Vol. 10 No. 6 PROGRESS IN NATURAL SCIENCE Jane 2000

RESEARCH NOTES

Gauss-Newton methods for a class of nonsmooth optimization problems *

LI Chong (李 冲)[1] and WANG Xinghua (王兴华)[2]

(1. Department of Applied Mathematics, Southeast University, Nanjing 210096, China; 2. Department of Mathematics, Hangzhou University, Hangzhou 310028, China)

Received July 22, 1999; revised October 25, 1999

Abstract The local quadratic convergence of the Gauss-Newton method for convex composite optimizations is established for any convex function with a minima set. This work extends Burke and Ferris' results when this minima set is a set of weak sharp minima for the convex function.

Keywords: Gauss-Newton method, weak sharp minima, regular point, quadratic convergence.

The famous Gauss-Newton method, which was proposed to find the least-squares solutions of nonlinear equations by Gauss in the early nineteenth century, is now extended to solve the following convex composite optimization:

$$(P)\ \min f(x) := h(F(x)),$$

where $h: R^m \to R$ is convex and $F: R^n \to R^m$ continuously differentiable. This problem has recently received a great deal of attention (see, for example, ref. [1] and the references therein), and is justifiable since many class of problems in optimization theory can be cast within its framework, e.g., convex inclusion, minimax problems, penalization methods and goal programming. Moreover, this method provides a unifying framework for the development and analysis of algorithmic solution techniques.

In 1985, Womersley[2] proved the local quadratic convergence of Gauss-Newton methods under the assumption of strong uniqueness, extending the work of Jittorntrum and Osborne[3] for the case where h is a norm. However, Womersley's assumption just ensures the Gauss-Newton sequence converges to a local minimum of (P). Recently, Burke and Ferris[1] have made a great progress in the study of convergence of Gauss-Newton methods. An important distinction from Womersley's work is that they do not require the minima set for h be a singleton or even a bounded set. Their research is based on two assumptions:

(i) the set of minima for the function h; denoted by C, is a weak sharp minima for h; that is, there is $\lambda > 0$ such that $h(y) \geqslant h_{\min} + \lambda d(y, C)$ holds $\forall y \in R^m$, where $h_{\min} = \min_y h(y)$ while

* Project supported by the National Natural Science foundation of China (Grant No. 19971013), Natural Science Foundation of Jiangsu Province (Grant No. BK99001) and partly by National Fundamental Research Project of China.

$d(y, C)$ denotes the distance from x to the set C;

(ii) there is a regular point $\bar{x} \in R^n$ for the inclusion

$$F(x) \in C; \tag{1}$$

that is,

$$\ker(F'(\bar{x})^T) \cap \Gamma_C(F(\bar{x})) = \{0\},$$

where the set-valued mapping $\Gamma_C : R^m \to R^m$ is given by

$$\Gamma_C(y) = (\mathrm{cone}(C - y))^0 = \{x^* \in R^m : \langle x^*, x\rangle \leqslant 1,\ \forall x \in \mathrm{cone}(C - y)\},\ \forall y \in R^m.$$

Under the above assumptions, they established the local quadratic convergence of the Gauss-Newton sequence. Based on the work of Burke and Ferris, we continue our investigation in this direction. It is unexpected that we find the local quadratic convergence of this method to be independent of the other properties of the convex function h. Our purpose is to relax the weak sharp minima assumption for h and establish the quadratic convergence of the Gauss-Newton method. In addition, we will propose a relaxation version of the Gauss-Newton method and give the superlinear convergence of this method.

1 Gauss-Newton method and its convergence

For $\Delta > 0$ and $x \in R^m$, let $D_\Delta(x)$ represent the set of solutions to the minimization problem

$$\min\{h(F(x) + F'(x)d) : \| d \| \leqslant \Delta\}. \tag{2}$$

Thus the basic algorithm considered in ref. [1] is as follows.

Algorithm 1. Let $\eta \geqslant 1$, $\Delta \in (0, +\infty]$ and $x^0 \in R^n$ be given. For $k = 1, 2, \cdots$, having x^k, we determine x^{k+1} as follows.
\ ;If $h(F(x^k)) = \min\{h(F(x^k) + F'(x^k)d) : \| d \| \leqslant \Delta\}$, then stop; otherwise, choose $d^k \in D_\Delta(x^k)$ to satisfy $\| d^k \| \leqslant \eta\, d(0, D_\Delta(x^k))$, and set $x^{k+1} = x^k + d^k$.

Let B denote the closed ball in R^m or R^n. It is helpful to recall the result due to Burke and Ferris in ref. [1] before giving our theorems.

Theorem BF[1]. *Let $\bar{x} \in R^n$ be a regular point of inclusion* (1) *where C is a set of weak sharp minima for h. Let* $0 < \bar{\delta} < \Delta$, *which exists from Proposition* 3.3 *of ref.* [1] *such that*

$$d(0, D_\Delta(x)) \leqslant \beta d(F(x), C) \text{ and } \{d \in R^n : \| d \| \leqslant \Delta,\ F(x) + F'(x)d \in C\} \neq \phi \tag{3}$$

hold on $\bar{x} + \bar{\delta}B$ for some β. Assume that F' is Lipschitz continuous on $\bar{x} + \bar{\delta}B$ with Lipschitz constant L and h is Lipschitz continuous on $F(\bar{x} + \bar{\delta}B) + (1/8)LB$ with Lipschitz constant M. If there is $\delta > 0$ such that

(i) $\delta < \min\{\bar{\delta}/2, 1\}$,

(ii) $d(F(\bar{x}), C) < \delta/(2\eta\beta)$ *and*

(iii) $\theta := \eta LM\delta\beta/\lambda < 1$,

then there is a neighborhood $M(\bar{x})$ *of* $\bar{x}$ *such that the sequence* $\{x^k\}$ *generated by Algorithm* 1 *with initial point in* $M(\bar{x})$ *converges at a quadratic rate to some* x^* *with* $F(x^*) \in C$, *then* x^* *is an optimal solution of* (P).

It should be noted that the weak sharp minima assumption for h is a very strong assumption. In fact, for any convex function f on R^m, the function $h(y) = (f(y) - f_{\min})^s$ on R^m with $s > 1$ does not have a set of weak sharp minima. Another important class for which there is no set of weak sharp minima is the class of Gateaux differentiable convex functions. Therefore, it is very interesting to relax the weak sharp minima assumption in Theorem BF. We have

Theorem 1. *Let* $\bar{x} \in R^n$ *be a regular point of inclusion* (1) *where* C *is a minima set for* h. *Let* $0 < \bar{\delta} < \Delta$ *such that* (3) *holds on* $\bar{x} + \bar{\delta}B$. *Assume that* F' *is Lipschitz continuous on* $\bar{x} + \bar{\delta}B$ *with Lipschitz constant* L. *If there is* $\delta > 0$ *such that*

(i) $\delta < \min\{\bar{\delta}/2, 1\}$,

(ii) $d(F(\bar{x}), C) < \delta/(2\eta\beta)$ *and*

(iii) $\eta L\delta\beta < 2$,

then there is a neighborhood $M(\bar{x})$ *of* $\bar{x}$ *such that the sequence* $\{x^k\}$ *generated by Algorithm* 1 *with initial point in* $M(\bar{x})$ *converges at a quadratic rate to some* x^* *with* $F(x^*) \in C$, *then* x^* *is an optimal solution of* (P).

2 Relaxation of the Gauss-Newton method and its convergence

Considering the background of the numerical computation, the following relaxation algorithm and the convergence result are of practical importance.

Let $D_\Delta^k(x^k)$ represent the set of all $d \in R^n$ satisfying $\| d \| \leqslant \Delta$ and

$$h(F(x^k) + F'(x^k)d) \leqslant \min\{h(F(x^k) + F'(x^k)d) : \| d \| \leqslant \Delta\} + \| d^{k-1} \|^\alpha.$$

Relaxation Algorithm 2. Let $\eta \geqslant 1$, $\alpha > 1$, $\Delta \in (0, +\infty]$, $x^0 \in R^n$ and $d^{-1} \in R^n$, $d^{-1} \neq 0$ be given. For $k = 1, 2, \cdots$, having x^k, we determine x^{k+1} as follows.

If $h(F(x^k)) \leqslant \min\{h(F(x^k) + F'(x^k)d) : \| d \| \leqslant \Delta\} + \| d^{k-1} \|^\alpha$, take $d^k = \| d^{k-1} \|^\alpha d^{k-1}$; otherwise, choose $d^k \in D_\Delta^k(x^k)$ to satisfy $\| d^k \| \leqslant \eta d(0, D_\Delta^k(x^k))$, and set $x^{k+1} = x^k + d^k$.

Theorem 2 *Let* $\bar{x} \in R^n$ *be a regular point of inclusion* (1) *where* C *is a set of weak sharp minima for* h. *Let* $0 < \bar{\delta} < \Delta$ *such that* (3) *holds on* $\bar{x} + \bar{\delta}B$. *Assume that* F' *is Lipschitz continuous on* $\bar{x}$

$+ \bar{\delta} B$ with Lipschitz constant L. If there is $\delta > 0$ such that

(i) $\delta < \min\{\bar{\delta}/(c+1), 1\}$,

(ii) $d(F(\bar{x}), C) < \delta/(2\eta\beta)$ *and*

(iii) $\eta L\delta\beta/2 + \eta\beta\delta^{\alpha-1}/\lambda < 1$,

where $p = \min\{2, \sqrt{\alpha}\}$, and $c = \sum_{i=0}^{+\infty}(1/2)^{p^i}$, *then there is a neighborhood* $M(\bar{x})$ *of* $\bar{x}$ *such that the sequence* $\{x^k\}$ *generated by Relaxation Algorithm* 2 *with initial point in* $M(\bar{x})$ *and* $\| d^{-1} \| \leqslant \delta/2$ *converges at a rate of* p *degree to some* x^* *with* $F(x^*) \in C$, *then* x^* *is an optimal solution of* (P).

Remark. Comparing Theorem 1 with Theorem 2, we note that Theorem 2 requires C be a set of weak sharp minima for h. It is not surprising, since in Relaxation Algorithm 2 the correction term d^k is an approximating solution to problem (2) such that the approximation error of the function value is controlled within a suitable bound. Thus it is closely related to the extent of the sharpness of h.

References

1 Burke, J. V., Ferris, M. C., A Gauss-Newton method for convex composite optimization, *Mathematical Programming*, 1995, 71(2): 179.

2 Womersley, R. S., Local properties of algorithms for minimizing nonsmooth composite functions, *Mathematical Programming*, 1985, 32(1): 69.

3 Jittorntrum, K., Osborne, M. R., Strong uniqueness and second order convergence in nonlinear discrete approximation, *Numerische Mathematik*, 1980, 34(3): 439.