}
\|x_{k+1} - x^*\| &\leq \|U_k^{-1}\| \, [\, \|F(x_k) - F(x^*) - F'(x^*, x_k - x^*)\| \\
&\quad + \|F'(x_k, x - x^*) - F'(x^*, x - x^*)\| \\
&\quad + \|F'(x_k, x - x^*) - V_k(x_k - x^*)\| + \|(U_k - V_k)(x_k - x^*)\| \,] \\
&= o(\|x_k - x^*\|)
\end{aligned}$$

The last equality follows from Lemma 2.1, Lemma 2.3, and Lemma 3.2. The proof is complete. $\square$

The methods for the construction of $\partial_\epsilon F(x)$ vary from problem to problem. For some problems, a detailed knowledge of $\partial_\epsilon F(x)$ is needed, while for some others maybe one element of $\partial_\epsilon F(x)$ is enough. In what follows, we present an integral method which can be regarded as a generalization of results of Ref. 20.

Let $\|\cdot\|_\infty$ denote the infinity norm of vectors of R^n, set $P_\epsilon(x) \equiv [\, y \in R^n : \|y-x\|_\infty \leq \epsilon \,]$. Define

$$F_\epsilon(x) \equiv (f_\epsilon^1(x), \cdots, f_\epsilon^n(x))^T, \tag{24}$$

where $f_\epsilon^i(x) \equiv \sigma(\epsilon) \int_{P_\epsilon(x)} f_i(y)dy$ with $\sigma(\epsilon) = 1/(2\epsilon)^n$.

Then we have

Proposition 3.3. $F_\epsilon(x)$ is continuously differentiable with respect to x, and

$$\nabla_j f_\epsilon^i(x) = \frac{1}{(2\epsilon)^n}[\int_{D^j_{\epsilon+}(x)} f_i(y)dy - \int_{D^j_{\epsilon-}(x)} f_i(y)dy], \quad \text{for } i,j = 1, \cdots, n, \tag{25}$$

where

$$\begin{aligned} D^j_{\epsilon+}(x) &= [\, y \in R^n : \|y-x\|_\infty = \epsilon, (y)_j = (x)_j + \epsilon \,] \\ &= [\, x + s : (s)_j = \epsilon, |(s)_i| \leq \epsilon, i \neq j, 1 \leq i \leq n \,], \\ D^j_{\epsilon-}(x) &= [\, y \in R^n : \|y-x\|_\infty = \epsilon, (y)_j = (x)_j - \epsilon \,] \\ &= [\, x - s : (s)_j = \epsilon, |(s)_j| \leq \epsilon, i \neq j, 1 \leq i \leq n \,], \end{aligned}$$

$(x)_j$ denotes the j-th component of vector x.

Proof. The proof can be obtained by a simple calculation. □

Lemma 3.3. Let $F_\epsilon(x)$ be defined by (24). Then for every $h \in R^n$ with $\|h\| = 1$,

$$\nabla F_\epsilon(x)h \in \partial_{(\sqrt{n}+1)\epsilon} F(x)h, \tag{26}$$

Proof. Given $h \in R^n$ with $\|h\| = 1$. Since $F_\epsilon(x)$ is differentiable,

$$\begin{aligned} \nabla F_\epsilon(x)h &= \lim_{t\to 0}[F_\epsilon(x+th) - F_\epsilon(x)]/t \\ &= \lim_{t\to 0}\sigma(\epsilon)\int_{P_\epsilon(0)}[(F_\epsilon(x+y+th) - F_\epsilon(x+y))/t]dy \\ &= \lim_{t\to 0}\sigma(\epsilon)\int_{P_\epsilon(0)}\int_0^1 \nabla F(x+s+t\theta h)hd\theta ds. \end{aligned}$$

Note that

$$\|s+\theta th\| \le \|s\| + t \le \sqrt{n}\|s\|_\infty + t \le \sqrt{n}\epsilon + t.$$

Hence, by letting $t < \epsilon$, we have

$$\nabla F(x+s+th) \in \partial_{(\sqrt{n}+1)\epsilon}F(x),$$

and consequently

$$\nabla F(x+s+th)h \in \partial_{(\sqrt{n}+1)\epsilon}F(x)h,$$

which implies (26). $\square$

Lemma 3.4. Let $\nabla F_\epsilon(x)$ be given by (25). Suppose $\partial F(x^*)$ is nonsingular. Then there exist a constant $C_2 > 0$ and a $\delta > 0$, such that $\nabla F_\epsilon(x)$ is nonsingular and $\|\nabla F_\epsilon(x)^{-1}\| \le C_2$ for all $x \in S(x^*, \delta)$, $\epsilon < \delta$.

Proof. By Proposition 3.1 (d), for every $\mu > 0$, there exists a $\delta > 0$ such that

$$\partial_{(\sqrt{n}+1)\epsilon}F(x) \subset \partial F(x^*) + \mu B, \quad \forall x \in S(x^*, \delta),\ \epsilon < \delta.$$

Since $\partial F(x^*)$ is nonsingular, for μ sufficiently small, every matrix of $\partial F(x^*) + \mu\bar{B}$ is nonsingular. Now let $\mathcal{A} \equiv \partial F(x^*) + \mu\bar{B}$. Then

$$\sigma \equiv \min_{A\in\mathcal{A}}\min_{\|h\|=1}\|Ah\| > 0.$$

On the other hand, it follows from Lemma 3.3 that

$$\nabla F_\epsilon(x)h \in \partial_{(\sqrt{n}+1)\epsilon} F(x)h, \quad \forall h \in R^n.$$

Hence

$$\nabla F_\epsilon(x)h \in \mathcal{A}h,$$

and

$$\min_{\|h\|=1} \|\nabla F_\epsilon(x)h\| \geq \min_{A\in\mathcal{A}} \min_{\|h\|=1} \|Ah\| = \sigma > 0,$$

which implies $\nabla F_\epsilon(x)$ is nonsingular. Letting $C_2 = 1/\sigma$, we have

$$\|\nabla F_\epsilon(x)^{-1}\| = 1/\min_{\|h\|=1} \|\nabla F_\epsilon(x)h\| \leq 1/\sigma = C_2, \quad \forall x \in S(x^*, \delta),\ \epsilon < \delta,$$

which completes the proof. $\square$

Theorem 3.3. Let $\nabla F_\epsilon(x)$ be given by (25). Then Theorem 3.2 holds for $U_k = \nabla F_{E(\|x-x^*\|)}(x_k)$ in (23).

Proof. Since $\partial F(x^*)$ is nonsingular, it follows from Lemma 3.4 that there exists a $\delta > 0$ such that $\nabla F_\epsilon(x)$ is nonsingular and $\|\nabla F_\epsilon(x)^{-1}\|$ is bounded for all $x \in S(x^*, \delta)$, $\epsilon < \delta$.

On other hand, for all $h \in R^n$ with $\|h\| = 1$, it follows from Lemma 3.3 that $\nabla F_\epsilon(x)h \in \partial_{(\sqrt{n}+1)\epsilon} F(x)h$, $\forall h \in R^n$. Since $\partial_{(\sqrt{n}+1)\epsilon} F(x)$ is closed, for every h there exists a $U_h \in \partial_{(\sqrt{n}+1)\epsilon} F(x)$ such that $\nabla F_\epsilon(x)h = U_h h$. By Lemma 3.2 and Remark 3.2, U_h satisfies (22) for $\epsilon = E(\|x - x^*\|)$. Therefore, by simulating the proof of Theorem 3.2, we can easily get the conclusion. $\square$

4 Norm Reducing Technique

In this section, we turn to discuss briefly the techniques which are likely to be combined with local convergence methods when a starting point is far from a solution point of (1). It is well known that solving the nonlinear simultaneous equations (1) is equivalent to finding the global minimizer of

$$f(x) = \frac{1}{2} F(x)^T F(x).$$

Suppose that $F(x)$ is locally Lipschitz continuous. Then $f(x)$ is also locally Lipschitz continuous, and

$$\partial f(x) = \partial F(x)^T F(x). \tag{27}$$

A proof of (27) was given by Clarke (Theorem 2.6.6, Ref. 14).

General methods for minimizing $f(x)$ may be summarized as:

$$\begin{aligned} x_+ &= x_c + \alpha d_c, \\ d_c &= -\arg \min_{v \in \partial f(x_c)} \|v\|, \end{aligned}$$

where x_c and x_+ denote the current and next approximate root of (1) respectively, α is the stepsize, and d_c is the search direction.

Substituting (27) into (28), we have

$$d_c = -\arg \min_{V \in \partial F(x_c)} \| V F(x_c) \| . \tag{28}$$

By Caratheodory theorem, there exist an integer m, and $V_i \in \partial F(x_c)$, $i \in \bar{m}$, such that every $V_c \in \partial F(x_c)$ may be expressed as

$$V_c = \sum_{i \in \bar{m}} \alpha_i V_i$$

with

$$\sum_{i \in \bar{m}} \alpha_i = 1; \quad \alpha_i \geq 0; \quad V_i \in \partial F(x_c).$$

As a consequence, (28) is equivalent to

$$\min \| \sum_{i \in \bar{m}} \alpha_i V_i^T F(x_c) \|, \qquad (29)$$
$$\text{s.t.} \quad \sum_{i \in \bar{m}} \alpha_i = 1;\ \alpha_i \geq 0;\ V_i \in \partial F(x_c), \quad \text{for } i \in \bar{m},$$

and

$$d_c = -\sum_{i \in \bar{m}} \bar{\alpha}_i V_i^T F(x_c).$$

where $(\bar{\alpha}_1, \cdots, \bar{\alpha}_m)$ is the solution of (29).

Therefore, if we already know $V_i \in \partial F(x_c)$, the search direction may be computed through (29). However, in practice, it is unrealistic to compute all $V_i \in \partial F(x_c)$. One possible substitute is to use the former iteration matrices computed at the previous steps. Let $V_i \in \partial F(x_i)$, and $I = [\, i : \ \|x_c - x_i \| \leq \epsilon]$. Then we may compute an approximate search direction by solving

$$\min \| \sum_{i \in I} \alpha_i V_i^T F(x_c) \|,$$
$$\text{s.t.} \quad \sum_{i \in I} \alpha_i = 1;\ \alpha_i \geq 0;\ V_i \in \partial F(x_i), \quad \text{for } i \in I$$

instead of (29). This method is called bundle-like method.

An alternative suggested by R. Fletcher is to use the L_1-norm, i.e., to convert (1) into the following global minimization problem:

$$\min \phi(x) = \sum_{i \in \bar{n}} |f_i(x)|.$$

Suppose that $f_i(x), i \in \bar{m}$, are regular (Ref. 14). Then for every $h \in R^n$,

$$\phi'(x,h) = \sum_{i \in \alpha(x)} f_i'(x,h) - \sum_{i \in \beta(x)} f_i'(x,h) + \sum_{i \in \gamma(x)} |f_i'(x,h)|,$$

where $\alpha(x) = [\, i \in \bar{n} : f_i(x) > 0\,]$, $\beta(x) = [\, i \in \bar{n} : f_i(x) < 0\,]$ and $\gamma(x) = [\, i \in \bar{n} : f_i(x) = 0\,]$.

We may use a trust region method to find a descent direction by solving

$$\begin{aligned}
&\min \phi(x) + \phi'(x,h) + \frac{1}{2}h^T Bh, \\
&\text{s.t} \quad \|h\| \le \Delta.
\end{aligned}$$

We omit the further discussion since it is not our main topic in this paper.

5 Numerical Experiments

We implemented the finite difference approximation method based on (4) and (18), and ϵ-generalized Jacobian methods based on (23) for some typical examples. The computations were performed in Matlab. The results show that our methods are efficient.

Example 5.1. Consider the following nonsmooth equations:

$$F(x) = \begin{pmatrix} |x_1| + (x_2-1)^2 - 1 \\ (x_1-1)^2 + |x_2| - 1 \end{pmatrix} = 0.$$

This problem has two solutions: $x^* = (1,1)^T$, $x^{**} = (0,0)^T$. F is differentiable at x^* but nondifferentiable at x^{**}, and it satisfies the conditions of Lemma 3.1. We used finite difference approximation method to solve this problem. The numerical results are shown in Table 1, in which s is the parameter in (18), NIT denotes the number of iterations needed to reach the specified precision, "Solution" is the exact solution of the equations, to which the iteration sequence converges, and $F(\hat{x})$ is the function evaluated at the computed solution $\hat{x}$.

Table 1. Results for Example 5.1

Initial point	s	NIT	Solution	$\|F(\hat{x})\|_2$
$(1,\,0)^T$	1.e-1	11	x^*	6.5087e-07
$(1,-1)^T$	1.e-5	6	x^{**}	7.6141e-08
$(-1,\,0)^T$	1.e-5	4	x^{**}	1.8981e-08
$(-1,-1)^T$	1.e-5	4	x^{**}	2.3110e-09

Example 5.2. Consider the function $F:\ R^n \to R^n$

$$F_i(x) = \begin{cases} c_1 g_i(x) & \text{if } g_i(x) \geq 0, \\ c_2 g_i(x) & \text{if } g_i(x) \leq 0, \end{cases}$$

where

$$g_i(x) = i - \sum_{j=1}^{i} [\cos(x_j - 1) + j(1 - \cos(x_j - 1)) - \sin(x_j - 1)].$$

When $c_1 = c_2$, $F(x)$ is continuously differentiable. Therefore $|c_1 - c_2|$ may be interpreted as the degree of nondifferentiability of F. As it is easily verified, $F(x)$ has the solutions $(1+2k_1\pi, \ldots, 1+2k_n\pi)^T$, where $k_1, \ldots, k_n$ are arbitrary integers. The example was considered by Martinez and Qi in Ref. 21. We used the ϵ-generalized Jacobian method to compute this example. An obvious difficulty is how to compute the elements of ϵ-generalized Jacobian. Let x be the current iterate at which F is nondifferentiable. In the neighborhood of x, find a point y at which F is differentiable and take $U_k = \nabla F(y)$. However, it is difficult to design a deterministic algorithm for identifying y. In our implementation we get y by giving a random small perturbation to x. If F is still nondifferentiable at y, then repeat the above process until finally F is differentiable at y. By the Rademacher theorem, F is differentiable almost every where. Therefore, the probability that y lies in D_F is one. Clearly running the same program for the same problem at different time results in different numbers of iternations due to different perturbations. In the test we ran our program 100 times for each case. The termination criterion is $\|F(x_k)\|_2 \leq 1.0e-6$. Numerical results are displayed in Table 2 for $n = 20$ with initial point $x_0 = (0, \ldots, 0)^T$. In Table 2, ANIT denotes the average number of iterations, and ANP the average number of perturbations in the whole iteration.

Table 2. Results for Example 5.2

c_1	c_2	ANIT	ANP
1	-1	8.48	4.55
20	-20	9.71	5.76
100	-100	10.60	6.63

Example 5.3. Consider the following nonlinear complementarity problem: find $x \in R^4$ such that $x \geq 0$, $f(x) \geq 0$, and $x^T f(x) = 0$, where $f : R^4 \to R^4$ is given by

$$f_1(x) = 3x_1^2 + 2x_1x_2 + 2x_2^2 + x_3 + 3x_4 - 6,$$
$$f_2(x) = 2x_1^2 + x_1 + x_2^2 + 10x_3 + 2x_4 - 2,$$
$$f_3(x) = 3x_1^2 + x_1x_2 + 2x_2^2 + 2x_3 + 9x_4 - 9,$$
$$f_4(x) = x_1^2 + 3x_2^2 + 2x_3 + 3x_4 - 3.$$

The problem can be transformed into the solution of nonsmooth equations:

$$F(x) = \min\{f(x), x\} = 0,$$

where min refers to the componentwise minimum. The problem has two solutions: $x^* = (1,0,3,0)^T$ and $x^{**} = (\sqrt{6}/2, 0, 0, 0.5)^T$. $F(x)$ is differentiable at x^* but nondifferentiable at x^{**}. We used the ϵ-generalized Jacobian method to solve the problem. Some technique for the implementation of the method has been presented in Example 5.2. Also we ran the same program 100 times for each case as we did in Example 5.1. The termination criterion is still $\|F(x_k)\|_2 \leq 1.0e-6$. The results are given in Table 3, in which ANIT and ANP are just the same as those in Table 2. Note in Table 3 for initial point $(1,0,1,0)^T$, sometimes the iteration sequence converged to x^*, sometimes to x^{**}, which was due to different random perturbations.

Table 3. Results for Example 5.3

Initial point	ANIT	ANP	Solution
$(1,0,0,0)^T$	3	0	x^{**}
$(1,0,1,-5)^T$	3	0	x^{**}
$(1,0,1,0)^T$	1.96	1	x^*, x^{**}
$(1,0,0,1)^T$	3.52	1	x^{**}
$(0,0,0,1)^T$	failed		

References

1. ROBINSON, S. M., Local Structure of Feasible Sets in Nonlinear Programming, part III: Stability and Sensitivity, Mathematical Programming Study, Vol. 30, pp. 45-66, 1987.

2. ROBINSON, S. M., Newton's Method for A Class of Nonsmooth Functions, Industrial Engineering Working Paper, University of Wisconsin, Madison, WI, 1988.

3. ROBINSON, S. M., An Implicit Function Theorem for B-differentiable Functions, Industrial Engineering Working Paper, University of Wisconsin, Madison, WI, 1988.

4. HARKER, P. T., and XIAO, B., Newton's Method for Nonlinear Complementarity Problem: A B-differentiable Equation Approach, Mathematical Programming, Vol. 48, pp.339-35 7, 1990.

5. PANG, J. S., Newton's Method for B-differentiable Equations, Mathematics of Operation Research, Vol. 15, pp.311-341, 1990.

6. PANG, J. S., A B-differentiable Equation-based, Globally, and Locally Quadratically Convergent Algorithm for Nonlinear Programs, Complementarity and Variational Inequality Problems, Mathematical Programming, Vol. 51, pp.101-131, 1991.

7. QI, L., Superlinearly Convergent Approximate Newton Methods for LC^1 Optimization Problems, Mathematical Programming, Vol. 64, pp.277-294, 1994.

8. SUN, J., and QI, L., An Interior Point Algorithm of $\circ(\sqrt{m} \mid ln\epsilon \mid)$ Iterations for C^1-Convex Programming, Mathematical Programming, Vol. 57, pp.239-257, 1992.

9. XIAO, B., and HARKER, P. T., A Nonsmooth Newton Method for Variational Inequalities, I: Theory, Mathematical Programming, Vol. 65, pp.151-194, 1994.

10. XIAO, B., and HARKER, P. T., A Nonsmooth Newton Method for Variational Inequalities, II: Numerical Results, Mathematical Programming, Vol. 65, pp.195-216, 1994.

11. KOJIMA, M., and SHINDO, S., Extensions of Newton and Quasi-Newton Methods to Systems of PC^1 Equations, Journal of Operation Research Society of Japan, Vol. 29, pp.352-374, 1986.

12. IP, C. M., and KYPARISIS, J., Local Convergence of Quasi-Newton Methods for B-differentiable Equations, Mathematical Programming, Vol. 58, pp.71-89, 1992.

13. QI, L., and SUN, J., A Nonsmooth Version of Newton's Method, Mathematical Programming, Vol. 58, pp.353-367, 1993.

14. CLARKE, F. K., Optimization and Nonsmooth Analysis, John Wiley and Sons, New York, 1983.

15. PANG, J. S., and QI, L., Nonsmooth Equations: Motivation and Algorithms, SIAM, Journal on Optimization, Vol. 3, pp.443-465, 1993.

16. QI, L., Convergence Analysis of Some Algorithms for Solving Nonsmooth Equations, Mathematics of Operation Research, Vol. 18, pp.227-244, 1993.

17. CHEN, X., and QI, L., A Parameterized Newton Method and A Quasi-Newton Method for Solving Nonsmooth Equations, Computational Optimization and Applications, Vol. 3, pp.157-179, 1994.

18. XU, H., A New Version of Newton's Method for Nonsmooth Equations, Dundee University, Numerical Report No. 158, 1994.

19. ORTEGA, J. M., and RHEINBOLDT, W. C., Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York, London, 1970.

20. MAYNE, D. Q., and POLAK, E., Non-differentiable Optimization Adaptive Smoothing, Journal of Optimization Theory and Applications, Vol. 43, pp.601-613, 1984.

21. MARTINEZ, J. M. and QI, L., Inexact Newton Methods for Solving Nonsmooth Equations, Journal of Computational and Applied Mathematics (to appear).