Journal of Computational Mathematics, Vol.20, No.3, 2002, 289–300.

GLOBALLY CONVERGENT INEXACT GENERALIZED NEWTON METHODS WITH DECREASING NORM OF THE GRADIENT*1)

Ding-guo Pu
(*Department of Mathematics, Tongji University, Shanghai 200331, China*)

Abstract

In this paper, motivated by the Martinez and Qi methods[1], we propose one type of globally convergent inexact generalized Newton methods to solve unconstrained optimization problems in which the objective functions are not twice differentiable, but have LC gradient. They make the norm of the gradient decreasing. These methods are implementable and globally convergent. We prove that the algorithms have superlinear convergence rates under some mile conditions.

The methods may also be used to solve nonsmooth equations.

Key words: Nonsmooth optimization, Inexact Newton method, Generalized Newton method, Global convergence, Superlinear rate.

1. Introduction

Let $f: R^n \to R$ be a nonlinear function and $g = \bigtriangledown f$. The usual Newton method for solving nonlinear equations $g(x) = 0$ involves making a first order approximation at the current trial point x_k:

$$\bigtriangledown^2 f(x_k) s_k + g_k = 0 \tag{1}$$

where $g_k = g(x_k)$. We then solve this equation calling the solution x_{k+1}, namely $x_{k+1} = x_k - \bigtriangledown^2 f(x_k)^{-1} g_k$ and repeat the process until obtaining a solution of the original optimization problem.

Newton method is attractive because it converges rapidly from any sufficiently good initial guess. Indeed, it is often taken as a standard convergent method, since one way of characterizing superlinear convergence is that the step should approach Newton step asymptotically in both magnitude and direction (Dennis and Moré [2]).

If the number of variables is large, or if second derivative information is difficult to compute, Newton method may be prohibitively expensive to use. If f is not twice differentiable, Newton method can not be used. For this reason, special methods have been developed to solve large-scale problems. One method is the inexact Newton method, see [3] and [4]. That is, we do not need to solve the Newton equation (1) accurately and as long as the magnitude of the residual vector for an approximate solution s_k:

$$r_k = \bigtriangledown^2 f(x_k) s_k + g_k \tag{2}$$

is asymptotically smaller comparing with $\|g_k\|$, inexact Newton method can work well. In order to stabilize the behavior of inexact Newton method, Eisenstat and Walker in [5] introduced and analyzed the following class of methods: For $k = 0$ step 1 until convergence do

* Received April 2, 1999; Final revised February 23, 2000.
1) This research is supported by Ministry of Education P.R.C. and Asia-Pacific Operations Research Center (APORC).

1. Obtain a s_k such that

$$\frac{\|g_k + \nabla^2 f(x_k)s_k\|}{\|g_k\|} = \frac{\|r_k\|}{\|g_k\|} \leq \eta_k, \tag{3}$$

and

$$\|g(x_k + s_k)\| \leq [1 - \theta(1 - \eta_k)]\|g_k\|. \tag{4}$$

2. Set $x_{k+1} = x_k + s_k$.

On the other hand, Newton method has been extended to non-twice differentiable case. Particularly, in recent years some superlinearly convergent generalized Newton methods for solving nonsmooth equations

$$F(x) = 0 \tag{5}$$

have been developed which are based upon Clarke's generalized Jacobian $\partial F(x)$ (see [6]) or B-differentials $\partial_B F(x)$ which used to replace $\nabla^2 f_k$ in (1) or (2), see e.g. [7], [8], [9] and [1].

Stimulated by the progress in these two aspects, in this paper we propose one type of globally convergent inexact generalized Newton methods to solve unconstrained optimization problems in which the objective functions are not twice differentiable, but have LC gradient. Comparing with the globally convergent method in [1], we construct the methods without using the iteration function and prove their convergence and superlinearly convergence under some weaker conditions.

We assume that function f has LC gradient, *i.e.*, there exists an $L > 0$ such that, for any $x, y \in R^n$,

$$\|g(x) - g(y)\| \leq L\|x - y\|. \tag{6}$$

If $g(x)$ is Lipschitzian then Rademacher's theorem implies that $\nabla g(x) = \nabla^2 f(x)$ exists almost everywhere and we can define

$$\partial_B^2 f(x) = \{\lim_{x_k \to x} \nabla^2 f(x_k)\}, \tag{7}$$

where the limit is taken for the x_k at which $f(x_k)$ is twice differentiable. Similar to Clarke's generalized Jacobian we define

$$\partial^2 f(x) = co\{\partial_B^2 f(x)\}. \tag{8}$$

We have (see [6])

(**R1**) $\partial^2 f(x)$ is nonempty, compact and bounded;

(**R2**) $\partial^2 f(x)$ is upper semicontinuous at x;

(**R3**) If $f(x)$ is uniformly convex at x_* and $g_* = g(x_*) = 0$, then x_* is a locally strict minimum point of $f(x)$.

We shall propose one type of globally convergent inexact generalized Newton methods for functions which have LC gradient. These methods are implementable. We shall prove that under some mild assumptions the algorithms are linearly convergent, and they are superlinearly convergent or even quadratically convergent for uniformly convex functions.

The paper is organized as follows. In Section 2, we propose the inexact generalized Newton algorithm which makes the norms of the gradients decreasing, and prove global convergence and superlinear convergence of the algorithm. We also point out that the method can be also used to solve the nonsmooth equations. In Section 3, we discuss the main assumption of the paper and show that the functions with semismooth gradients or C-differentiable gradients must meet the assumption. Section 4 lists some applications for the algorithm and gives some numerical results.

Throughout this paper the vector norms are Euclidean.

2. Inexact Generalized Newton Algorithm with Decreasing Gradient Norms

Algorithm 2.1. *Given an initial guess x_0, $\beta_0 \in (0,1)$, $\tau \in (0,1)$, $\theta \in (0,1)$ and a sequence of numbers $\{\bar\eta_k\}$, $0 < \bar\eta_k < 1$, we compute a sequence of steps $\{s_k\}$ and iterates $\{x_k\}$ as follows:*
For $k = 0$ step 1 until convergence do
1. *Choose a $V_k \in \partial^2 f(x_k)$, find an $\bar s_k$ such that $\|g_k + V_k \bar s_k\| \le \bar\eta_k \|g_k\|$.*
2. *If the following conditions (9) and (10) hold for $s_k = \bar s_k$ and $\eta_k = \bar\eta_k$ then let $s_k = \bar s_k$ and $\eta_k = \bar\eta_k$, otherwise let $\beta_k = min\{1,\ \|g_k\|\}\beta_0$, $\hat\eta_k = \|g_k + V_k \bar s_k\|/\|g_k\|$, $\eta_k = 1 - \tau^i(1-\hat\eta_k)$ and $s_k = [-\beta_k(1-\eta_k)^{1/2} g_k + (1-\eta_k)\bar s_k]/(1-\hat\eta_k)$, where i is the smallest non-negative integer such that*

$$\|r_k\| = \|g_k + V_k s_k\| \le \eta_k \|g_k\| \tag{9}$$

and

$$\|g(x_k + s_k)\| \le [1 - \theta(1-\eta_k)]\|g_k\|. \tag{10}$$

3. *Set $x_{k+1} = x_k + s_k$.*

Theorem 2.1. *Assume that Algorithm 2.1 is implemented with $\bar\eta_k < 1$. If $s_k = \bar s_k$, $\eta_k = \bar\eta_k$ and $\sum_{k=1}^{\infty}(1-\eta_k)$ is divergent then $lim_{k\to\infty} g_k = 0$.*

Proof. By Algorithm 2.1, $s_k = \bar s_k$ and $\eta_k = \bar\eta_k$ imply

$$\begin{aligned}\|g_{k+1}\| &\le [1-\theta(1-\eta_k)]\|g_k\| \\ &\le \|g_0\| \prod_{j=1}^{k}[1-\theta(1-\eta_j)] \\ &\le \|g_0\| \ \ exp\left[-\theta\sum_{j=1}^{k}(1-\eta_j)\right].\end{aligned}$$

Since $\theta > 0$ and $1 - \eta_j \ge 0$, the divergence of $\sum_{j=0}^{\infty}(1-\eta_j)$ implies that $\lim_{k\to\infty} g_k = 0$.

Assumption A1. Assume function f has LC gradient. We say that f satisfies A1 at x if for any $y \in R^n$ and any $V_y \in \partial^2 f(y)$, the following holds.

$$g(y) - g(x) = V_y(y-x) + o(\|y-x\|).$$

We say that f satisfies A1 at x with degree ρ if f has LC gradient and the following holds,

$$g(y) - g(x) = V_y(y-x) + O(\|y-x\|^{\rho}).$$

We say that f satisfies A1 (with degree ρ) if f satisfies A1(with degree ρ) at any $x \in R^n$.

Theorem 2.2. *Assume that Algorithm 2.1 is implemented with $\bar\eta_k \to 0$ to generate a sequence $\{x_k\}$, and $\|V_k\| \le M$ and $\|V_k^{-1}\| \le M$, $M > 1$, for all k. If $s_k = \bar s_k$, $\eta_k = \bar\eta_k$ and x_* is an accumulation point of $\{x_k\}$ and f satisfies A1 at x_*, then x_k converges to x_* superlinearly.*

Proof. Theorem 2.1 implies that $\lim_{k\to\infty}\|g_k\| = 0$ and $\|g_*\| = 0$. Given any $C > 1$, $\eta_k \to 0$ implies that there exists K_C such that for any $k > K_C$, $\eta_k \le (3CM^2)^{-1}$. Because $f(x)$ satisfies A1, there exists a neighborhood $B_\delta(x_*)$ of x_* such that for any $x \in B_\delta(x_*)$ and $V_x \in \partial^2 f(x)$,

$$\|g(x) - V_x(x - x_*)\| \le \|x - x_*\|/(2CM). \tag{11}$$

Since x_* is an accumulation point of $\{x_k\}$, there exists $K = K(\delta,\ C) > K_C$ such that $x_K \in B_\delta(x_*)$. We have

$$\|x_K + s_K - x_*\| \le \|V_K^{-1}\|\|V_K(x_K - x_*) + V_K s_K\|$$

$$
\begin{aligned}
&\leq M[\|V_K(x_K - x_*) - g_K\| + \|r_K\|] \\
&\leq M[(1+\eta_K)\|V_K(x_K - x_*) - g_K\| + \eta_K\|V_K(x_K - x_*)\|] \\
&\leq [(1+1/3)/2 + 1/3]\|x_K - x_*\|/C = \|x_K - x_*\|/C, \qquad (12)
\end{aligned}
$$

that is $x_K + s_K \in B_\delta(x_*)$. So we have proved that, for all $k > K(\delta, C)$, $x_k \in B_\delta(x_*)$ and $\|x_{k+1} - x_*\|/\|x_k - x_*\| \leq 1/C$. Since this is true for any large number C, we know that this theorem holds.

Lemma 2.1. *Given x with $g(x) \neq 0$, $V_x \in \partial^2 f(x)$ and $\theta \in (0,1)$. Suppose f is uniformly convex and satisfies A1 at x. If there exists an $\bar{s}$ such that $\|g(x) + V_x\bar{s}\| \leq \bar{\eta}\|g(x)\|$ for an $\bar{\eta} \in (0,\ 1)$, then there exists a $\lambda \in (0,\ 1)$ such that for any $\eta \in (\lambda,\ 1)$, the vector $s = [-\beta(1-\eta)^{1/2}g(x) + (1-\eta)\bar{s}]/(1-\hat{\eta})$ satisfies*

$$\|g(x) + V_x s\| \leq \eta\|g(x)\| \quad \textit{and} \quad \|g(x+s)\| \leq [1 - \theta(1-\eta)]\|g(x)\|, \qquad (13)$$

and in the above expression of s, $\beta = min\{1, \|g(x)\|\}\beta_0$ and

$$\hat{\eta} = \frac{\|g(x) + V_x\bar{s}\|}{\|g(x)\|}. \qquad (14)$$

Proof. Set

$$\varepsilon = \frac{\beta(1-\theta)(1-\hat{\eta})\|g(x)\|}{4M(\|g(x)\| + \|\bar{s}\|)}. \qquad (15)$$

Because f is uniformly convex at x, there exist $M > 1$ and $\delta > 0$ such that for any $y \in B_\delta(x)$ and any $V_y \in \partial^2 f(y)$, $\|V_y\| \leq M$, $\|V_y^{-1}\| \leq M$ and V_y is a positive definite matrix by convexity. $g(x) \neq 0$ implies $\bar{s} \neq 0$. As function f satisfies A1 at x, we can choose small enough δ such that for any $\|d\| \leq \delta$ and any $V_{x+d} \in \partial^2 f(x+d)$,

$$\|g(x+d) - g(x) - V_{x+d}d\| \leq \varepsilon\|d\|, \qquad (16)$$

where ε is defined in (15). Let

$$\lambda = max\left\{\frac{1+\hat{\eta}}{2},\ 1 - \left[\frac{(1-\hat{\eta})\delta}{\|g(x)\| + \|\bar{s}\|}\right]^2,\ 1 - \left[\frac{\beta(1-\hat{\eta})}{8M^3}\frac{\|\bar{s}\|}{\|g(x)\| + \|\bar{s}\|}\right]^2\right\} \qquad (17)$$

and $s = [-\beta(1-\eta)^{1/2}g(x) + (1-\eta)\bar{s}]/(1-\hat{\eta})$ for arbitrary $\eta \in (\lambda, 1)$. (17) implies $(1-\hat{\eta}) \geq (\eta - \hat{\eta}) \geq (1-\hat{\eta})/2$. For any $y \in B_\delta(x)$ and $V_y \in \partial^2 f(y)$, as V_y is a positive definite matrix, $g(x)^T V_y V_y g(x) \leq M g(x)^T V_y g(x)$ and $g(x)^T V_y g(x) \geq \|g(x)\|^2/M$. (17) implies

$$(1-\eta)^{1/2} < (1-\lambda)^{1/2} \leq \beta(1-\hat{\eta})/(8M^3) \leq (1-\hat{\eta})/(2\beta M), \qquad (18)$$

and hence we have

$$
\begin{aligned}
&\left\|g(x) - \frac{\beta(1-\eta)^{1/2}V_y g(x)}{1-\hat{\eta}}\right\|^2 \\
= \ &\|g(x)\|^2 - \frac{2\beta(1-\eta)^{1/2}g(x)^T V_y g(x)}{1-\hat{\eta}} + \frac{\beta^2(1-\eta)g(x)^T V_y V_y g(x)}{(1-\hat{\eta})^2}
\end{aligned}
$$

$$\begin{aligned}
&\le \|g(x)\|^2 - \frac{\beta(1-\eta)g(x)^T V_y g(x)}{1-\hat{\eta}}\left[2(1-\eta)^{-1/2} - \frac{\beta M}{1-\hat{\eta}}\right] \\
&\le \|g(x)\|^2 - \frac{\beta(1-\eta)g(x)^T V_y g(x)}{1-\hat{\eta}}\left[2(1-\eta)^{-1/2} - (1-\eta)^{-1/2}/2\right] \\
&\le \|g(x)\|^2\left[1 - \frac{\beta(1-\eta)^{1/2}}{2M(1-\hat{\eta})}\right] \le \|g(x)\|^2\left[1 - \frac{\beta(1-\eta)^{1/2}}{4M(1-\hat{\eta})}\right]^2. \qquad (19)
\end{aligned}$$

(17) also implies

$$\begin{aligned}
\|s\| &= \frac{\left\|-\beta(1-\eta)^{1/2}g(x) + (1-\eta)\bar{s}\right\|}{1-\hat{\eta}} \\
&< \frac{(1-\eta)^{1/2}(\|g(x)\| + \|\bar{s}\|)}{1-\hat{\eta}} \qquad (20) \\
&\le \delta.
\end{aligned}$$

So, $x+s \in B_\delta(x)$. We have

$$\begin{aligned}
\|\bar{s}\| &\le \|V_x^{-1}\|\|g(x) + V_x\bar{s} - g(x)\| \le M(\|g(x) + V_x\bar{s}\| + \|g(x)\|) \\
&\le M(1+\bar{\eta})\|g(x)\| \le 2M\|g(x)\|. \qquad (21)
\end{aligned}$$

(17), (19), (21) and the fact $(1-\eta)^{1/2} \le \beta/(8M^3)$ (see (21)) mean

$$\begin{aligned}
&\|g(x) + V_y s\| \\
&= \left\|g(x) + \frac{-\beta(1-\eta)^{1/2}V_y g(x) + (1-\eta)V_y\bar{s}}{1-\hat{\eta}}\right\| \\
&\le \|g(x)\|\left[1 - \frac{\beta(1-\eta)^{1/2}}{4M(1-\hat{\eta})}\right] + \frac{2(1-\eta)M^2\|g(x)\|}{1-\hat{\eta}} \\
&\le \|g(x)\|\left[1 - \frac{\beta(1-\eta)^{1/2}}{2M(1-\hat{\eta})}\right] \qquad (22) \\
&\le \eta\|g(x)\| + \|g(x)\|\left[\frac{\beta(1-\eta)^{1/2}}{8M} - \frac{\beta(1-\eta)^{1/2}}{2M(1-\hat{\eta})}\right] \\
&\le \|g(x)\|\left[\eta - \frac{\beta(1-\eta)^{1/2}}{4M(1-\hat{\eta})}\right]. \qquad (23)
\end{aligned}$$

Let $y = x$, (23) implies that the first inequality of (13) holds, *i.e.*, $\|g(x) + V_x s\| \le \eta\|g(x)\|$. Furthermore, let $y = x+s$, from (22), we obtain

$$\|g(x) + V_{x+s}s\| \le \|g(x)\|\left[1 - \frac{\beta(1-\eta)^{1/2}}{4M(1-\hat{\eta})}\right]. \qquad (24)$$

(16), (15), (21), (24) and $\beta(1-\eta)^{1/2} \ge 4M(1-\eta)$ (see (18)) imply

$$\|g(x+s)\| \le \|g(x+s) - g(x) - V_{x+s}s\| + \|g(x) + V_{x+s}s\|$$

$$
\begin{aligned}
&\leq \varepsilon\|s\| + \|g(x) + V_{x+s}s\| \\
&\leq \frac{\beta(1-\theta)(1-\eta)^{1/2}\|g(x)\|}{4M} + \|g(x)\|\left[1 - \frac{\beta(1-\eta)^{1/2}}{4M(1-\hat{\eta})}\right] \\
&\leq \left[1 - \frac{\beta\theta(1-\eta)^{1/2}}{4M}\right]\|g(x)\| \\
&\leq [1-\theta(1-\eta)]\|g(x)\|. \qquad (25)
\end{aligned}
$$

Therefore, the second inequality in (13) is also true.

If for some x_k, $\theta \in (0,1)$ and $V_k \in \partial^2 f(x_k)$, we can use an algorithm to obtain η_k and s_k satisfying (9) and (10), then we say that the algorithm can be implemented for (x_k, θ, V_k). Lemma 2.1 implies that if $f(x)$ is uniformly convex at x_k, then Algorithm 2.1 can be implemented for (x_k, θ, V_k) with any $V_k \in \partial^2 f(x_k)$.

We now prove global convergence and superlinear convergence for the algorithm.

Theorem 2.3. *Suppose f is uniformly convex. Let Algorithm 2.1 be implemented with $\bar{\eta}_k \leq \eta_{max} < 1$ to generate a sequence $\{x_k\}$. If x_* is an accumulation point of $\{x_k\}$ and f satisfies A1 at x_*, then $g_* = g(x_*) = 0$.*

Proof. Assume that $g_* = 0$ is not true. Because x_* is an accumulation point of $\{x_k\}$, there exists a subsequence $\{x_{k(i)}\}$ which converges to x_*. We can choose $k(i)$ properly such that $\bar{s}_{k(i)} \to \bar{s}_*$, $\beta_{k(i)} \to \beta_*$ and $V_{k(i)} \to V_*$, where $V_{k(i)} \in \partial^2 f(x_{k(i)})$ and $\|g_{k(i)}\| \geq C_1 > 0$ for some C_1. The upper semicontinuity of $\partial^2 f(x)$ implies $V_* \in \partial^2 f(x_*)$. Let $\hat{\eta}_{k(i)} = \|g_{k(i)} + V_{k(i)}\bar{s}_{k(i)}\|/\|g_{k(i)}\|$ and $\hat{\eta}_* = \|g_* + V_*\bar{s}_*\|/\|g_*\|$, it is clear that $\lim_{i\to\infty}\hat{\eta}_{k(i)} = \hat{\eta}_*$. $g_* \neq 0$ and $g_{k(i)} \neq 0$ imply $\bar{s}_* \neq 0$ and $\bar{s}_{k(i)} \neq 0$.

As $f(x)$ is uniformly convex at x_*, there exist $M > 1$ and $\delta_0 > 0$ such that for any $y \in B_{\delta_0}(x_*)$ and $V_y \in \partial^2 f(y)$, $\|V_y\| \leq M$, $\|V_y^{-1}\| \leq M$. Taking any $\theta_1 > \theta$, by Lemma 2.1, $g_* \neq 0$ implies that Algorithm 2.1 can be implemented for (x_*, θ_1, V_*). (23) (in which let $y = x = x_*$) and (25) imply that there exists $\lambda \geq (1+\hat{\eta}_*)/2$ (see (17)) such that for any $\eta \in (\lambda,\ 1-\tau(1-\lambda)]$ and corresponding $s_*(\eta) = [-\beta_*(1-\eta)^{1/2}g_* + (1-\eta)\bar{s}_*]/(1-\hat{\eta}_*)$, where $\beta_* = min\{1,\ \|g_*\|\}\beta_0$, the following results hold

$$
\begin{aligned}
&\|g_* + V_*s_*(\eta)\| \\
&\leq \eta\|g_*\| - \frac{\beta_*(1-\eta)^{1/2}}{4M(1-\hat{\eta}_*)}\|g_*\| \leq \eta\|g_*\| - \varepsilon_1 \qquad (26)
\end{aligned}
$$

and

$$
\begin{aligned}
&\|g(x_* + s_*(\eta))\| \leq [1-\theta_1(1-\eta)]\|g_*\| \\
&\leq [1-\theta(1-\eta)]\|g_*\| - \varepsilon_1, \qquad (27)
\end{aligned}
$$

where $\varepsilon_1 > 0$ is a constant defined as

$$
\varepsilon_1 = min\left\{\frac{\beta_*\tau(1-\lambda)^{1/2}}{4M(1-\hat{\eta}_*)}\|g_*\|,\ \tau(\theta_1-\theta)(1-\lambda)\|g_*\|\right\}.
$$

We now prove that when $i \to \infty$, $s_{k(i)}(\eta)$ is uniformly convergent to $s_*(\eta)$ for all $\eta \in (\lambda,\ \ 1-\tau(1-\lambda)]$.

It is easy to know that $|\beta_{k(i)} - \beta_*| \le |\|g_{k(i)}\| - \|g_*\||\beta_0 \le \|g_{k(i)} - g_*\|$. Because $\hat{\eta}_{k(i)} \to \hat{\eta}_*$, there exists a K_0 such that $\lambda \ge (1+\hat{\eta}_*)/2 \ge \hat{\eta}_{k(i)}$ for all $k(i) \ge K_0$. By definition, $s_{k(i)}(\eta) = [-\beta_{k(i)}(1-\eta)^{1/2} g_{k(i)} + (1-\eta)\bar{s}_{k(i)}]/(1-\hat{\eta}_{k(i)})$, we have

$$\begin{aligned}
&\|s_{k(i)}(\eta) - s_*(\eta)\| \\
= \;& \left\| \frac{-\beta_{k(i)}(1-\eta)^{1/2} g_{k(i)} + (1-\eta)\bar{s}_{k(i)}}{1-\hat{\eta}_{k(i)}} - \frac{-\beta_*(1-\eta)^{1/2} g_* + (1-\eta)\bar{s}_*}{1-\hat{\eta}_*} \right\| \\
\le \;& \left\| \frac{-\beta_{k(i)}(1-\eta)^{1/2} g_{k(i)} + (1-\eta)\bar{s}_{k(i)}}{1-\hat{\eta}_{k(i)}} - \frac{-\beta_*(1-\eta)^{1/2} g_* + (1-\eta)\bar{s}_*}{1-\hat{\eta}_{k(i)}} \right\| \\
+ \;& \| -\beta_*(1-\eta)^{1/2} g_* + (1-\eta)\bar{s}_*\| \left| \frac{1}{1-\hat{\eta}_{k(i)}} - \frac{1}{1-\hat{\eta}_*} \right| \\
\le \;& \frac{|\beta_{k(i)} - \beta_*|\|g_*\| + \beta_{k(i)}\|g_{k(i)} - g_*\| + \|\bar{s}_{k(i)} - \bar{s}_*\|}{1-\hat{\eta}_{k(i)}} \\
+ \;& \frac{\| -\beta_*(1-\eta)^{1/2} g_* + (1-\eta)\bar{s}_*\| \; |\hat{\eta}_{k(i)} - \hat{\eta}_*|}{(1-\hat{\eta}_{k(i)})(1-\hat{\eta}_*)} \\
\le \;& \frac{(\|g_*\|+1)\|g_{k(i)} - g_*\| + \|\bar{s}_{k(i)} - \bar{s}_*\|}{(1-\lambda)} + \frac{(\|g_*\| + \|\bar{s}_*\|)|\hat{\eta}_{k(i)} - \hat{\eta}_*|}{(1-\lambda)(1-\hat{\eta}_*)}. \qquad (28)
\end{aligned}$$

So, when $i \to \infty$, $s_{k(i)}(\eta)$ uniformly converges to $s_k(\eta)$ for all $\eta \in (\lambda, \;\; 1-\tau(1-\lambda)]$.

Based on this fact, there exists $K_1 > K_0$ such that for all $k(i) \ge K_1$, $\|g_{k(i)} - g_*\| \le \varepsilon_1/4$, $\|s_{k(i)}(\eta) - s_*(\eta)\| \le \varepsilon_1/(4M)$, $\|V_{k(i)} - V_*\|\|s_*(\eta)\| \le \|V_{k(i)} - V_*\|(\|\bar{s}_*\| + \|g_*\|)/(1-\hat{\eta}_*) \le \varepsilon_1/4$ and $\|g(x_{k(i)} + s_{k(i)}(\eta)) - g(x_* + s_*(\eta))\| \le \varepsilon_1/2$ for $\eta \in (\lambda, \;\; 1-\tau(1-\lambda)]$. For such $k(i)$ and η we have

$$\begin{aligned}
&\|g_{k(i)} + V_{k(i)} s_{k(i)}(\eta)\| \\
\le \;& \|g_{k(i)} - g_*\| + \|V_{k(i)} s_{k(i)}(\eta) - V_* s_*(\eta)\| + \|g_* + V_* s_*(\eta)\| \\
\le \;& \varepsilon_1/4 + \varepsilon_1/2 + \eta\|g_*\| - \varepsilon_1 \\
\le \;& \eta\|g_{k(i)}\|, \qquad (29)
\end{aligned}$$

and

$$\begin{aligned}
&\|g(x_{k(i)} + s_{k(i)}(\eta))\| \\
\le \;& \|g(x_{k(i)} + s_{k(i)}(\eta)) - g(x_* + s_*(\eta))\| + \|g(x_* + s_*(\eta))\| \\
\le \;& \varepsilon_1/2 + [1-\theta(1-\eta)]\|g_*\| - \varepsilon_1 \\
\le \;& [1-\theta(1-\eta)]\|g_{k(i)}\|. \qquad (30)
\end{aligned}$$

On the other hand, we can prove that for any $k(i) > K_1$, there must be a j such that $1-\tau^{j+1}(1-\hat{\eta}_k) \in (\lambda, \;\; 1-\tau(1-\lambda)]$. In fact, as we have seen, $\hat{\eta}_{k(i)} \le \lambda$ for all $k(i) \ge K_1$ and hence there must be a j satisfying $1-\tau^j(1-\hat{\eta}_{k(i)}) \le \lambda < 1-\tau^{j+1}(1-\hat{\eta}_{k(i)})$, then $\tau^j(1-\hat{\eta}_{k(i)}) \ge 1-\lambda > \tau^{j+1}(1-\hat{\eta}_{k(i)})$ and $1-\tau^{j+1}(1-\hat{\eta}_{k(i)}) \le 1-\tau(1-\lambda)$, *i.e.*, $1-\tau^{j+1}(1-\hat{\eta}_{k(i)}) \in (\lambda,\ 1-\tau(1-\lambda)]$.

We are now ready to derive a contradiction. In fact the above facts mean that for any $k(i) \ge K_1$, when we use Algorithm 2.1, we will find $\eta_{k(i)}$ and $s_{k(i)}$ to satisfy conditions (9) and (10) with the property that $\eta_{k(i)} \le 1-\tau(1-\lambda)$. Therefore, we have $\|g_{k(i)+1}\| \le [1-\theta\tau(1-\lambda)]\|g_{k(i)}\|$ and

$$\|g_{k(i)}\| - \|g_{k(i)+1}\| \ge \theta\tau(1-\lambda)\|g_{k(i)}\| \ge \theta\tau(1-\lambda)C_1$$

for $k(i) \ge K_1$. As $\|g_k\|$ is nonincreasing, the above result implies $\|g_k\| \to -\infty$ which is impossible. So, this theorem has been proved.

Theorem 2.4. *Suppose f is uniformly convex. Let Algorithm 2.1 be implemented with $\bar{\eta}_k \le \eta_{max} < 1$ and $\bar{\eta}_k \to 0$ to generate a sequence $\{x_k\}$. If x_* is an accumulation point of $\{x_k\}$ and f satisfies A1 at x_*, then x_k converges to x_* superlinearly.*

Proof. By Theorem 2.3 we know that $g_* = 0$. According to the proof of Theorem 2.2, it will suffice if we can show that for sufficiently large i Algorithm 2.1 will choose $\eta_{k(i)} = \bar{\eta}_{k(i)}$ and $s_{k(i)} = \bar{s}_{k(i)}$, *i.e.*, $\bar{\eta}_{k(i)}$ and $\bar{s}_{k(i)}$ satisfy (13).

As f satisfies A1 at x_*, there exist a Lipschitz constant L and a $\delta > 0$ such that $\|g(x)\| \le L\|x - x_*\|$ for all $x \in B_\delta(x_*)$. Let $C = \max\{1,\ 2LM(1-\theta)^{-1}\}$. From (16) we may reduce δ, if necessary, so that for any $\|d\| \le \delta$ and $V_{x_*+d} \in \partial^2 f(x_*+d)$, $\|g(x_*+d) - V_{x_*+d}d\| \le \|d\|/(2CM^2)$.

Now we consider sufficiently large $k(i)$ such that $x_{k(i)} \in B_\delta(x_*)$ and $\hat{\eta}_{k(i)} \le \bar{\eta}_{k(i)} \le (3CM^2)^{-1}$.

$$\begin{aligned} \|g_{k(i)}\| &\ge \|V_{k(i)}(x_{k(i)} - x_*)\| - \|V_{k(i)}(x_{k(i)} - x_*) - g_{k(i)}\| \\ &\ge \|x_{k(i)} - x_*\|/(2M). \end{aligned} \tag{31}$$

Following the proof of (12), we can obtain

$$\begin{aligned} \|x_{k(i)} + \bar{s}_{k(i)} - x_*\| &\le M[\|V_{k(i)}(x_{k(i)} - x_*) - g_{k(i)}\| + \hat{\eta}_{k(i)}\|g_{k(i)}\|] \\ &\le \|x_{k(i)} - x_*\|/C. \end{aligned} \tag{32}$$

(31) and (32) imply that $x_{k(i)} + \bar{s}_{k(i)} \in B_\delta(x_*)$ and

$$\begin{aligned} &\|g(x_{k(i)} + \bar{s}_{k(i)})\| \le L\|x_{k(i)} + \bar{s}_{k(i)} - x_*\| \\ \le\ & L\|x_{k(i)} - x_*\|/C \le (1-\theta)\|g_{k(i)}\| \\ \le\ & (1 - \theta(1 - \bar{\eta}_{k(i)}))\|g_{k(i)}\|. \end{aligned} \tag{33}$$

So, $\bar{\eta}_{k(i)}$ and $\bar{s}_{k(i)}$ meet condition (13), *i.e.*, $\bar{\eta}_{k(i)} = \eta_{k(i)}$ and $\bar{s}_{k(i)} = s_{k(i)}$. The theorem holds .

3. Discussion for Assumption A1

In the previous section, the assumption A1 has been used. What types of functions can meet this assumption? In this section, we shall show that the following two classes of functions satisfy A1.

1. $f(x)$ has semismooth gradient $g(x)$, *i.e.*,

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

exists for any $h \in R^n$, where $\partial g(x + th')$ is the generalized Jacobian of g at $x + th'$ in the sense of Clarke (1990).

Let $g'(x;h)$ denote the directional derivative. The following result is contained in Lemma 2.2 of [8].

Lemma 3.1. *Suppose that $g: R^n \to R^n$ is directionally differentiable at a neighborhood of x. The following statements are equivalent:*

(1) *g is semismooth at x;*

(2) *$g'(.;.)$ is semicontinuous at x;*

(3) *for any $V \in \partial g(x+h)$ and $h \to 0$,*

$$Vh - g'(x;h) = o(\|h\|). \tag{34}$$

According to this lemma, it is clear that if $g(x)$ is semismooth, then $f(x)$ satisfies A1.

2. Second order C-differentiable function. Qi [11] gave:

Definition 3.1. *Suppose that $T: R^n \to R^{m\times n}$ is a set-valued operator, i.e., for any $x \in R^n$, any $V \in T(x)$ is an $m \times n$ matrix. We say that the function F: $R^n \to R^m$ is C-differentiable at $x \in R^n$ if*

(1) *$T(y)$ is nonempty and compact for any y in a neighborhood of x;*

(2) *T is upper semicontinuous at x;*

(3) *for any $V \in T(x+d)$,*

$$F(x+d) = F(x) + Vd + o(\|d\|).$$

In this case, T is called a C-differential operator of F.

Pu and Zhang [10] extended the above conception to second order C-differentiability, they gave.

Definition 3.2. *A first order differentiable function $f: R^n \to R$ is said second order C-differentiable at x with a second order C-differential operator (or called C2 operator for short) T if the gradient g of f is C-differentiable at x with T. Furthermore, we say that f is second order C-differentiable at x with a C2 operator T and degree ρ, $\rho > 1$, if f is second order C-differentiable at x with T and for any $V \in T(x+d)$,*

$$g(x+d) = g(x) + Vd + O(\|d\|^\rho).$$

We say that f is second order C-differentiable in D with a C2 operator T (and degree ρ) if f is second order C-differentiable at each point $x \in D$ with T (and degree ρ).

It is clear that if f is second order C-differentiable with C2 operator $\partial^2 f$, then f satisfies A1.

Applications and Numerical Tests

The algorithms proposed in Chapter 2 and this chapter can be applied to the following problems.

1. For the constrained optimization problems:

$$\min \quad f_1(x), \quad x \in R^n$$

$$s.t. \quad g_i(x) \le 0, \quad i = 1, 2, ..., n \tag{35}$$

where f_1 and g_i are twice differentiable, we can define an unconstrained optimization problem as following:

$$\min \ f(x) \overset{def}{=} f_1(x) + \mu \sum_{i=1}^{m} (\max\{0, \ -g_i(x)\})^2/2, \tag{36}$$

where f is a penalty function. We know that under some condition problems (35) and (36) are equivalent. f may not be twice differentiable at the x with $g_i(x) = 0$ for some i, but has LC gradient at these points. We can use the algorithms to solve such problems.

2. The methods may be used to solve the nonsmooth equations which arise from reformulation of nonlinear complementarity problems, variational inequalities and some optimization problems (see [9] and [11]).

We have conducted some numerical tests for using the inexact generalized Newton methods to solve unconstrained optimization problems in which the objective functions are not twice differentiable, but have LC gradients. The results of these tests show that the inexact generalized Newton methods are effective. Some examples are as follows.

The objective functions are:

$$f(x) = f_1(x) + \mu \sum_{i=1}^{m} (\max\{0, \ -g_i(x)\})^2/2, \tag{37}$$

and $f_1(x)$ and $g_j(x)$ are defined in following problems.

Problem 1. (problem 227, [12])

$$f_1(x) = (x_1 - 2)^2 + (x_2 - 1)^2,$$

$$g_1(x) = x_1^2 - x_2,$$
$$g_2(x) = x_2^2 - x_1.$$

Problem 2. (problem 215, [12])

$$f_1(x) = x_2,$$
$$g_1(x) = -(x_2 - x_1^2).$$

Problem 3. (problem 232, [12])

$$f_1(x) = -(9 - (x_1 - 3)^2)x_2^3/(27\sqrt{3}),$$

$$g_1(x) = -(x_1/\sqrt{3} - x_2),$$
$$g_2(x) = -(x_1 + \sqrt{3}x_2),$$
$$g_3(x) = -(6 - x_1 - \sqrt{3}x_2).$$

Problem 4. (problem 250, [12])

$$f_1(x) = -x_1x_2x_3,$$

$$g_1(x) = -(x_1 + 2x_2 + 2x_3),$$
$$g_2(x) = -(72 - x_1 - 2x_2 - 2x_3).$$

Problem 5. (problem 264, [12])

$$f_1(x) = x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5(x_1 + x_2) - 21x_3 + 7x_4,$$
$$g_1(x) = -(-x_1^2 - x_2^2 - x_3^2 - x_4^2 - x_1 + x_2 + x_3 + x_4 + 8),$$
$$g_2 = -(-x_1^2 - 2x_2^2 - x_3^2 - 2x_4^2 + x_1 + x_4 + 9),$$
$$g_3 = -(-2x_1^2 - x_2^2 - x_3^2 - 2x_1 + x_2 + x_4 + 5).$$

We calculated the above problems by two methods:

M1 (Method 1). The inexact generalized Newton methods without using any stabilization technique, *i.e.*, Given an initial guess x_0 and a $\eta \in (0,1)$, we compute a sequence of steps $\{s_k\}$ and iterates $\{x_k\}$ as follows:

For $k=0$ step 1 until convergence do:

Obtain s_k such that $r_k = V_k s_k + g_k$, where $\|r_k\|/\|g_k\| \leq \eta_k$, and $V_k \in \partial^2 f(x_k)$. let $x_{k+1} = x_k + s_k$ (see [1], [11] and [10]).

M2 (Method 2). The globally convergent inexact generalized Newton methods with decreasing gradient norms, see Section 2.

The termination criterion is $\|g\| \leq 10^{-5}$. The parameters are chosen as: $\eta_k = (10k)^{-1}$; in method 2, $\tau = 0.7$, $\theta = 0.3$ and $\beta_0 = 0.00001$;

In the "NIT/NG" entry of the table below,

NIT=the number of iterations,

NG=the number of gradient evaluations.

μ is a parameter given in (37).

The data in Table 1 show that the inexact generalized Newton methods can be used to solve non-twice differentiable optimization problems effectively, and the stabilization technique, reducing gradient norms, can be used to improve convergence behavior.

			M1	M2		M1	M2
problem	μ	Initial points	NIT/ NG	NIT/ NG	Initial points	NIT/ NG	NIT/ NG
1	20	0.5, 0.5	9/9	9/201	1, 1	9/9	7/27
1	20	10, 10	F	13/31	-10, -10	F	12/21
1	20	10, -10	F	11/20	2, 2	6/6	11/20
1	100	0.5, 0.5	11/11	14/36	1, 1	11/11	12/32
1	100	6,6	F	16/29	-6, -6	F	20/34
2	20	1, 1	6/6	15/28	2, 2	6/6	18/34
2	20	5, 5	F	17/46	8, 8	F	16/44
2	40	5, 5	F	12/22	15,15	F	15/27
3	20	2, 0.5	5/5	4/6	4, 1	6/6	7/13
3	20	4, 2	8/8	9/16	6 1.5	6/6	8/14
3	100	2, 0.5	5/5	5/6	4, 1	6/6	7/12
3	100	6, 1.5	5/5	10/13	10, 5	F	9/19
4	100	10, 10, 10	10/10	14/24	-10, -10, -10	11/11	11/20
4	100	15, 15, 15	5/5	9/16	-15, -15, -15	10/10	16/33
4	200	10, 10, 10	10/10	14/24	-10, -10, -10	12/12	13/30
4	200	15, 15, 15	5/5	7/11	-15, -15, -15	13/13	11/23
5	20	0, 0, 0, 0	14/14	12/18	0,0.5,1.5,-0.5	14/14	11/16
5	20	0,0.8,1.8,-0.8	15/15	11/15	0, 0, 0, 0	F	35/55
5	50	0,0.5,1.5,-0.5	F	20/30	0,0.8,1.8,-0.8	F	15/35

Table 1

References

[1] Martinez, J.M., Qi, L., Inexact Newton's method for solving nonsmooth equations, *Journal of Computer Applied Mathematics,* **60** (1995), 127-145.

[2] Dennis Jr.J.J., Moré, J.J., Quasi-Newton method, motivation and theory, *SIAM Reviews,* **19** (1997), 46-89.

[3] Dembo, R.S., Eisenstat, S.C., Steihaug, T., Inexact Newton's method, *SIAM Journal on Numerical Analysis,* **14** (1982), 400-408.

[4] Dembo, R.S., Steihaug, T., Truncated-Newton's method for large-scale unconstrained optimization, *Mathematical Programming,* **26** (1982), 190-212.

[5] Eisenstat, S.C., Walker, H.F., Globally convergent inexact Newton's method, *SIAM Journal on Optimization,* **4** (1994), 393-422.

[6] Clarke, F.H., Optimization and nonsmooth analysis, SIAM Philadelphia, 1990.

[7] Qi, L., Sun, J., A nonsmooth version of Newton's method, *Mathematical Programming,* **58,** 353-368.

[8] Qi, L., Convergence analysis of some method for solving nonsmooth equations, *Mathematics of Operations Research,* **18** (1993), 227-243.

[9] Pang, J.S., Qi, L., A globally convergent Newton's method for convex SC^1 minimization problems, *Journal of Optimization Theory Applications,* **85** (1995), 633-648.

[10] Pu, D., Zhang, J., An inexact generalized Newton's method for second order C-Differentiable optimization, *Journal of Computational and Applied Mathematics,* **93** (1998), 107-122.

[11] Qi, L., C-differential operators, C-differentiability and generalized Newton method, AMR 96/5, Applied Mathematics Report, University of New South Wales, 1996.

[12] Schittkowski, K., More test examples for nonlinear programming codes, Springer-Verlag, 1988.