JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 93, No. 2, pp. 395–415, MAY 1997

New Version of the Newton Method for Nonsmooth Equations[1]

H. XU[2] AND B. M. GLOVER[3]

Communicated by D. Q. Mayne

Abstract. In this paper, an inexact Newton scheme is presented which produces a sequence of iterates in which the problem functions are differentiable. It is shown that the use of the inexact Newton scheme does not reduce the convergence rate significantly. To improve the algorithm further, we use a classical finite-difference approximation technique in this context. Locally superlinear convergence results are obtained under reasonable assumptions. To globalize the algorithm, we incorporate features designed to improve convergence from an arbitrary starting point. Convergence results are presented under the condition that the generalized Jacobian of the problem function is nonsingular. Finally, implementations are discussed and numerical results are presented.

Key Words. Nonsmooth mappings, weak Jacobians, semismooth functions, finite-difference approximations, inexact Newton methods, global convergence.

1. Introduction

We will study the following system of nonsmooth equations:

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

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

[1]The authors are grateful to two referees and Professor D. Q. Mayne for insightful comments and constructive suggestions. They are also thankful to Professors L. Qi, M. Kojima, and X. Chen for providing additional references, and to R. Fletcher, S. Leyffer, S. C. Eisentat, and H. Walker for many helpful discussions.

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

[3]Senior Lecturer, School of Information Technology and Mathematical Sciences, University of Ballarat, Ballarat, Victoria, Australia.

The problem of finding solutions to such systems arises, in particular, in the study of nonlinear complementarity problems and both smooth and nonsmooth optimization problems; see Refs. 1–7.

There has been considerable recent attention in the literature on numerical methods for solving (1). The main methods of solution are based on the classical Newton method and quasi-Newton methods. Quasi-Newton methods are efficient in the smooth case as they do not require the computation of a Jacobian at each iterate. Unfortunately, they cannot be easily generalized to the nonsmooth case. Ip and Kyparisis (Ref. 8) obtained local linear convergence of general quasi-Newton methods under the assumptions that a bound on the deterioration of the updating matrix can be maintained and that F is B-differentiable at a solution point. They also obtained local superlinear convergence results under the condition that F is Fréchet differentiable at a solution point. This is of course restrictive for nonsmooth functions. Chen and Qi (Ref. 9) obtained stronger results for some specific classes of mappings. However, it is unclear yet whether linear and superlinear convergence results can be obtained when the Broyden method is applied to general locally Lipschitz mappings.

In contrast, the research on generalized Newton methods is seemingly more promising; see Refs. 4 and 10–14 for details. Among the available surveys, the recent one by Qi and Sun (Ref. 14) seems to be most interesting. They have developed a generalized Newton iterative scheme for (1); the iteration matrix is explicitly taken from the generalized Jacobian defined by Clarke (Ref. 15). Similar to the classical Newton method, there are essentially two questions to be answered for the generalized Newton method: (a) How can the iteration matrix or even the generalized Jacobian be computed? and (b) How can the algorithm be globalized? In relation to (b), Qi (Ref. 13) developed a globally convergent hybrid method. However, there are two conditions inherent in implementing this algorithm: One is that an element of $\partial_B F(x)$ is available; the other is that $\partial_B F(x)$ must be nonsingular. Chen and Qi (Ref. 9) successfully avoided the difficulty caused by the possible singularity of $\partial_B F(x)$ by using a parameterized Newton method. However, the problem of how to compute an element of the generalized Jacobian remains unresolved. In Ref. 16, Xu defined a type of ϵ-generalized Jacobian to approximate the actual generalized Jacobian, but the computation is slightly complicated, since the Lipschitz continuity of F has not as yet been fully exploited.

The main purpose of this paper is to present algorithms which attempt to avoid the computation of the generalized Jacobian. For this purpose, we introduce an inexact Newton iterative scheme in which the iterates produce points at which F is differentiable. This is, at least theoretically, guaranteed by the Rademacher theorem, which states that a locally Lipschitz mapping is almost everywhere differentiable. In this way, we make the generalized Newton method implementable when it is applied to problems in which an element of the generalized Jacobian is not readily available. We are also able to avoid the exact solution of a linear Newton system at each iteration, which may need considerable effort when the problem is of large scale or the current point is far from the solution point. Indeed, this was the original motivation of Dembo, Eisenstate, and Steihaug, who first proposed an inexact Newton method in Ref. 17. To improve our algorithm further, we use a classical finite-difference approximation technique in this context. We also incorporate features designed to improve convergence from an arbitrary starting point by referring to recent work of Eisenstat and Walker (Ref. 18).

We should comment briefly on the inexact Newton method for nonsmooth equations proposed recently by Martinez and Qi (Ref. 19). They investigated two inexact Newton schemes and employed the Broyden scheme to solve the linear Newton systems at each iterate. The fundamental difference of the work presented here from that of Ref. 19 is that their results are based on the assumption that an element of the generalized Jacobian is available.

The remainder of this paper is organized as follows. In Section 2, we define a weak Jacobian, which was first considered by Qi in Ref. 13, and the concept of semismoothness for Lipschitz mappings. We show that the theory developed by Qi and Sun can be explained from an alternative point of view. In Section 3, we develop an inexact Newton method and present an analysis of local convergence results under reasonable assumptions. In Section 4, we present a further version of our algorithm by using a classic finite-difference approximation technique. In Section 5, we establish a global inexact Newton algorithm for (1). We show that the algorithm is globally convergent under the condition that the generalized Jacobian is nonsingular at the limiting point of the iterates. Initial numerical results are presented and discussed in Section 6.

2. Weak Jacobian and Corresponding Newton Iteration

Throughout the paper, we use $\|\cdot\|$ to denote the standard Euclidean norm in R^n and the induced matrix norm in the real $n \times n$ matrix space $L(R^n)$. We denote by D_F the set of points in R^n at which F is differentiable. We let $S(x, \delta)$ denote an open ball in R^n (sometimes in R^m, when $x \in R^m$) with center x and radius δ. The closure of a set S is denoted by $\bar{S}$. For convenience, the open unit ball in $L(R^n)$ will be denoted by B. If $\{\alpha_k\}$ and $\{\beta_k\}$ are two real sequences, and if $\alpha_k/\beta_k \to 0$, as $k \to \infty$, then we say

$\alpha_k = o(\beta_k)$. If there exist positive constants C', C'' such that

$$|\alpha_k| \leq C'|\beta_k|, \qquad |\beta_k| \leq C''|\alpha_k|,$$

for sufficiently large k, then we say that $\alpha_k = O(\beta_k)$.

We assume throughout that F is locally Lipschitz continuous on R^n in the sense that, for every x, there exist a positive constant L and a sufficiently small positive constant δ such that

$$\|F(y) - F(z)\| \leq L\|y - z\|, \qquad \forall y, z \in S(x, \delta). \tag{2}$$

Here, L is called the Lipschitz constant of F near x.

According to the Rademacher theorem, F is differentiable almost everywhere in R^n.

Definition 2.1. The set

$$WF(x) = \lim_{x_i \in D_F, x_i \to x} \{JF(x_i)\}$$

is called the weak Jacobian of F at x, where $JF(x)$ denotes the Jacobian of F at the differentiable point x.

This set was first considered by Qi in Ref. 13. Generally $WF(x) \neq JF(x)$, even when $x \in D_F$.

Proposition 2.1. For every $x \in R^n$, the following results hold:

(a) $WF(x)$ is nonempty, bounded, and closed;
(b) $WF(x)$ is upper semicontinuous at x in the sense that, for every $\epsilon > 0$, there exists $\delta > 0$ such that

$$WF(y) \subset WF(x) + \epsilon B, \qquad \forall y \in S(x, \delta). \tag{3}$$

Proof.

(a) The proof is similar to that of Proposition 2.6.2 in Ref. 15. We include it for completeness. Let δ be sufficiently small so that (2) holds. For each $y \in D_F \cap S(x, \delta/2)$, $h \in R^n$, we have

$$F'(y, h) = JF(y)h = \lim_{t \to 0}[F(y + th) - F(y)]/t.$$

Let t be sufficiently small so that $y + th \in S(x, \delta)$. Then, by (2),

$$\|[F(y + th) - F(y)]/t\| \leq L\|h\|,$$

and hence,

$$\|JF(y)h\| \leq L\|h\|,$$

which implies

$$\|JF(y)\| \le L, \qquad \forall y \in S(x, \delta).$$

Since F is differentiable almost everywhere in $S(x, \delta)$, by Definition 2.1, $WF(x)$ must be nonempty and bounded. We now show that $WF(x)$ is closed. Let $\bar{W} \in \overline{WF(x)}$. Then, there exist a sequence $\{W_i\} \subset WF(x)$ such that $W_i \to \bar{W}$. For every W_i, there is a sequence of Jacobians $\{JF(x_{i_n})\}$ such that $JF(x_{i_n}) \to W_i$ as $x_{i_n} \to x$. Hence, there exists a subsequence $JF(x_{i_{n_k}}) \to \bar{W}$. By definition, $\bar{W} \in WF(x)$.

(b) We proceed by contradiction. For some $\epsilon_0 > 0$ and every $\delta_n > 0$, there exists $y_n \in S(x, \delta_n)$ and $V_n \in WF(y_n)$ such that $V_n \notin WF(x) + \epsilon_0 B$. For each V_n, there exists by definition a subsequence of Jacobians $\{JF(y_n^k)\}$ such that when $y_n^k \to y_n$, $JF(y_n^k) \to V_n$. Now let $\delta_n = 1/n$. Then, there exists $y_n^{k_n} \in S(y_n, 1/n^2) \cap D_F$ such that $\|JF(y_n^{k_n}) - V_n\| \le 1/n$. Since $JF(y_n^{k_n})$ is bounded, and since $y_n^{k_n} \to x$ as $n \to \infty$, there exists a subsequence of $\{JF(y_n^{k_n})\}$ converging to some $V' \in WF(x)$, which implies that a subsequence of $\{V_n\}$ tends to V'. This leads to a contradiction. The proof is complete. □

Clarke (Ref. 15) introduced a generalized Jacobian of F as follows:

$$\partial F(x) = \operatorname{conv}\{\lim_{x_i \in D_F, x_i \to x} JF(x_i)\},$$

where conv denotes convex hull. It is clear that

$$\partial F(x) = \operatorname{conv}\{WF(x)\}.$$

Definition 2.2. $WF(x)$ is said to be nonsingular if all $V \in WF(x)$ are nonsingular.

In Ref. 13, F is called strongly BD-regular at x if $WF(x)$ is nonsingular.

Assumption 2.1. For $x \in R^n$ and for all $h \in R^n$, the following limit exists:

$$\lim_{V \in WF(x+th), t \to 0} Vh.$$

Proposition 2.2. Assumption 2.1 is equivalent to the existence of the following limit:

$$\lim_{V \in \partial F(x+th), t \to 0} Vh. \tag{4}$$

Proof. It suffices to prove that Assumption 2.1 implies (4). For every $V \in \partial F(x+th)$, it follows from the Caratheodory theorem and the definition of $\partial F(x)$ that there exist

$$\lambda_i^t \in [0, 1], \quad \sum_{i=0}^{m} \lambda_i^t = 1, \qquad V_i^t \in WF(x+th), \quad i=1, \ldots, m,$$

such that

$$V = \sum_{i=0}^{m} \lambda_i^t V_i^t \quad \text{and} \quad Vh = \sum_{i=0}^{m} \lambda_i^t V_i^t h.$$

Letting $t \to 0$, we have $V_i^t h \to a$, where a is a constant. Hence, $Vh \to a$. This completes the proof. □

Definition 2.3. We say that F is semismooth if the following limit exists for any $h \in R^n$:

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

It is obvious that this definition is equivalent to that of Ref. 14. Due to the equivalence of the two definitions, we can present another version of some important results of ref. 14, which are stated without proof.

Theorem 2.1. Under Assumption 2.1, the following results hold:

(a) the classic directional derivative $F'(x, h)$ exists;
(b) if in addition, F is semismooth at x, we have
  (i) for any $V \in WF(x+h)$,

$$\lim_{h \to 0} [F'(x+h, h) - F'(x, h)]/\|h\| = 0;$$

  (ii) for $V \in WF(x+h)$, $x+h \in D_F$, $h \to 0$,

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

Now we recall briefly the version of the generalized Newton method developed by Qi (Ref. 13). Let x_c denote the current iterate and x_+ the next iterate. Suppose that $WF(x_c)$ is nonsingular. A modification of the generalized Newton iteration was given by Qi (Ref. 13) as follows:

$$x_+ = x_c - V_c^{-1} F(x_c), \tag{5}$$

where $V_c \in WF(x_c)$.

Theorem 2.2. See Lemma 2.6, Ref. 13. If $WF(x)$ is nonsingular, then there exist $\delta > 0$, $C > 0$ such that $WF(y)$ is nonsingular for all $y \in S(x, \delta)$

and

$$\|V^{-1}\| \leq C, \qquad \forall V \in WF(y),\ y \in S(x, \delta).$$

Theorem 2.3. Local Convergence. See Theorem 3.1, Ref. 13. Suppose that F is semismooth at the solution point x^* and that $WF(x^*)$ is nonsingular. Then, for x_0 sufficiently close to x^*, the iteration (5) is well defined and convergent.

3. Inexact Generalized Newton Method

It is clear that both the generalized Newton iteration in Ref. 14 and its variation (5) depend on the condition that one element of $\partial F(x)$ or $WF(x)$ is available. In general, it is difficult to calculate either $WF(x)$ or $\partial F(x)$. However, when $x \in D_F$, $JF(x)$ exists and it is computable. This motivates us to design a new version of the Newton method for (1).

Let x_c denote the current point. Suppose that $x_c \in D_F$. Compute

$$y_c = x_c - JF(x_c)^{-1}F(x_c).$$

If $y_c \in D_F$, then set $x_+ = y_c$ and repeat the iteration. Otherwise, choose a proper perturbation r_c so that

$$x_+ = y_c + r_c \in D_F.$$

The existence of such an r_c is theoretically guaranteed by the Rademacher theorem. In this way, we keep the remaining iterates in the set D_F. Accordingly, we have the following new algorithm.

Algorithm 3.1.

Step 0. Given $x_0 \in D_F$, set $k = 0$.
Step 1. If $F(x_k) = 0$, stop.
Step 2. Solve the Newton system

$$JF(x_k)s_k = -F(x_k); \tag{6}$$

let $y_k = x_k + s_k$ and

$$x_{k+1} = \begin{cases} y_k, & \text{if } y_k \in D_F, \\ y_k + r_k, & \text{otherwise,} \end{cases}$$

where r_k is appropriately chosen.
Step 3. Set $x_k = x_{k+1}$, $k = k + 1$; go to Step 1.

We have another motivation for introducing the modified scheme. For some large-scale problems, it can be very expensive to solve (6) exactly using a direct method such as Gaussian elimination; this may not be justified when x_k is far from x^*; see Ref. 17. It is often reasonable to solve the subproblem inexactly by an iterative method such as the conjugate gradient method, while the convergence rate is not significantly reduced. In this case, we regard r_k as a residual. For both the purposes stated above, we modify Algorithm 3.1 as follows.

Algorithm 3.2.

Step 0. Given $x_0 \in D_F$, set $k=0$.
Step 1. If $F(x_k)=0$, stop.
Step 2. Find some $\eta_k \in [0, 1)$ and s_k such that

$$JF(x_k)s_k = -F(x_k) + r_k, \tag{7a}$$

$$x_k + s_k \in D_F, \tag{7b}$$

$$\|r_k\|/\|F(x_k)\| \le \eta_k. \tag{7c}$$

Step 3. Step $x_{k+1} = x_k + s_k$, $k=k+1$; go to Step 1.

Here, η_k may depend on x_k.

We now present an analysis of the convergence results for Algorithm 3.2.

Theorem 3.1. Let x^* be a solution point of (1). Suppose that F is semismooth at x^* and that $WF(x^*)$ is nonsingular. Assume also that $\eta_k \le \eta_{\max} < 1/(2+L)C$, where L is the Lipschitz constant of F at x^* and C is some constant. Then, there exists $\delta > 0$ such that, if $x_0 \in S(x^*, \delta) \cap D_F$, then the sequence $\{x_k\}$ generated by Algorithm 3.2 is well defined and converges to x^*. Moreover, the convergence is linear, in the sense that

$$\|x_{k+1} - x^*\| \le t\|x_k - x^*\|, \qquad 0 < t < 1. \tag{8}$$

Proof. By assumption, $WF(x^*)$ is nonsingular. It follows from Theorem 2.2 that there exist $C > 0$ and $\delta_1 > 0$ such that $JF(x_0)$ is nonsingular and

$$\|JF(x_0)^{-1}\| \le C,$$

for $x_0 \in S(x^*, \delta_1) \cap D_F$. Hence,

$$\begin{aligned} \|x_1 - x^*\| &= \|x_0 - JF(x_0)^{-1}F(x_0) + JF(x_0)^{-1}r_0 - x^*\| \\ &\le C(\|r_0\| + \|F(x_0) - F(x^*) - F'(x^*, x_0 - x^*)\| \\ &\quad + \|F'(x^*, x_0 - x^*) - JF(x_0)(x_0 - x^*)\|). \end{aligned} \tag{9}$$

On the other hand, it follows from (2) and (7) that there exist constants L, δ_2 (with $0<\delta_2\leq\delta_1$) such that

$$\|r_0\|\leq\eta_0\|F(x_0)\|=\eta_0\|F(x_0)-F(x^*)\|\leq\eta_0 L\|x_0-x^*\|, \tag{10}$$

for $x_0\in S(x^*, \delta_2)\subset D_F$. By Theorem 2.1, for $\eta_0>0$, there exists, with $0<\delta<\delta_2$, such that

$$\|F(x_0)-F(x^*)-F'(x^*, x_0-x^*)\|\leq\eta_0\|x_0-x^*\|, \tag{11}$$

$$\|F'(x^*, x_0-x^*)-JF(x_0)(x_0-x^*)\|\leq\eta_0\|x_0-x^*\|, \tag{12}$$

for $x_0\in S(x^*, \delta)\cap D_F$. Now, letting

$$\eta_0\leq\eta_{\max}, \qquad x_0\in S(x^*, \delta)\subset D_F,$$

and combining (9)–(12), we have

$$\|x_1-x^*\|\leq\eta_{\max}(L+2)C\|x_0-x^*)\|.$$

Setting

$$t=\eta_{\max}(L+2)C, \qquad \eta_{\max}(L+2)C<1,$$

we obtain (8) for $k=0$. The rest can be easily obtained by induction. The proof is complete. □

From the proof of the theorem, we observe that η_k influences the convergence rate of Algorithm 3.2. We can accelerate the convergence rate by reducing $\|r_k\|$.

Theorem 3.2. Let x^* be a solution of (1). Suppose that F is semismooth at x^* and that $WF(x^*)$ is nonsingular. The points $\{x_k\}$ generated by Algorithm 3.2 converge to x^*. Then, $\{x_k\}$ converges to x^* superlinearly iff

$$\|r_k\|=o(\|F(x_k)\|). \tag{13}$$

Proof. Assume that x_k converges superlinearly to x^*. Note that $x_k\in D_F$. Then,

$$\begin{aligned} r_k &= JF(x_k)(x_{k+1}-x_k)+F(x_k)\\ &= F(x_k)-F(x^*)-F'(x^*, x_k-x^*)+F'(x^*, x_k-x^*)\\ &\quad -F'(x_k, x_k-x^*)+F'(x_k, x_{k+1}-x^*). \end{aligned}$$

It follows from Theorem 2.1 that

$$\begin{aligned}\|r_k\| &\le \|F(x_k) - F(x^*) - F'(x^*, x_k - x^*)\| \\ &\quad + \|F'(x^*, x_k - x^*) - F'(x_k, x_k - x^*)\| + \|JF(x_k)\| \, \|x_{k+1} - x^*\| \\ &= o(\|x_k - x^*\|) + o(\|x_k - x^*\|) + O(\|x_{k+1} - x^*\|),\end{aligned}$$

for sufficiently large k. Since $\{x_k\}$ converges superlinearly to x^*, we have

$$O(\|x_{k+1} - x^*\|) = o(\|x_k - x^*\|^{1+p}),$$

for some $p > 0$, which proves (13). Conversely, suppose that (13) holds. The proof proceeds in a similar way to that of Theorem 3.1, and hence the details are omitted. The proof is complete. □

4. Finite Difference Approximations

Just as in the smooth case, following Ref. 20 we now proceed to discuss how to find a substitute for $JF(x)$ in Algorithm 3.2 when the computation of the derivatives is very complicated or even impossible. We propose to use a finite-difference matrix to approximate the Jacobian $JF(x)$.

We first give a general definition of difference approximation.

Definition 4.1. Let $A: S_A \times S_h \subset R^n \times R^m \to L(R^n)$. Then, A is called a discrete consistent approximation (DCA) to $JF(x)$ on $S_0 \cap D_F \subset S_A$ if $0 \in R^m$ is a limiting point of S_h and

$$\lim_{h \in S_h, h \to 0} A(x, h) = JF(x), \qquad \text{uniformly for } x \in S_0 \cap D_F. \tag{14}$$

If there are constants M and $\delta > 0$ such that

$$\|A(x, h) - JF(x)\| \le M\|h\|, \qquad \forall x \in S_0 \cap D_F, \forall h \in S_h \cap S(0, \delta), \tag{15}$$

then A is a discrete strongly consistent approximation (DSCA) to $JF(x)$ on $S_0 \cap D_F$.

The definition is a generalization of Definition 11.2.1 of Ortega and Rheinboldt (Ref. 21). The word discrete is employed to emphasize that (14) or (15) only holds in $S_0 \cap D_F$ rather than S_0. The former may be discrete.

We say that $F(x)$ is first-order discrete consistently (FODC) approximatable in R^n if, for every $x \in R^n$, there exist a mapping $A: S_A \times S_h \subset R^n \times R^m \to L(R^n)$ and a neighborhood of x (S_0 say) such that (14) holds. If in addition (15) holds, then $F(x)$ is called first-order discrete strongly consistently (FODSC) approximatable in R^n.

Many mappings arising in applications are FODSC approximatable in R^n. For example, consider the standard nonlinear complementarity problem: Find x solving

$$h(x) \geq 0, \qquad f(x) \geq 0, \qquad h(x)^T f(x) = 0,$$

where h and f are two continuously differentiable functions from R^n to R^m. This problem can be formulated as

$$H(x) = \min\{f(x), h(x)\} = 0.$$

Clearly $H(x)$ is FODSC approximatable in R^n.

More generally, the piecewise C^1 mapping defined by Kojima and Shindo (Ref. 10) is FODSC approximatable in R^n.

We now consider another version of Algorithm 3.2.

Algorithm 4.1.

Step 0. Given $x_0 \in D_F$, set $k = 0$.
Step 1. If $F(x_k) = 0$, stop.
Step 2. Find some $\eta_k \in [0, 1)$ and s_k such that

$$A(x_k, h_k)s_k = -F(x_k) + r_k,$$

$$\|r_k\| \leq \eta_k \|F(x_k)\|,$$

$$x_k + s_k \in D_F,$$

where h_k is appropriately chosen.
Step 3. Set $x_{k+1} = x_k + s_k$, $k = k + 1$; go to Step 1.

Theorem 4.1. Let x^* be a solution point of (1). Suppose that $F(x)$ is FODSC approximatable in R^n and semismooth at x^*, that $WF(x^*)$ is nonsingular, and that $A(x, h)$ is a DSCA to $JF(x)$ in the neighborhood of x_k, with $\|h\| = O(\|F(x_k)\|)$. Assume also that $\eta_k \leq \eta_{\max} < 1/2(L+2)C$, where L is the Lipschitz constant of F at x^* and C is some constant. Then, there exists $\delta > 0$ such that, if $x_0 \in S(x^*, \delta) \cap D_F$, the sequence $\{x_k\}$ generated by Algorithm 4.1 converges to x^*. Moreover, the convergence is linear in the sense that

$$\|x_{k+1} - x^*\| \leq t\|x_k - x^*\|, \qquad 0 < t < 1. \tag{16}$$

Before we present the proof, we give a simple preliminary result.

Lemma 4.1. Suppose that $F(x)$ is FODSC approximatable in R^n and that $A(x, h)$ is a DSCA to $JF(x)$ in the neighborhood of $x \in D_F$, with $\|h\| =$

$O(\|F(x)\|)$; suppose that x^* is a solution of (1) and that $WF(x^*)$ is nonsingular. Then, there exist $\delta > 0$, $C > 0$, such that $A(x, h)$ is nonsingular and

$$\|A(x, h)^{-1}\| \leq C, \qquad \forall x \in S(x^*, \delta) \cap D_F.$$

Proof. Notice that, for $x \in D_F$, $A(x, h)$ is a DSCA to $JF(x)$ in the neighborhood of x. Hence,

$$A(x, h) = JF(x) + O(\|h\|) = JF(x) + O(\|F(x)\|), \qquad \text{as } h \to 0.$$

By Lipschitz continuity, there exists $\delta_1 > 0$ such that

$$\|F(x)\| = \|F(x) - F(x^*)\| \leq L\|x - x^*\|, \qquad \forall x \in S(x^*, \delta_1).$$

Hence,

$$A(x, h) = JF(x) + O(\|x - x^*\|), \qquad \text{as } x \to x^*; \tag{17}$$

equivalently, for every $\epsilon > 0$, there exists $0 < \delta_2 < \delta_1$ such that

$$\|A(x, h) - JF(x)\| \leq \epsilon, \qquad \forall x \in S(x^*, \delta_2) \cap D_F.$$

On the other hand, by Proposition 2.1, for $\epsilon > 0$, there exists $0 < \delta \leq \delta_2$ such that

$$JF(x) \in WF(x^*) + \epsilon B;$$

that is, $JF(x)$ is nonsingular and there exists $C > 0$ such that

$$\|JF(x)^{-1}\| \leq C,$$

for sufficiently small $\epsilon > 0$ and $x \in S(x^*, \delta) \cap D_F$. By the Banach perturbation theorem, it follows that $A(x, h)$ is nonsingular and that

$$\|A(x, h)^{-1}\| \leq \|JF(x)^{-1}\| / [1 - \epsilon \|JF(x)^{-1}\| \|B\|],$$

where

$$\|B\| = \max_{E \in B} \|E\|$$

Letting

$$0 < \epsilon \leq \max[1, 1/2C\|B\|],$$

we have

$$\|A(x, h)^{-1}\| \leq 2C, \qquad \forall x \in S(x^*, \delta) \cap D_F.$$

The proof is thus complete. □

Proof of Theorem 4.1. By assumption, $WF(x^*)$ is nonsingular. It follows from Lemma 4.1 that there exists $\delta_1 > 0$ and a constant $C > 0$ such that $A(x_0, h_0)$ is nonsingular and

$$\|A(x_0, h_0)^{-1}\| \leq 2C, \qquad \forall x_0 \in S(x^*, \delta_1) \cap D_F.$$

Letting x_0 be such a point, we have

$$\begin{aligned}\|x_1 - x^*\| &= \|x_0 - A(x_0, h_0)^{-1}(F(x_0) + r_0) - x^*\| \\ &\leq \|A(x_0, h_0)^{-1}\|(\|r_0\| + \|F(x_0) - A(x_0, h_0)(x_0 - x^*)\|) \\ &\leq 2C[\|r_0\| + \|F(x_0) - F(x^*) - JF(x_0)(x_0 - x^*)\| \\ &\quad + \|(JF(x_0) - A(x_0, h_0))(x_0 - x^*)\|].\end{aligned}$$

Similar to the proof of Theorem 3.1, for every $\eta_0 > 0$, there exists $0 < \delta_2 \leq \delta_1$ such that

$$\|F(x_0) - F(x^*) - JF(x_0)(x_0 - x^*)\| \leq \eta_0 \|x_0 - x^*\|,$$

$$\|r_0\| \leq L\eta_0 \|x_0 - x^*\|,$$

for all $x_0 \in S(x^*, \delta_2) \cap D_F$.

On the other hand, it follows from (17) that, for η_0, there exists $0 < \delta \leq \delta_2$ such that

$$\|JF(x_0) - A(x_0, h_0)\| \leq \eta_0, \qquad x_0 \in S(x^*, \delta) \cap D_F.$$

Consequently,

$$\begin{aligned}\|x_1 - x^*\| &\leq 2C\eta_0(L\|x_0 - x^*\| + \|x_0 - x^*\| + \|x_0 - x^*\|) \\ &= 2(L+2)C\eta_0 \|x_0 - x^*\|.\end{aligned}$$

By letting

$$\eta_0 \leq \eta_{\max} < 1/2(L+2)C, \qquad t = 2(L+2)C\eta_{\max},$$

we obtain (16) for $k = 0$. The rest can be easily proven by induction. The proof is complete. □

Theorem 4.2. Suppose that the conditions in Theorem 4.1 are satisfied. Then, $\{x_k\}$ converges to x^* superlinearly iff

$$\|r_k\| = o(\|F(x_k)\|).$$

We shall not present the proof, since it is straightforward.

5. Global Inexact Newton Method

In the previous sections, only local convergence of the algorithms was discussed. The iterates may not necessarily converge if a starting point is far from the solution. Therefore, it should be interesting to globalize the algorithms.

For the sake of convenience, we only globalize Algorithm 3.2. In Ref. 18, Eisenstat and Walker develop a globally convergent inexact Newton method for smooth equations. They employ the following algorithm.

Algorithm 5.1. Global Inexact Newton Method.

Step 0. Given a starting point x_0 and $t \in [0, 1)$, set $k=0$.
Step 1. If $F(x_k)=0$, stop.
Step 2. Find some $\eta_k \in [0, 1)$ and s_k such that

$$\|F(x_k)+JF(x_k)s_k\| \leq \eta_k \|F(x_k)\|,$$

$$\|F(x_k+s_k)\| \leq [1-t(1-\eta_k)]\|F(x_k)\|.$$

Step 3. Set $x_{k+1}=x_k+s_k$, $k=k+1$; go to Step 1.

For our purpose, we modify Algorithm 5.1 as follows.

Algorithm 5.2. Global Inexact Newton Method.

Step 0. Given a starting point x_0 and $t \in [0, 1)$, set $k=0$.
Step 1. If $F(x_k)=0$, stop.
Step 2. Find some $\eta_k \in [0, 1)$ and s_k such that

$$x_k+s_k \in D_F, \tag{18}$$

$$F(x_k)+JF(x_k)s_k=r_k, \tag{19}$$

$$\|r_k\| \leq \eta_k \|F(x_k)\|, \tag{20}$$

$$\|F(x_k+s_k)\| \leq [1-t(1-\eta_k)]\|F(x_k)\|. \tag{21}$$

Step 3. Set $x_{k+1}=x_k+s_k$, $k=k+1$; go to Step 1.

Before analyzing the convergence of the new algorithm, we discuss the existence of an inexact Newton step s_k in Step 2.

Proposition 5.1. There exists s_k satisfying (18)–(20) if and only if either $F(x_k)=0$ or $F(x_k) \not\perp R(JF(x_k))$, where $R(A)$ denotes the range of A.

Proof. We consider only the case where $F(x_k) \neq 0$. Inequality (20) is equivalent to

$$\|JF(x_k)s_k\|^2+2F(x_k)^T JF(x_k)s_k+(1-\eta_k^2)\|F(x_k)\|^2 \leq 0;$$

letting $s_k=\alpha d$, and substituting it into the above, we have

$$\alpha^2\|JF(x_k)d\|^2+2F(x_k)^T JF(x_k)d\alpha+(1-\eta_k^2)\|F(x_k)\|^2 \leq 0. \tag{22}$$

If $F(x_k) \perp R(JF(x_k))$, then there are no real α, d satisfying (22); i.e., there is no s_k satisfying (19)–(20). If $F(x_k) \not\perp R(JF(x_k))$, then there exists $\bar{d}$ satisfying

$$(F(x_k)^T JF(x_k)\bar{d})^2 - (1 - \eta_k^2)\|F(x_k)\|^2 \|JF(x_k)\bar{d}\|^2 > 0;$$

thus, there exists a real number $\bar{\alpha}$ satisfying (22) for $d = \bar{d}$; equivalently, there exists $\bar{s}_k$ satisfying (19)–(20). An arbitrary small perturbation of $\bar{s}_k$ can be made under the same condition and s_k satisfying (18)–(20) exists. The proof is complete. □

Lemma 5.1. Let $x \in D_F$ and $t \in [0, 1)$ be given. Assume that there exists $\bar{s}$ satisfying $\|F(x) + JF(x)\bar{s}\| < \|F(x)\|$. Then, there exists $\eta_{\min} \in [0, 1)$ such that, for any $\eta \in [\eta_{\min}, 1)$, there exists s satisfying

$$\|F(x) + JF(x)s\| \leq \eta \|F(x)\|, \tag{23}$$

$$\|F(x+s)\| \leq [1 - t(1 - \eta)]\|F(x)\|, \tag{24}$$

and $x + s \in D_F$.

Proof. There are two differences between this lemma and Lemma 3.1 of Ref. 18. First, F is assumed to be continuously differentiable in Lemma 3.1 of Ref. 18, while here we only require F to have a Fréchet derivative at x. Second, we require $x + s \in D_F$. Actually, Lemma 3.1 of Ref. 18 can be proved without any modifications under the condition that F has a Fréchet derivative at x. We do not restate it here, but this result does not guarantee that $x + s \in D_F$. Suppose now that there is s satisfying (23)–(24), but $x + s \notin D_F$. It is clear that (23)–(24) still holds when an arbitrary small perturbation on s is made and η is replaced by $(1 + 1/M)\eta$, where M is sufficiently large such that $\eta_{\min}(1 + 1/M) < 1$. By the Rademacher theorem, we can find δs such that $x + s + \delta s \in D_F$ and $x + s + \delta s$ satisfies (23)–(24). The proof is complete. □

Remark 5.1. Algorithm 5.2 breaks down at some x_k if and only if $F(x_k) \neq 0$ and x_k is a stationary point of $\|F(x)\|$.

Theorem 5.1. Assume that $\{x_k\} \subset D_F$ is a sequence such that $F(x_k) \to 0$ and, for each k,

$$\|F(x_k) + JF(x_k)s_k\| \leq \eta \|F(x_k)\|,$$

$$\|F(x_{k+1})\| \leq \|F(x_k)\|,$$

where $s_k = x_{k+1} - x_k$ and η is independent of k. If x^* is a limit point of $\{x_k\}$ such that F is semismooth at x^* and $\partial F(x^*)$ is nonsingular, then $F(x^*) = 0$ and $x_k \to x^*$.

Proof. Clearly, $F(x^*)=0$. It follows from Proposition 3.1 of Ref. 14 that there exist $\delta>0$, $C>0$ such that $\partial F(y)$ is nonsingular and

$$\|V^{-1}\|\leq C, \qquad \forall V\in\partial F(y),\ y\in S(x,\delta).$$

On the other hand, by Theorem 2.1,

$$\|F(y)-F(x^*)-F'(x^*,y-x^*)\|\leq(1/2C)\|y-x^*\|,$$

for δ sufficiently small. Consequently,

$$\begin{aligned}\|F(y)\|&\geq\|F'(x^*,y-x^*)\|-\|F(y)-F(x^*)-F'(x^*,y-x^*)\|\\&\geq\|F'(x^*,y-x^*)\|-(1/2C)\|y-x^*\|.\end{aligned}$$

By Lemma 2.2 of Ref. 14, there exists $V\in\partial F(x^*)$ such that

$$F'(x^*,y-x^*)=V(y-x^*).$$

Thus,

$$\|F'(x^*,y-x^*)\|\geq(1/\|V^{-1}\|)\|y-x^*\|\geq(1/C)\|y-x^*\|.$$

Consequently,

$$\|F(y)\|\geq(1/2C)\|y-x^*\|;$$

equivalently,

$$\|y-x^*\|\leq 2C\|F(y)\|, \qquad \forall y\in S(x^*,\delta). \tag{25}$$

Let $\epsilon\in(0,\delta/4)$. Since x^* is a limit point of x_k and $F(x^*)=0$, there exists k such that

$$x_k\in S_\epsilon\equiv[y:\ \|F(y)\|<\epsilon/[C(1+\eta)],\ y\in S(x^*,\delta/2)].$$

Then,

$$\begin{aligned}\|s_k\|&=\|F'(x_k)^{-1}[-F(x_k)+(F(x_k)+JF(x_k)s_k)]\|\\&\leq C(\|F(x_k)\|+\|r_k\|)\\&\leq C(1+\eta)\|F(x_k)\|\\&\leq\epsilon<\delta/4.\end{aligned}$$

So, $x_{k+1}\in S(x^*,\delta)$. Since

$$\|F(x_{k+1})\|\leq\|F(x_k)\|<\epsilon/[C(1+\eta)],$$

by (25),

$$\begin{aligned}\|x_{k+1}-x^*\|&\leq 2C\|F(x_{k+1})\|\\&\leq 2C\epsilon/[C(1+\eta)]<\delta/2,\end{aligned}$$

which implies that $x_{k+1} \in S_\epsilon$. Therefore, $x_k \in S_\epsilon$ for sufficiently large k. Since $\|F(x_k)\| \to 0$, it follows that $x_k \to x^*$. The proof is complete. □

Remark 5.2. We do not require that $\eta < 1$ in the proof.

Theorem 5.2. Global Convergence of Algorithm 5.2. Suppose that Algorithm 5.2 does not break down, if $\sum_{k \geq 0} (1 - \eta_k)$ is divergent. Then, $F(x_k) \to 0$. If in addition, x^* is a limit point of $\{x_k\}$ such that F is semismooth at x^* and $\partial F(x^*)$ is nonsingular, then $F(x^*) = 0$ and $\{x_k\} \to x^*$.

Proof. By (21),

$$\begin{aligned} \|F(x_k)\| &\leq [1 - t(1 - \eta_{k-1})] \|F(x_{k-1})\| \\ &\leq \|F(x_0)\| \prod_{0 \leq j \leq k} [1 - t(1 - \eta_j)] \\ &\leq \|F(x_0)\| \prod_{0 \leq j \leq k} \exp[-t(1 - \eta_j)]. \end{aligned}$$

It follows from the assumptions that $\|F(x_k)\| \to 0$. The rest follows from Theorem 5.1 and Remark 5.2. The proof is complete. □

6. Computational Experiments

As stated in the previous sections, our method is to maintain the iterates generated by the algorithms in D_F. One challenging issue is how to implement it practically for general locally Lipschitz continuous functions. One possible way that we suggest is to make a random perturbation r_c if $y_c = x_c + s_c \notin D_F$. Theoretically, via the Rademacher theorem, the probability that $y_c + r_c$ is in D_F is one. The process is repeated when $y_c + r_c \notin D_F$.

We implemented two typical examples taken from Refs. 8–9 and 19 with $r_k = \eta_k \|F(x_k)\| d_k$, where d_k is a random vector satisfying $\|d_k\|_\infty < 1$. The numerical results are satisfactory as we shall demonstrate; however, further numerical experimentation is still required and will be the focus of further research.

Example 6.1. Consider the following function $F: R^n \to R^n$:

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

Table 1. Computational results, Example 6.1, Algorithm 3.2.

N_v	c_1	c_2	η_k	N_{out}	N_{in}
1	1	−1	0.5	5	0
2	1	−1	0.5	5	0
3	1	−1	0.5	5	0
4	1	−1	1/(2+k)	7	1
5	1	−1	1/(2+k)	7	1
6	1	−1	1/(2+k)	7	1
7	100	−100	0.5	64	15
8	100	−100	0.5	41	10
9	100	−100	0.5	51	12
10	100	−100	1/(2+k)	40	10
11	100	−100	1/(2+k)	35	9
12	100	−100	1/(2+k)	32	8
20	100	−100	1/(2+k)	44	12
30	100	−100	1/(2+k)	104	23
40	100	−100	1/(2+k)	228	54

where

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

This example was considered by Martinez and Qi in Ref. 19. As it is easily verified, the solution of $F(x)=0$ is $(1+2k_1\pi,\ldots,1+2k_n\pi)^T$, where $k_1,\ldots,k_n$ are arbitrary integers. By adjusting the difference $|c_1-c_2|$, one may get different degrees of nondifferentiability of F.

We implemented Algorithm 3.2 for this example with the starting point $x_0=(0,\ldots,0)^T$. The numerical results are presented in Table 1. In the table, N_v denotes the number of variables of the problem, N_{out} denotes the number of Newton iterations needed to reach the required precision, which is specified as exp(−6), while N_{in} represents the number of times that the random correction in Step 2 was carried out.

In comparison with Table 1 of Ref. 19, we find that generally the number of outer iterations is greater than in Ref. 19 when the problem dimension increases. Although the two N_{in} columns have different meanings, we can still conclude that the policy of a random correction in Step 2 is successful.

Computational experiments were also conducted for another set of problems arising from nonlinear complementarity problems.

Example 6.2. Consider the following nonlinear complementarity problem: Find $x\in R^4$ such that

$$x\geq 0,\qquad f(x)\geq 0,\qquad x^Tf(x)=0,$$

Table 2. Computational results, Example 6.2, Algorithm 3.2.

Starting point	N_{out}	N_{in}	$\|F(x_k)\|$	Solution
$(1, 0, 1, -5)^T$	5	1	3.7007E−09	$(\sqrt{6}/2, 0, 0, 0.5)^T$
$(1, 0, 1, 0)^T$	4	1	8.8413E−10	$(\sqrt{6}/2, 0, 0, 0.5)^T$
$(1, 0, 0, 1)^T$	4	1	5.0936E−07	$(1, 0, 3, 0)^T$
$(1, 0, 0, 0)^T$	5	0	3.7007E−09	$(\sqrt{6}/2, 0, 0, 0.5)^T$
$(0, 0, 0, 1)^T$	Failed	—	—	—

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 is equivalent to solving the nonsmooth equations

$$F(x) = 0, \qquad \text{where } F(x) = \min[f(x); x],$$

and min denotes the componentwise minimum. This problem has two solutions,

$$x^* = (1, 0, 3, 0)^T, \qquad x^{**} = (\sqrt{6}/2, 0, 0, 0.5)^T.$$

Also, $F(x)$ is differentiable at x^* but nondifferentiable at x^{**}. Numerical results are displayed in Table 2. The algorithm failed at the starting point $(0, 0, 0, 1)^T$, since it is only a local algorithm.

Furthermore, we applied Algorithm 4.1 to Example 6.2. Numerical results are displayed in Table 3. In Step 2, h_k was simply $(0.01, 0.01, 0.01, 0.01)^T$. From Tables 2 and 3, we find that there is little significant difference between the two algorithms.

Finally it should be noted that Algorithm 5.2 is yet to be implemented due to the practical difficulties inherent in Step 2 of the algorithm. This will be the focus of future research.

Table 3. Computational results, Example 6.2, Algorithm 4.1.

Starting point	N_{out}	N_{in}	$\|F(x_k)\|$	Solution
$(1, 0, 1, -5)^T$	6	0	2.3747E−07	$(\sqrt{6}/2, 0, 0, 0.5)^T$
$(1, 0, 1, 0)^T$	5	1	9.6286E−08	$(\sqrt{6}/2, 0, 0, 0.5)^T$
$(1, 0, 0, 1)^T$	5	1	6.5154E−08	$(1, 0, 3, 0)^T$
$(1, 0, 0, 0)^T$	6	0	2.3747E−07	$(\sqrt{6}/2, 0, 0, 0.5)^T$
$(0, 0, 0, 1)^T$	Failed	—	—	—

References

1. Qi, L., and Jiang, H., *Semismooth Karush–Kuhn–Tucker Equations and Convergence Analysis of Newton and Quasi-Newton Methods for Solving These Equations*, Report 5, Applied Mathematics Department, University of New South Wales, 1994.
2. Harker, P. T., and Xiao, B., *Newton's Method for Linear Complementarity Problem: A B-Differentiable Equation Approach, Mathematical Programming*, Vol. 48, pp. 339–357, 1990.
3. 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.
4. Pang, J. S., and Qi, L., *Nonsmooth Equations: Motivation and Algorithms*, SIAM Journal on Optimization, Vol. 3, pp. 443–465, 1993.
5. Qi, L., *Superlinearly Approximate Newton Methods for* LC^1*-Optimization Problems*, Mathematical Programming, Vol. 64, pp. 277–294, 1994.
6. Sun, J., and Qi, L., *An Interior-Point Algorithm of* $O(\sqrt{m}|\log \epsilon|)$ *Iterations for* C^1*-Convex Programming*, Mathematical Programming, Vol. 57, pp. 239–257, 1992.
7. Xiao, B., and Harher, P. T., *A Nonsmooth Newton Method for Variational Inequalities, Part 1: Theory*, Mathematical Programming, Vol. 65, pp. 151–194, 1994.
8. Ip, C. M., and Kyparisis, J., *Local Convergence of Quasi-Newton Methods for B-Differentiable Equations*, Mathematical Programming, Vol. 58, pp. 71–89, 1992.
9. Chen, X., and Qi, L., *A Parametrized Newton Method and a Quasi-Newton Method for Nonsmooth Equations*, Computational Optimization and Applications, Vol. 3, pp. 157–179, 1994.
10. Kojima, M., and Shindo, S., *Extensions of Newton and Quasi-Newton Methods to Systems of* PC^1*-Equations*, Journal of the Operation Research Society of Japan, Vol. 29, pp. 352–374, 1986.
11. Kummer, B., *Newton's Method for Nondifferentiable Functions*, Advances in Mathematical Optimization, Edited by J. Guddat et al. Akademie Verlag, Berlin, Germany, pp. 114–125, 1988.
12. Pang, J. S., *Newton's Method for B-Differentiable Equations*, Mathematics of Operation Research, Vol. 15, pp. 311–341, 1990.
13. Qi, L., *Convergence Analysis of Some Algorithms for Solving Nonsmooth Equations*, Mathematics of Operation Research, Vol. 18, pp. 227–244, 1993.
14. Qi, L., and Sun, J., *A Nonsmooth Version of Newton's Method*, Mathematical Programming, Vol. 58, pp. 353–367, 1993.
15. Clarke, F. H., *Optimization and Nonsmooth Analysis*, John Wiley and Sons, New York, New York, 1983.
16. Xu, H., *Approximate Newton Methods for Nonsmooth Equations*, Report 158, Mathematics and Computer Sciences Department, Dundee University, Dundee, Scotland, 1994.

17. DEMBO, R. S., EISENTAT, S. C., and STEIHAUG, T., *Inexact Newton Methods*, SIAM Journal on Numerical Analysis, Vol. 19, pp. 400–408, 1982.
18. EISENSTAT, S. C., and WALKER, H. F., *Globally Convergence Inexact Newton Methods*, SIAM Journal on Optimization, Vol. 4, pp. 393–422, 1994.
19. MARTINEZ, J. M., and QI, L., *Inexact Newton Methods for Solving Nonsmooth Equations*, Journal of Computational and Applied Mathematics, Vol. 60, pp. 127–145, 1995.
20. DENNIS, J. E., and MORE, J. J., *Quasi-Newton Methods: Motivation and Theory*, SIAM Review, Vol. 19, pp. 46–89, 1977.
21. ORTEGA, J. M., and RHEINBOLDT, W. C., *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, New York, 1970.