Journal of Computational Mathematics, Vol.9, No.3, 1991, 222-228.

A GLOBALLY CONVERGENT ALGORITHM FOR A LOCALLY LIPSCHITZ FUNCTION*1)

Zhang Lian-sheng   Tian Wei-wen
(*Shanghai University of Science and Technology, Shanghai, China*)

Abstract

In this paper, an algorithm of global convergence is proposed for a locally Lipschitz function, which is strictly differentiable at almost all differentiable points, and several examples are computed on an IBM PC.

§1. Introduction

A great deal of effort has been devoted to nondifferentiable optimization in recent years. Many researches have gone into implementable algorithms mainly by Wolfe, Lemarechal, Zowe, Polak, Kiwiel, et al., besides various definitions of subgradient and corresponding first-order optimality conditions. However, in these algorithms the functions are required to be convex or semi-smooth in order to guarantee global convergence of the algorithms. It seems that there is no globally convergent algorithm for the locally Lipschitz function without additional condition. In this paper, we propose an algorithm of global convergence for a locally Lipschitz function, which is strictly differentiable at almost all differentiable points. In addition, we describe some concepts and properties of the locally Lipschitz function.

Definition 1.1. Let $f : R^n \to R$ be locally Lipschitz continuous. The generalized gradient of f at x is defined by

$$\partial f(x) = \mathrm{co}\{\lim_{v_i \to 0} \nabla f(x+v_i)\}$$

where $\nabla f(x)$ denotes the gradient of f at x, co denotes the convex hull of a set, and v_i are such that $\nabla f(x+v_i)$ exists and $\lim_{v_i \to 0} \nabla f(x+v_i)$ exists. We recall that a locally Lipschitz function $f(x), x \in R^n$, is differentiable almost everywhere.

Definition 1.2. Let $f : R^n \to R$ be locally Lipschitz continuous. The generalized directional derivative of f at x in the direction h is defined by

$$f^0(x;h) = \overline{\lim_{\substack{y\to 0\\ t\to 0}}} \frac{f(x+y+th)-f(x+y)}{t}.$$

Proposition. Let $f : R^n \to R$ be locally Lipschitz continuous. Then
1. $\partial f(x)$ exists and is compact at all $x \in R^n$.
2. $\partial f(x)$ is bounded on bounded sets.

* Received April 12, 1988.
1) The Project Supported by National Natural Science Foundation of China.

3. $\partial f(x)$ is upper semi-continuous (u.s.c.) in the sense that

$$\{x_i \to \hat{x}, y_i \in \partial f(x_i) \text{ and } y_i \to \hat{y}\} \Rightarrow \{\hat{y} \in \partial f(\hat{x})\}.$$

4. $f^0(x;h)$ exists for all $x, h \in R^n$, and

$$f^0(x;h) = \max_{\xi \in \partial f(x)} \langle \xi, h \rangle.$$

Definition 1.3. For any $\varepsilon > 0$, the ε-smeared generalized gradient is defined by

$$\partial_\varepsilon f(x) = \operatorname{co}\Big\{ \bigcup_{x' \in x + \varepsilon B(\theta,1)} \partial f(x') \}, \quad B(\theta,1) = \{y : \|y\| \le 1\Big\}.$$

$\partial_\varepsilon f(x)$ has properties 1–3 above.

§2. Several Lemmas

In this section, we demonstrate several lemmas concerning the algorithm.

Let $f : R^n \to R$ be locally Lipschitzian. For any $\varepsilon > 0$, we define

$$h_\varepsilon(x) = -\mathrm{Nr}(\partial_\varepsilon f(x)) = -\mathrm{Argmin}\{\|h\| : h \in \partial_\varepsilon f(x)\} \tag{2.1}$$

and $\varepsilon : R^n \to R$ by

$$\varepsilon(x) = \max\{\varepsilon \in \mathcal{E} : \|h_\varepsilon(x)\|^2 \ge \delta\varepsilon\} \tag{2.2}$$

where

$$\mathcal{E} = \{\varepsilon : \varepsilon = \varepsilon_0 v^k,\ k \in N^+\} U\{0\} \tag{2.3}$$

and $v \in (0,1), \varepsilon_0 > 0, \delta > 0$ are given.

Lemma 1. *The function $\|h_\varepsilon(x)\|^2$ defined by (2.1) is lower semi-continuous with respect to x.*

Proof. We know that the point-set mapping $\partial_\varepsilon f(x)$ is u.s.c. and the set $\bigcup_{x \in A} \partial_\varepsilon f(x)$ is bounded on the bounded set A. Since $h_\varepsilon(x) = -$ Argmin $\{\|h\| : h \in \partial_\varepsilon f(x)\}$ implies that $-h_\varepsilon(x) \in \partial_\varepsilon f(x)$, it follows that the set $\{-h_\varepsilon(x)\}$ is bounded. Hence, we can suppose that $-h_\varepsilon(x) \xrightarrow[x \to x_0]{} h_0$ and $h_0 \in \partial_\varepsilon f(x_0)$ holds by the upper semi-continuity of $\partial_\varepsilon f(x)$. Next, since $-h_\varepsilon(x_0) \in \partial_\varepsilon f(x_0)$ and $-h_\varepsilon(x_0) =$ Argmin $\{\|h\| : h \in \partial_\varepsilon f(x_0)\}$, it follows that $\|h_0\| \ge \| - h_\varepsilon(x_0)\| = \|h_\varepsilon(x_0)\|$. Moreover, since $\lim_{x \to x_0} -h_\varepsilon(x) = h_0$ implies $\lim_{x \to x_0} \|h_\varepsilon(x)\| = \|h_0\| \ge \|h_\varepsilon(x_0\|$, it follows that $\lim_{x \to x_0} \|h_\varepsilon(x)\|^2 \ge \|h_\varepsilon(x_0)\|^2$. Consequently, $\lim_{x \to x_0} \inf \|h_\varepsilon(x)\|^2 \ge \|h_\varepsilon(x_0)\|^2$, that is, $\|h_\varepsilon(x)\|^2$ is l.s.c..

Lemma 2. *For every $\bar{x} \in R^n$ such that $\theta \bar{\in} \partial f(\bar{x})$, there exists a $\rho(\bar{x}) > 0$ such that*

$$\varepsilon(x_i) \ge v\varepsilon(\bar{x}) > 0, \quad \textit{for all } x_i \in B(\bar{x}, \rho(\bar{x}))$$

where $B(\bar{x}, \rho(\bar{x})) = \{x : \|x - \bar{x}\| \le \rho(\bar{x})\}$.

Proof. Let $\bar{x}$ be such that $\theta \overline{\in} \partial f(\bar{x})$. Then, since $\partial_\varepsilon f(x)$ is u.s.c., there exists an $\varepsilon_1 > 0$ such that $\|h_{\varepsilon_1}(\bar{x})\|^2 \geq \frac{1}{2}\|h_0(\bar{x})\|^2 > 0$. Moreover, $\varepsilon' < \varepsilon''$ implies that $\|h_{\varepsilon'}(\bar{x})\|^2 \geq \|h_{\varepsilon''}(\bar{x})\|^2$ by definition of $h_\varepsilon(x)$. Since

$$\begin{aligned} \varepsilon(\bar{x}) &= \max\{\varepsilon \in \mathcal{E} : \|h_\varepsilon(\bar{x})\|^2 \geq \delta\varepsilon\}, \\ \mathcal{E} &= \{\varepsilon : \varepsilon = \varepsilon_0 v^k,\ k \in N^+\} U\{0\} \end{aligned} \tag{2.4}$$

and $\|h_{\varepsilon_1}(\bar{x})\|^2 \geq \frac{1}{2}\|h_0(\bar{x})\|^2 = \delta \cdot \frac{1}{\delta}\|h_0(\bar{x})\|^2$, it follows that if $\varepsilon \leq \min\left\{\varepsilon_1, \frac{1}{2}\frac{1}{\delta}\|h_0(\bar{x})\|^2\right\}$, then $\|h_\varepsilon(\bar{x})\|^2 \geq \|h_{\varepsilon_1}(\bar{x})\|^2$, and $\frac{1}{2}\frac{1}{\delta}\|h_0(\bar{x})\|^2 \geq \varepsilon$. Therefore,

$$\|h_\varepsilon(\bar{x})\|^2 \geq \|h_{\varepsilon_1}(\bar{x})\|^2 \geq \delta \cdot \frac{1}{2}\frac{1}{\delta}\|h_0(\bar{x})\|^2 \geq \delta\varepsilon. \tag{2.5}$$

It follows from (2.4) and (2.5) that

$$\varepsilon(\bar{x}) \geq \max\left\{\varepsilon \in \mathcal{E} : \varepsilon \leq \min\left\{\varepsilon_1, \frac{1}{2}\frac{1}{\delta}\|h_0(\bar{x})\|^2\right\}\right\} > 0.$$

Next, since $\|h_\varepsilon(\bullet)\|^2$ is l.s.c. by Lemma 1, there exists a $\rho(\bar{x}) > 0$ such that

$$\|h_{\varepsilon(\bar{x})}(x_i)\|^2 \geq \|h_{\varepsilon(\bar{x})}(\bar{x})\|^2, \quad \text{for all } x_i \in B(\bar{x}, \rho(\bar{x})).$$

Moreover, since $\|h_{\varepsilon(\bar{x})}(\bar{x})\|^2 \geq \delta\varepsilon(\bar{x})$,

$$\|h_{v\varepsilon(\bar{x})}(x_i)\|^2 \geq \|h_{\varepsilon(\bar{x})}(x_i)\|^2 \geq \|h_{\varepsilon(\bar{x})}(\bar{x})\|^2 \geq \delta\varepsilon(\bar{x}) \geq \delta v\varepsilon(\bar{x})$$

for all $x_i \in B(\bar{x}, \rho(\bar{x}))$.

Hence, by definition of $\varepsilon(x_i)$, one has

$$\varepsilon(x_i) \geq v\varepsilon(\bar{x}) > 0, \text{ for all } x_i \in B(\bar{x}, \rho(\bar{x})).$$

Lemma 3. *Let S' be a compact, convex subset of a compact convex set S and let $\bar{\alpha} \in (0,1)$. Let $h' = \mathrm{Nr}(S') = \mathrm{Argmin}\{\|h\| : h \in S'\}$ and let $g \in S$ be such that*

$$\langle g, h'\rangle \leq \bar{\alpha}\|h'\|^2. \tag{6}$$

Then $h'' = \mathrm{Nr}(\mathrm{co}\{g, S'\})$ satisfies

$$\|h''\|^2 \leq \max\{\bar{\alpha}^2, 1 - (1-\bar{\alpha})^2\|h'\|^2/2c^2\}\|h'\|^2,$$

where $c \geq \max\{\|g\| : g \in S\}$, $0 < \bar{\alpha} < 1$.

Proof. Since $\|h''\| = \min\{\|h\| : h \in \mathrm{co}\{g, S'\}\} \leq \min\{\|h\| : h \in \mathrm{co}\{g, h'\}\}$, it follows that

$$\begin{aligned} \|h''\|^2 &\leq \min_{0\leq\lambda\leq 1} \|\lambda g + (1-\lambda)h'\|^2 = \min_{0\leq\lambda\leq 1} \lambda^2\|g\|^2 + 2\lambda(1-\lambda)\langle g, h'\rangle + (1-\lambda)^2\|h'\|^2 \\ &\leq \min_{0\leq\lambda\leq 1} \lambda^2 c^2 + 2\lambda(1-\lambda)\bar{\alpha}\|h'\|^2 + (1-\lambda)^2\|h'\|^2 \\ &= \min_{0\leq\lambda\leq 1} (c^2 - 2\bar{\alpha}\|h'\|^2 + \|h'\|^2)\lambda^2 + (2\bar{\alpha}\|h'\|^2 - 2\|h'\|^2)\lambda + \|h'\|^2 \\ &\leq \min_{0\leq\lambda\leq 1} 2c^2(\lambda - (1-\bar{\alpha})\|h'\|^2/2c^2)^2 - (1-\bar{\alpha})^2\|h'\|^4/2c^2 + \|h'\|^2. \end{aligned}$$

Therefore, as $0 < \lambda = (1-\bar{\alpha})\|h'\|^2/2c^2 < 1$, we obtain

$$\|h''\|^2 \leq (1 - (1-\bar{\alpha})^2\|h'\|^2/2c^2)\|h'\|^2.$$

On the other hand, let g and h' be collinear. It implies that either $\langle g, h'\rangle = \|g\|\,\|h'\|$ or $\langle g, h'\rangle = -\|g\|\,\|h'\|$. In the case of $\langle g, h'\rangle = -\|g\|\,\|h'\|$, we obtain $\mathrm{Nr}(\mathrm{co}\{g, S'\}) = \theta$. In the case of $\langle g, h'\rangle = \|g\|\,\|h'\|$, since $\langle g, h'\rangle \leq \bar{\alpha}\|h'\|^2$ implies $\|g\| \leq \bar{\alpha}\|h'\|$, it follows that $\|h''\|^2 \leq \|g\|^2 \leq \bar{\alpha}^2\|h'\|^2$, which completes our proof.

Lemma 4. *Let $f : R^n \to R$ be locally Lipschitzian and be strictly differentiable at almost all differentiable points. Let $\varepsilon_i > 0$, $x_i \in R^n$ be such that $\theta \bar{\in} \partial_{\varepsilon_i} f(x_i), Y_s \subset \partial_{\varepsilon_i} f(x_i)$ be a closed convex set and $\eta_s = -\mathrm{Nr}(Y_s)$.*

(1) Suppose that for any strictly differentiable point $x_i' \in x_i + \varepsilon_i B(\theta, 1)$ satisfying

$$\langle \nabla f(x_i'), \eta_s\rangle \leq -\alpha\|\eta_s\|^2, \quad \alpha \in (0,1) \tag{2.7}$$

we have

$$\eta_s \in -\mathrm{int}(\mathrm{con}\,\partial_{\varepsilon_i} f(x_i))^*$$

and

$$f(x_i + \beta^{k_i}\eta_s) - f(x_i) \leq -\alpha\beta^{k_i}\|\eta_s\|^2,$$

where $\beta^{k_i}\|\eta_s\| \leq \varepsilon_i$.

(2) If $f(x_i + \beta^{k_i}\eta_s) - f(x_i) > -\alpha\beta^{k_i}\|\eta_s\|^2, \beta^{k_i}\|\eta_s\| \leq \varepsilon_i$, then there exists a strictly differentiable point $x_i' \in x_i + \varepsilon_i B(\theta, 1)$ and a $\delta_i' > 0$ such that for all $\bar{x}_i \in x_i' + \delta_i' B(\theta, 1)$ we have

$$\langle g, \eta_s\rangle \geq -\bar{\alpha}\|\eta_s\|^2, \qquad \text{for all } g \in \partial f(\bar{x}_i) \tag{2.8}$$

where $\bar{\alpha} \in (\alpha, 1)$.

Proof. (1) Since any strictly differentiable point $x_i' \in x_i + \varepsilon_i(\theta, 1)$ satisfies

$$\langle \nabla f(x_i'), \eta_s\rangle \leq -\alpha\|\eta_s\|^2, \quad \alpha \in (0,1), \tag{2.9}$$

it follows from $\partial f(x) = \mathrm{co}\{\lim_{x_i \to x} \nabla f(x_i)\}$ and (2.9) that

$$\langle \eta, \eta_s\rangle \leq -\alpha\|\eta_s\|^2, \qquad \text{for all } \eta \in \partial_{\varepsilon_i} f(x_i) \tag{2.10}$$

where $\alpha \in (0,1)$, that is, $\eta_s \in$ -int $(\mathrm{con}\ \partial_{\varepsilon_i} f(x_i))^*$.

Moreover, since $f(x_i + \beta^{k_i}\eta_s) - f(x_i) = \langle \xi_{iu}, \beta^{k_i}\eta_s\rangle$ where $\xi_{iu} \in \partial f(x_i + u\beta^{k_i}\eta_s), u \in (0,1)$, and $\beta^{k_i}\|\eta_s\| \leq \varepsilon_i$, it follows from (2.10) that

$$f(x_i + \beta^{k_i}\eta_s) - f(x_i) \leq -\alpha\beta^{k_i}\|\eta_s\|^2.$$

(2) If $f(x_i + \beta^{k_i}\eta_s) - f(x_i) > -\alpha\beta^{k_i}\|\eta_s\|^2, \beta^{k_i}\|\eta_s\| \leq \varepsilon_i$, then by the proof of case (1) there exists a strictly differentiable point $x_i' \in x_i + \varepsilon_i B(\theta, 1)$ such that

$$\langle \nabla f(x_i'), \eta_s\rangle > -\alpha\|\eta_s\|^2, \quad \alpha \in (0,1). \tag{2.11}$$

Next, by the upper semi-continuity of $\partial f(x)$, for any $\varepsilon > 0$, there exists a $\delta_i' > 0$ such that

$$x_i' + \delta_i' B(\theta, 1) \subset x_i + \varepsilon_i B(\theta, 1)$$

and as $\bar{x}_i \in x_i' + \delta_i' B(\theta, 1)$,

$$\partial f(\bar{x}_i) \subset \partial f(x_i') + \varepsilon B(\theta, 1) = \nabla f(x_i') + \varepsilon B(\theta, 1). \tag{2.12}$$

Hence

$$\|\eta - \nabla f(x_i')\| \le \varepsilon, \quad \text{for all } \eta \in \partial f(\bar{x}_i). \tag{2.13}$$

If we suitably choose an $\varepsilon > 0$ and a $\delta_i' > 0$, it follows from (2.11) and (2.13) that for all $\bar{x}_i \in x_i + \delta_i' B(\theta, 1)$,

$$\langle \eta, \eta_s \rangle \ge -\bar{\alpha}\|\eta_s\|^2, \quad \text{for all } \eta \in \partial f(\bar{x}_i),$$

where $\bar{\alpha} \in (\alpha, 1)$.

§3. The Algorithm and its Global Convergence

Based on the lemmas in Section 2, we can propose an implementable algorithm for the locally Lipschitz function, which is globally convergent.

Algorithm.

Parameters : $\varepsilon_0 > 0, \alpha \in (0,1), \bar{\alpha} \in (\alpha, 1), \beta, \nu \in (0,1)$ and positive integers s, N are given.

Date : $x_0 \in R^n$.

Step 0: Set $i = 0$.

Step 1: Set $\varepsilon = \varepsilon_0$.

Step 2: Compute $Y_s \subset \partial_\varepsilon f(x_i)$, a convex hull of s points in $\partial_\varepsilon f(x_i)$.

Step 3: Compute $\eta_s = -\mathrm{Nr}(Y_s)$.

Step 4: If $\|\eta_s\| < \varepsilon$, set $\varepsilon = \nu\varepsilon$ and go to step 2. Else compute $k_s \in N^+$ such that $\beta\varepsilon \le \beta^{k_s}\|\eta_s\| \le \varepsilon$.

Step 5: If

$$f(x_i + \beta^{k_s}\eta_s) - f(x_i) \le -\alpha\beta^{k_s}\|\eta_s\|^2. \tag{3.1}$$

(i) Set $h_i = \eta_s$ and compute the smallest $k_i \in N^+$ such that

$$f(x_i + \beta^{k_i}h_i) - f(x_i) \le -\alpha\beta^{k_i}\|h_i\|^2. \tag{3.2}$$

(ii) Set $x_{i+1} = x_i + \beta^{k_i}h_i$, $i := i+1$, and go to step 1.

Step 6: Set

$$\delta_s = \beta^{k_s}\|\eta_s\|/2N, \quad e_l = \underset{(0,\cdots,0,1,0,\cdots,0)}{l-\text{th}}.$$

Step 7: Choice $x_{i,s+1}^{(j_1,j_2,\cdots,j_n)} \in B(x_i + \sum_l (2j_l - 1)\delta_s e_l; \delta_s)$, where $j_l = 0, \cdots, N$, $l = 1, \cdots, n$.

Step 8: Compute $g_f\left(x_{i,s+1}^{(j_1,j_2,\cdots,j_n)}\right) \in \partial f(x_{i,s+1}^{(j_1,j_2,\cdots,j_n)})$.

Step 9: If

$$\langle g_f(x_{i,s+1}^{(j_1,j_2,\cdots,j_n)}), \eta_s \rangle \ge -\bar{\alpha}\|\eta_s\|^2 \tag{3.3}$$

set $Y_{s+1} = \mathrm{co}\{\{g_f(x_{i,s+1}^{(j_1,j_2,\cdots,j_n)})\} U Y_s\}$. Set $s := s+1$ and go to step 3. Else, set $N = 2N$ and go to step 6.

Remark. We use the Grid search method to choose the points $x_{i,s+1}^{(j_1,j_2,\cdots,j_n)}$ in step 7.

We shall now demonstrate the theorem for global convergence.

Theorem. *Let f be locally Lipschitzian and be strictly differentiable at almost all differentiable points. Then*

(1) *If the algorithm generates a finite sequence $\{x_i\}_{i=0}^{N}$, jamming at x_N, i.e., with construction stopping and the algorithm cycling in the loop defined by steps $2-4$, or steps $3-9$ or steps $6-9$, then $\theta \in \partial f(x_N)$.*

(2) *It the algorithm generates an infinite sequence $\{x_i\}_{i=0}^{\infty}$, then every accumulation point $\hat{x}$ of $\{x_i\}_{i=0}^{\infty}$ satisfies $\theta \in \partial f(\hat{x})$.*

Proof. (1) Suppose that the sequence $\{x_i\}$ is finite with the algorithm jamming up at x_N, cycling infinitely in one of the loops defined by steps 2 to 4 or steps 3 to 9 or steps 6 to 9. Suppose that $\theta \bar{\in} \partial f(x_N)$.

(i) Consider the loop defined by steps 2 to 4. Since $\theta \bar{\in} \partial f(x_N)$, by definition

$$\varepsilon(x_N) = \max\{\varepsilon \in \mathcal{E} / \|h_\varepsilon(x_N)\|^2 \geq \delta\varepsilon\} > 0$$

where $\mathcal{E} = \{\varepsilon/\varepsilon = \varepsilon_0 \nu^k,\ k \in N^+\} U\{0\}$, $h_\varepsilon(x_N) = -\mathrm{Nr}(\partial_\varepsilon f(x_N)) = -\mathrm{Argmin}\{\|h\|/h \in \partial_\varepsilon f(x_N)\}, \nu \in (0,1), \varepsilon_0,\ \delta > 0$. Hence for all $\varepsilon \leq \varepsilon(x_N), Y_s \subset \partial_\varepsilon f(x_N), \|\eta_s\|^2 = \|\mathrm{Nr}(Y_s)\|^2 \geq \|\mathrm{Nr}\partial_\varepsilon f(x_N)\|^2 \geq \|\mathrm{Nr}\partial_{\varepsilon(x_N)} f(x_N)\|^2 = \|h_{\varepsilon(x_N)}(x_N)\|^2 \geq \delta\varepsilon(x_N) \geq \delta\varepsilon$, and hence no infinite cycling can occur in this loop.

(ii) Consider the loop defined by steps 3 to 9. If $\|\eta_s\| \underset{s}{\rightarrow} c_0 > 0$, there exists an $s_0 > 0$ such that as $s \geq s_0$, $\|\eta_s\| \geq c_0 - \varepsilon > 0$, where $\varepsilon > 0$ is smaller. Set $h' = -\eta_s$, and by Lemma 3

$$\|\eta_{s+1}\|^2 \leq \max\{\bar{\alpha}^2,\ 1-(1-\bar{\alpha})^2\|\eta_s\|^2/2c^2\}\|\eta_s\|^2 \leq \max\{\bar{\alpha}^2,\ 1-(1-\bar{\alpha})^2(c_0-\varepsilon)^2/2c^2\}\|\eta_s\|^2$$

where $c \geq \max\{\|\eta\|/\eta \in \partial_{\varepsilon_0} f(x_N)\} \geq c_0$. It follows that $\|\eta_{s+m}\| \underset{m}{\rightarrow} 0$, which contradicts $\|\eta_s\| \underset{s}{\rightarrow} c_0 > 0$.

If $\|\eta_s\| \underset{s}{\rightarrow} 0$, the algorithm must go to step 2 from step 4, and hence no infinite cycling can occur in this loop.

(iii) Consider the loop defined by steps 6 to 9. In this case the expression (3.1) does not hold, and hence by Lemma 4, there exists a strictly differentiable point $x_i' \in x_i + \varepsilon_i B(\theta, 1)$ and a $\delta_i' > 0$ such that for all $\bar{x}_i \in x_i' + \delta_i' B(\theta, 1)$,

$$\langle g, \eta_s \rangle \geq -\bar{\alpha}\|\eta_s\|^2, \quad \text{for all } g \in \partial f(\bar{x}_i)$$

where $\bar{\alpha} \in (\alpha, 1)$. Hence no infinite cycling can occur in this loop.

(2) Now suppose that the sequence $\{x_i\}$ is infinite. Suppose that $x_i \xrightarrow{K} \hat{x}$, with $K \subset \{0, 1, 2, \cdots\}$ infinite, and that $\theta \bar{\in} \partial f(\hat{x})$. Then, by Lemma 2 there exists an i_0 such that for all $i \in K, i \geq i_0$, $\varepsilon(x_i) \geq \nu\varepsilon(\hat{x}) > 0$. Consequently, for all $i \in K$, $i \geq i_0$, the expression (3.2) of the algorithm is satisfied with $\|\eta_s\| \geq \nu\varepsilon(x)$ by $\|\eta_s\| \geq \varepsilon(x_i)$, and $\beta^{k_i}\|\eta_s\| \geq \beta\varepsilon(x_i) \geq \beta\nu\varepsilon(\hat{x})$ by the algorithm. Hence, by (3.2) for all $i \in K, i \geq i_0$,

$$f(x_{i+1}) - f(x_i) \leq -\alpha\beta^{k_i}\|h_i\|^2 \leq -\alpha\beta(\nu\varepsilon(x))^2. \tag{3.4}$$

Now $f(x_i) \xrightarrow{K} f(\hat{x})$ by continuity and $\{f(x_i)\}$ is monotonically decreasing. Hence, we must have $f(x_i) \underset{i}{\rightarrow} f(\hat{x})$, which contradicts (3.4). This completes our proof.

§4. Numerical Examples

This section describes the computational results of the algorithm in Section 3. The algorithm is coded in Fortran 77 on an IBM PC.

Example 1.

$$\min f(x) = 1 + \sum_{i=1}^{5}(|x_i| + ix_i^2).$$

The optimal point is $X = (0,0,0,0,0)$, and the optimal value is 1.

1) We choose an initial point $X_0 = (10,10,10,10,10)$. After 25 iterations, we obtain X^*=(2.401899E–004 , 3.978963E–005, 1.479298E–004, -3.129370E–004, 1.627195E–004) and the optimal value $f^* = 1.000904000$.

2) we choose $X_0 = (10,-24,35,18,-54)$. After 20 iterations, we obtain $X^* = ($ -3.102912E–004, -3.664491E–007, 2.227507E–004, 4.288445E–005, 1.320743E–004) and $f^* =$ 1.000709000.

Example 2.

$$\min f(x,y) = \begin{cases} 5(9x^2+16y^2)^{1/2}, & \text{for } x \geq |y|, \\ 9x+16|y|, & \text{for } 0 < x < |y|, \\ 9x+16|y|-x^9, & \text{for } x \leq 0. \end{cases}$$

The optimal point is $(x,y) = (-1,0)$ and the optimal value is -8. We choose an initial point $(x_0,y_0) = (1.4,0.8)$. After 26 iterations, we obtain $(x^*,y^*) = (-1.0000680$, 3.040496E–006) and $f^* = -7.999951000$.

The above examples illustrate that the algorithm in Section 3 is efficient.

References

[1] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley-Interscience, New York, 1983.

[2] C. Lemarechal, An extension of Davidon methods to nondifferentiable optimization problems, in Nondifferentiable Optimization, Mathematical Programming Study 3, North Holland, Amsterdam, 1975, 95–109.

[3] R. Mifflin, An algorithm for constrained optimization with semismooth functions, *Math. Oper. Res.*, **2** (1977), 191–207.

[4] E. Polak, Computational Methods in Optimization, A Unified Approach, Academic Press, New York, 1971.

[5] E. Polak, On the Extension of Constrained Optimization Algorithms from Differentiable to Nondifferentiable Problems, *SIAM J. Control and Optimization,* **21** : 2 (1983), 179–203.

[6] J. Zowe, Nondifferentiable Optimization, NATO ASI Series, Vol. F15, Computational Mathematical Programming.