Some Properties of Trust Region Algorithms for Nonsmooth Optimization

Y. Yuan
Department of Applied Mathematics and Theoretical Physics
University of Cambridge
Silver Street
Cambridge, CB3 9EW
England

Abstract

This paper discusses some properties of trust region algorithms for nonsmooth optimization. The problem is expressed as the minimization of a function $h(f(x))$, where $h(.)$ is convex and f is a continuously differentiable mapping from $\Re^n$ to $\Re^m$. Conditions for the convergence of a class of algorithms are discussed, and it is shown that the class includes minimax and L_1 problems.

Key words: Trust Region Algorithms, Nonsmooth Optimization,
Kuhn-Tucker Points.

Technical Report: DAMTP 1983/NA4

1. Introduction

Many papers have been published on trust region algorithms, for example see Moré(1982). Powell (1970, 1975, 1983) and Sorensen (1982), but most attention has been given to the smooth case. We investigate some properties of trust region algorithms for nonsmooth cases. The problem we went to solve is

$$\min_{x \in \Re^n} h(f(x)), \tag{1.1}$$

where $h(.)$ is a convex function defined on $\Re^m$ and is bounded below; $f(x) = (f_1(x), \ldots, f_m(x))^T$ is a map from $\Re^n$ to $\Re^m$ and $f_i(x)(i = 1, \ldots, m)$ are all continuously differentiable functions on $\Re^n$.

The trust region algorithms are iterative and an initial point $x_1 \in \Re^n$ should be given. The methods generate a sequence of points $x_k(k = 1, 2, \ldots)$ in the following way. At the beginning of $k-$th iteration, x_k, Δ_k and B_k are available, where $\Delta_k > 0$ is a step-bound and B_k is an $n \times n$ real symmetric matrix. Let d_k be a solution of

$$\min h(f(x_k) + \nabla^T f(x_k)d) + \frac{1}{2} d^T B_k d, \tag{1.2}$$

subject to

$$||d|| \leq \Delta_k. \tag{1.3}$$

Here $||.||$ may be any norm in $\Re^n$ space. Let

$$x_{k+1} = \begin{cases} x_k + d_k, & \text{if } h(f(x_k)) > h(f(x_k + d_k)) \\ x_k, & \text{otherwise .} \end{cases} \tag{1.4}$$

It is noted that our choice of x_{k+1} is different from Powell's (1983). He let $x_{k+1} = x_k + d_k$ if and only if the inequality

$$h(f(x_k + d_k) \leq h(f(x_k) - c[h(f(x_k)) - \phi_k(d_k)] \tag{1.5}$$

holds for some constant $c \in (0, 1)$, where $\phi_k(d)$ is defined by

$$\phi_k(d) = h(f(x_k) + \nabla^T f(x_k)d) + \frac{1}{2} d^T B_k d. \tag{1.6}$$

Thus, because our condition for letting $x_{k+1} = x_k + d_k$ is weaker than Powell's, our algorithms let $x_{k+1} = x_k + d_k$ more often, and we have the desirable property of accepting any trial vector of variables that reduces the objective function.

Similar to Powell (1975), let Δ_{k+1} satisfy

$$||d_k|| \leq \Delta_{k+1} \leq c_1 ||d_k|| \tag{1.7}$$

if

$$h(f(x_k)) - h(f(x_k + d_k)) \geq c_2[h(f(x_k)) - \phi_k(d_k)], \tag{1.8}$$

otherwise let

$$c_3 ||d_k|| \leq \Delta_{k+1} \leq c_4 ||d_k||, \tag{1.9}$$

where $c_i (i = 1, 2, 3, 4)$ are positive constant satisfying $c_1 \geq 1, c_2 < 1$ and $c_3 \leq c_4 < 1$. We also let $\Delta_k (k = 1, 2, \ldots)$ satisfy

$$\Delta_k \leq \bar{\Delta}. \tag{1.10}$$

for some positive constant $\bar{\Delta}$. Our theory applies to several techniques for generating B_{k+1}.

In section 2, conditions for convergence are discussed. Under the assumption that

$$||B_k|| \leq c_5 + c_6 \sum_{i=1}^{k} \Delta_i \tag{1.11}$$

and that $\{x_k\}$ are bounded, section 2 proves that

$$\liminf_{k \to \infty} \psi(x_k) = 0, \tag{1.12}$$

where c_5 and c_6 are two positive constants, $||B_k||$ is the matrix norm subordinate to the vector norm $||.||$, that is

$$||B_k|| = \sup_{x \neq 0} ||B_k x|| / ||x|| \tag{1.13}$$

(see Wilkinson (1965)) and $\psi(.)$ is defined by

$$\psi(x) = h(f(x)) - \min_{||d|| \leq 1} h(f(x) + \nabla^T f(x) d), \quad \forall x \in \Re^n. \tag{1.14}$$

Two conditions are given at the beginning of Section 3. If (1.11) and these two conditions are satisfied, and if the points $\{x_k\}$ are all in a small neighbourhood of the solution, it is proved in Section 3 that $||B_k|| (k = 1, 2, \ldots)$ are uniformly bounded and, not only does $\{x_k\}$ converge to the solution, but also $\sum ||d_k||$ is finite.

In Section 4 and Section 5, minimax and L_1 problems are analyzed. In either case, strict complementarity and second order sufficiency are assumed. If (1.11) holds and $\{x_k\}$ are in a small neighbourhood of the solution, it is proved that the conditions of Section 3 are satisfied.

2. Kuhn-Tucker Points

In this section, it is proved that the sequence $\{x_k\}$ generated by our algorithm is bounded away Kuhn-Tucker points. Here Kuhn-Tucker points are defined to be those at which the equation

$$\psi(x) = 0 \tag{2.1}$$

holds, where $\psi(.)$ is defined by (1.14).

Since any two norms in $\Re^n$ are equivalent, there exist positive constants c_7 and c_8 such that

$$c_7 ||d|| \leq ||d||_2 \leq c_8 ||d|| \tag{2.2}$$

holds for all $d \in \Re^n$.

Let us define

$$Max_L(x) = h(f(x)) - \min_{||d|| \leq L} h(f(x) + \nabla^T F(x)d) \tag{2.3}$$

for all $x \in \Re^n$ and all $L > 0$, than we have the following lemma.

Lemma 2.1 *If $Max_L(x)$ and $\phi_k(d)$ are defined by (2.3) and (1.6) respectively, than*

$$h(f(x_k)) - \phi_k(d_k) \geq \frac{1}{2}\min\{Max_{\Delta_k}(x_k), c_8^{-2}[Max_{\Delta_k}(x_k)]^2/||B_k||\Delta_k^2\}. \tag{2.4}$$

Proof By the definition of d_k, we have

$$h(f(x_k)) - \phi_k(d_k) \geq h(f(x_k)) - \phi_k(d), \quad \forall ||d|| \leq \Delta_k. \tag{2.5}$$

Let $||\bar{d}_k|| \leq \Delta_k$ satisfy

$$Max_{\Delta_k}(x_k) = h(f(x_k) - h(f(x_k) + \nabla^T f(x_k)\bar{d}_k) \tag{2.6}$$

Then, unity the convexity of $h(.)$, we have, for $\alpha \in [0, 1]$,

$$\begin{aligned} h(f(x_k)) - \phi_k(d_k) &\geq h(f(x_k)) - \phi_k(\alpha \bar{d}_k) \\ &\geq \alpha Max_{\Delta_k}(x_k) - \frac{1}{2}\alpha^2 \bar{d}_k^2 B_k \bar{d}_k \\ &\geq \alpha Max_{\Delta_k}(x_k) - \frac{1}{2}\alpha^2 ||\bar{d}_k||_2 ||B_k \bar{d}_k||_2 \\ &\geq \alpha Max_{\Delta_k}(x_k) - \frac{1}{2}c_8^2 ||B_k||\Delta_k^2 \alpha^2 \end{aligned} \tag{2.7}$$

The last line of the above inequality is due to (2.2) and $||\bar{d}_k|| \leq \Delta_k$. Therefore

$$\begin{aligned} h(f(x_k)) - \phi_k(d_k) &\geq \max_{0 \leq \alpha \leq 1}[\alpha Max_{\Delta_k}(x_k) - \frac{1}{2}c_8^2||B_k||\Delta_k^2\alpha^2] \\ &\geq \frac{1}{2}\min\{Max_{\Delta_k}(x_k), c_8^{-2}[Max_{\Delta_k}(x_k)]^2/||B_k||\Delta_k^2\}, \end{aligned} \tag{2.8}$$

which ensures(2.4). □

By definition,$\psi(x) = Max_1(x)$ for all $x \in \Re^n$, and by the convexity of $h(.)$, we have

$$Max_L(x) \geq \min\{L, 1\}\psi(x) \tag{2.9}$$

for all $x \in \Re^n$ and all $L > 0$.

The following theorem, similar to Powell's (1975), implies that $\{x_k\}$ is not bounded away from Kuhn-Tucker points.

Theorem 2.2 *If (1.11) is satisfied and $\{x_k\}$ are bounded, then*

$$\liminf_{k \to \infty} \psi(x_k) = 0 \tag{2.10}$$

Proof Assume that the theorem is invalid, then there exists $\delta_1 > 0$, such that

$$\psi(x_k) > \delta_1 \tag{2.11}$$

for all k. From (2.9) and the above inequality,

$$Max_{\Delta_k}(x_k) \geq \min\{\Delta_k, 1\}\delta_1 \quad . \tag{2.12}$$

By lemma 2.1 and (2.12), we have

$$h(f(x_k)) - \phi_k(d_k) \geq \frac{1}{2}\delta_1 \min\{\Delta_k, 1, c_8^{-2}\delta_1/||B_k||, c_8^{-2}\delta_1/||B_k||\Delta_k^2\} \tag{2.13}$$

Let $\sum'$ denote the sum over the iterations on which (1.8) holds. Then by the fact that $h(.)$ is bounded below, we have

$$\sum_k{}'[h(f(x_k)) - \phi_k(d_k)] \tag{2.14}$$

is convergent. From (2.13) we have

$$\sum_k{}'\Delta/(c_8 + c_6\sum_{i=1}^{k}\Delta_i) \tag{2.15}$$

is convergent. By the definition of Δ_k, we have, due to Powell (1975),

$$\sum_{i=1}^{k}\Delta_i \leq (1 + c_1/(1-c_4))[\Delta_1 + \sum_{i=1}^{k}{}'\Delta_i] \quad . \tag{2.16}$$

Therefore

$$\sum_k{}'\Delta/[c_5 + c_6(1 + \frac{c_1}{1-c_4})\Delta_1 + c_6(1 + \frac{c_1}{1-c_4})\sum_{i=1}^{k}{}'\Delta_i] \tag{2.17}$$

is also convergent. Hence, we obtain that

$$\sum_k \Delta_k \tag{2.18}$$

is convergent. Noticing (2.16), we have

$$\sum_{k=1}^{\infty}\Delta_k \tag{2.19}$$

is finite. Consequently, $||B_k||$ are uniformly bounded, and by (2.13),

$$h(f(x_k)) - \phi_k(d_k) \geq \frac{1}{2}\delta_1\Delta_k, \quad \text{for } k \geq k_1 \tag{2.20}$$

where k_1 is a constant integer. Since

$$h(f(x_k + d_k)) - \phi_k(d_k) = O(||d_k||^2) \tag{2.21}$$

we have, from (2.20) and (2.21),

$$\lim \frac{h(f(x_k)) - h(f(x_k + d_k))}{h(f(x_k)) - \phi_k(d_k)} = 1 > c_2 \tag{2.22}$$

Thus there exists $k_2 > 0$ such that

$$\Delta_{k+1} \geq ||d_k||, \quad \text{for } k \geq k_2 \tag{2.23}$$

Now, by the argument of the proof of lemma 6 of Powell (1983), there exists a positive constant η such that, if $||d_k|| < \Delta_k$, then $||d_k|| > \eta$. However, the convergence of the sum (2.19) implies that $||d_k|| \to 0$. Therefore there exists $k_3 > 0$ such that

$$\Delta_k = ||d_k||, \quad \text{for } k \geq k_3. \tag{2.24}$$

From (2.23) and (2.24), it follows that

$$\Delta_{k+1} \geq \Delta_k \quad \text{for } k \geq \max\{k_2, k_3\}. \tag{2.25}$$

This contradicts (2.19), which completes our proof. □

The conditions of theorem 2.3 are often satisfied. In fact many algorithms demand (1.11). The condition that $\{x_k\}$ are bounded is also usually satisfied, in particular when x_1 is chosen so that

$$\{x | h(f(x)) \leq h(f(x_1))\} \tag{2.26}$$

is a bounded set in $\Re^n$.

Since $\psi(x)$ is continuous (see Powell (1983)), $\{x_k\}$ can not be bounded from Kuhn-Tucker points if the conditions of theorem 2.2 are satisfied. Further, theorem 2.2 gives the following corollary.

Corollary 2.3 *If $\bar{x}$ is an isolated local minimum of the objective function (1.1) at which $\psi(\bar{x}) = 0$, if $\psi(.) \neq 0$ at every other point in the neighbourhood of $\bar{x}$, and if x_k is in this neighbourhood for all sufficiently large k , then $x_k \to \bar{x}$.*

Proof It follows directly from theorem 2.2 that the monotonically decreasing sequence $\{h(f(x_k)); k = 1, 2, \ldots\}$ converges to $h(f(\bar{x}))$, and it follows from continuity that, for any $\delta > 0$ there exists $\epsilon > 0$ such that, if $||x_k - \bar{x}|| \geq \delta$ then $h(f(x_k)) \geq h(f(\bar{x})) + \epsilon$. Therefore the corollary is true. □

3. Convergence Results

In this section, two conditions are given on a Kuhn-Tucker point, and under these conditions it is proved that, if $\{x_k\}$ are in a small neighbourhood of the Kuhn-Tucker point, then the conditions in corollary 2.3 hold and $\sum_{k=1}^{\infty} ||d_k||$ is finite. Hence, assuming (1.11), $||B_k||$ is uniformly bounded. The two conditions are as follows. $h(f)$ is said to satisfy condition(I) at $\bar{x}$ if and only if there exist $\epsilon_1 > 0$ and $c_9 > 0$ such that

$$h(f(x)) - h(f(\bar{x})) \geq c_9 ||x - \bar{x}||^2 \tag{3.1}$$

for all $||x - \bar{x}|| < \epsilon_1$. $h(f)$ is said to satisfy Condition (II) at $\bar{x}$ with respect to α_0 if and only if there exists $\epsilon_2 > 0$ such that

$$Max_{||x-\bar{x}||}(x) \geq \alpha_0 [h(f(x)) - h(f(\bar{x}))] \tag{3.2}$$

for all $||x - \bar{x}|| < \epsilon_2$, where $Max_{||x-\bar{x}||}(x)$ is defined by (2.3).

The following lemma shows that Condition(I) implies that $\bar{x}$ is a Kuhn-Tucker point, and that the assertion $\psi(\bar{x}) = 0$ in Corollary 2.3 is redundant.

Lemma 3.1 *If $\bar{x}$ is a local minimum of $h(f(x))$, then $\bar{x}$ is a Kuhn-Tucker point.*

Proof If the lemma is invalid, then there exists $d_0 \in \Re^n$ that

$$h(f(\bar{x}) + \nabla^T f(\bar{x}) d_0) < h(f(\bar{x})). \tag{3.3}$$

By the convexity of h(.) we have, for $\alpha \in [0, 1]$

$$h(f(\bar{x}) + \alpha \nabla^T f(\bar{x}) d_0) < h(f(\bar{x})) - \alpha [h(f(\bar{x})) - h(f(\bar{x}) + \nabla^T f(\bar{x}) d_0)]. \tag{3.4}$$

Hence for small $\alpha > 0$,

$$h(f(\bar{x} + \alpha d_0)) < h(f(\bar{x})) - \alpha [h(f(\bar{x})) - h(f(\bar{x}) + \nabla^T f(\bar{x}) d_0)] + o(\alpha), \tag{3.5}$$

which contradicts the assumption. Hence the lemma is true. □

Our main result that $\sum ||d_k||$ is convergent depends on the following two lemmas.

Lemma 3.2 *Let $\{x_k\}$ be generated by the algorithms stated in Section 1. Assume that $h(f)$ satisfies Condition (II) at $\bar{x}$ with respect to some $\alpha_0 \in (0, 1)$ and that x_k is in a neighbourhood of $\bar{x}$ such that (3.2) holds for sufficiently large k. Then we have*

$$h(f(x_k)) - \phi_k(d_k) \geq \frac{1}{2} \alpha_0 D_k \min\{\Delta_k, ||x_k - \bar{x}||, \alpha_0 |D_k| / c_8^2 ||B_k||\}, \tag{3.6}$$

where

$$D_k = [h(f(x_k)) - h(f(\bar{x}))] / ||x_k - \bar{x}|| \tag{3.7}$$

and $\phi_k(.)$ is defined by (1.6).

Proof Obviously the lemma is valid if $h(f(x_k)) \leq h(f(\bar{x}))$. So we assume

$$h(f(x_k)) > h(f(\bar{x})) \tag{3.8}$$

for all k. Define $\hat{d}_k$ satisfying $||\hat{d}_k|| \leq ||x_k - \bar{x}||$ and

$$h(f(x_k) + \nabla^T f(x_k)\hat{d}_k) = \min_{||d_k|| \leq ||x_k - \bar{x}||} h(f(x_k) + \nabla^T f(x_k)d), \tag{3.9}$$

then for all $\alpha \in [0, 1]$,

$$\begin{aligned} h(f(x_k)) &- \phi_k(\alpha \min\{\Delta_k/||x_k - \bar{x}||, 1\}\hat{d}_k) \\ &\geq \alpha \min\{\Delta_k/||x_k - \bar{x}||, 1\}[h(f(x_k)) - h(f(x_k) + \nabla^T f(x_k)\hat{d}_k)] \\ &\quad -\frac{1}{2}\alpha^2[\min\{\Delta/||x_k - \bar{x}||, 1\}]^2 \hat{d}_k^T B_k \hat{d}_k \\ &\geq \alpha\alpha_0 \min\{\Delta_k, ||x_k - \bar{x}||\} D_k \\ &\quad -\frac{1}{2}\alpha^2 c_8[\min \Delta_k, ||x_k - \bar{x}||\}]^2 ||B_k||. \end{aligned} \tag{3.10}$$

Hence, by the definition of d_k, we have

$$\begin{aligned} h(f(x_k)) &- \phi_k(d_k) \\ &\geq \max_{0 \leq \alpha \leq 1} \{h(f(x_k)) - \phi_k(\alpha \min\{\Delta/||x_k - \bar{x}||, 1\}\hat{d}_k)\} \\ &\geq \frac{1}{2} \min\{\alpha_0 \min\{\Delta_k, ||x_k - \bar{x}||\} D_k, \alpha_0^2 D_k^2 / c_8^2 ||B_k||\}, \end{aligned} \tag{3.11}$$

which ensures (3.6). $\square$

Lemma 3.3 *Assume that all the conditions in the previous lemma are satisfied, that $h(f)$ satisfies Condition(I) at $\bar{x}$ and that x_k is in a neighbourhood of $\bar{x}$ such that (3.1) holds for sufficiently large k. If*

$$\sum_k{}' \min\{||d_k||, ||x_k - \bar{x}||, D_k/(1 + ||B_k||)\} \tag{3.12}$$

is convergent, then

$$\sum_k{}' ||d_k||/(1 + ||B_k||) \tag{3.13}$$

is also convergent.

Proof The theorem is trivial if $||d_k||$ is the smallest term in the braces of expression (3.12) for all sufficiently large k. Therefore, if the lemma is invalid, there exist $x_{k_j} (j = 1, 2, \ldots)$ such that $k_j \in \sum'$ and

$$\sum_{j=1}^{\infty} \min\{||x_{k_j} - \bar{x}||, D_{k_j}/(1 + ||B_{k_j}||\} \tag{3.14}$$

is convergent but

$$\sum_{j=1}^{\infty} ||d_{k_j}||/(1+||B_{k_j}||) \tag{3.15}$$

is divergent. Without lose of generality, we assume

$$\lim_{j\to\infty} \frac{min\{||x_{k_j}-\bar{x}||, D_{k_j}/(1+||B_{k_j}||)\}}{||d_{k_j}||/(1+||B_{k_j}||)} = 0 \tag{3.16}$$

If there are only finitely many j such that the equation

$$\min\{||x_{k_j}-\bar{x}, D_{k_j}/(1+||B_{k_j}||)\} = ||x_{k_j}-\bar{x}|| \tag{3.17}$$

holds, then we have

$$\lim_{j\to\infty} D_{k_j}/||d_{k_j}|| = 0. \tag{3.18}$$

Hence, by the definition (3.7) of D_k,

$$h(f(x_{k_j})) - h(f(\bar{x})) = o(||x_{k_j}-\bar{x}||||d_{k_j}||). \tag{3.19}$$

It follows from (3.1) that

$$||x_{k_j}-\bar{x}|| = o(||d_{k_j}||). \tag{3.20}$$

On the other hand, because $k_j \in \sum'$ and because

$$h(f(x_{k_j}+d_{k_j})) - h(f(\bar{x})) \leq h(f(x_{k_j})) - h(f(\bar{x})) = o(||x_{k_j}-\bar{x}||||d_{k_j}||), \tag{3.21}$$

we obtain from (3.1) the equation

$$||x_{k_j}+d_{k_j}-\bar{x}||^2 = o(||x_{k_j}-\bar{x}||||d_{k_j}||). \tag{3.22}$$

Since (3.20) gives

$$\lim_{j\to\infty} ||x_{k_j}+d_{k_j}-\bar{x}||/||d_{k_j}|| = 1, \tag{3.23}$$

it follows that

$$||d_{k_j}||^2 = o(||x_{k_j}-\bar{x}||||d_{k_j}||), \tag{3.24}$$

which is a contradiction to (3.20). So we assume that there are infinitely many j such that (3.17) holds. Without loss of generality, we assume (3.17) holds for all j. Hence

$$\lim_{j\to\infty} ||x_{k_j}-\bar{x}||(1+||B_{k_j}||)/||d_{k_j}|| = 0 \quad . \tag{3.25}$$

So again we derive that

$$||x_k-\bar{x}|| = o(||d_{k_j}||). \tag{3.26}$$

By the definition of $\hat{d}_k$ in (3.9) and by (3.2), we have

$$\begin{aligned} h(f(_{k_j})) &- \phi_{k_j}(\hat{d}_{k_j}) \\ &\geq \alpha_0[h(f(x_{k_j})) - h(f(\bar{x})] - \frac{1}{2}c_8^2||x_{k_j} - \bar{x}||^2||B_{k_j}||. \end{aligned} \tag{3.27}$$

Since

$$\begin{aligned} h(f(x_{k_j})) &- h(f(\bar{x})) \\ &\geq [(h(f(x_{k_j})) - h(f(\bar{x}))(h(f(x_{k_j} + d_{k_j})) - h(f(\bar{x}))]^{\frac{1}{2}} \\ &\geq c_9||x_{k_j} - \bar{x}||||x_{k_j} + d_{k_j} - \bar{x}|| \end{aligned} \tag{3.28}$$

and by (3.25) and (3.26)

$$\begin{aligned} ||x_{k_j} - \bar{x}||^2||B_{k_j}|| &= o(||d_{k_j}||||x_{k_j} - \bar{x}||) \\ &= o(||x_{k_j} + d_{k_j} - \bar{x}||||x_{k_j} - \bar{x}||), \end{aligned} \tag{3.29}$$

there exists $j_1 > 0$ such that

$$h(f(x_{k_j})) - \phi_{k_j}(\hat{d}_{k_j}) \geq \frac{1}{2}\alpha_0[h(f(x_{k_j})) - h(f(\bar{x}))] \tag{3.30}$$

for all $j > j_1$. From (3.20) and (3.26), there exists j_2 such that

$$h(f(x_{k_j})) - \phi_{k_j}(d_{k_j}) \geq \frac{1}{2}\alpha_0[h(f(x_{k_j})) - h(f(\bar{x}))] \tag{3.31}$$

for all $j > j_1$. Therefore, from (1.8)

$$h(f(x_{k_j})) - h(f(x_{k_j} + d_{k_j})) \geq \frac{1}{2}\alpha_0 c_0[h(f(x_{k_j})) - h(f(\bar{x}))], \tag{3.32}$$

which we rewrite as

$$h(f(x_{k_j} + d_{k_j})) - h(f(\bar{x})) \leq (1 - \frac{1}{2}\alpha_0 c_2)[h(f(x_{k_j})) - h(f(\bar{x}))]. \tag{3.33}$$

Thus, since $(h(f(x_k)))$ decreases monotonically,

$$[h(f(x_{k_j} + d_{k_j})) - h(f(\bar{x}))]^{\frac{1}{2}} \leq (1 - \frac{1}{2}\alpha_0 c_2)^{\frac{1}{2}}[h(f(x_{k_j} + d_{k_j})) - h(f(\bar{x}))]^{\frac{1}{2}} \tag{3.34}$$

for $j \geq j_2$. Consequently,

$$\sum_{j=j_2}^{\infty}[h(f(x_{k_j} + d_{k_j})) - h(f(\bar{x}))]^{\frac{1}{2}} \tag{3.35}$$

is convergent. It follows from (3.1) that

$$\sum_{j=j_2}^{\infty} ||x_{k_j} + d_{k_j} - \bar{x}|| \tag{3.36}$$

is convergent. Noticing (3.26), we have

$$\sum_{j=j_2}^{\infty} ||d_{k_j}|| \tag{3.37}$$

is convergent. This contradicts (3.15), which demonstrates that our lemma is valid. □

Theorem 3.1 *If all the condition in lemma 3.3 are satisfied and if $B_k (k = 1, 2, \ldots)$ satisfies (1.11), then*

$$\sum_{k=1}^{\infty} ||d_k|| \tag{3.38}$$

is convergent.

Proof Since $h(f)$ is bounded below, we have

$$\sum{}' [h(f(x_k)) - h(f(x_k + d_k))]/[h(f(x_k)) - h(f(\bar{x}))]^{\frac{1}{2}} \tag{3.39}$$

is convergent (Powell, 1975). By (1.8), (3.6) and (3.7) we have that

$$\sum{}' \frac{1}{2}\bar{x} \min\{\Delta_k, ||x_k - \bar{x}||, \alpha_0 D_k / c_8^2 ||b_k||\} [h(f(x_k)) - h(f(\bar{x}))]^{\frac{1}{2}} / ||x_k - \bar{x}|| \tag{3.40}$$

is convergent, so (3.1) implies that

$$\sum{}' \min\{\Delta_k, ||x_k - \bar{x}||, \alpha_0 D_k / c_8^2 ||B_k||\} \tag{3.41}$$

is convergent. Therefore, since α_0 and c_8 are constants, we deduce that (3.12) is convergent. Hence, by lemma 3.3, (3.13) is convergent. Since $B_k (k = 1, 2, \ldots)$ satisfies (1.11),

$$\sum_k{}' ||d_k|| / (1 + \sum_{i=1}^{k} \Delta_i) \tag{3.42}$$

is convergent. Because $\Delta_i \leq c_1 ||d_{i-1}||$, it follows that

$$\sum_k{}' ||d_k|| / (1 + \sum_{i=1}^{k} ||d_i||) \tag{3.43}$$

is also convergent. By using an argument that is similar to the derivation of (2.16), we have

$$\sum_k{}' ||d_k|| / (1 + \sum_{i=1}^{k}{}' ||d_i||) \tag{3.44}$$

is convergent. Hence by the arguments of Powell (1975), we have

$$\sum_{k=1}^{\infty} ||d_k|| \tag{3.45}$$

is convergent. □

Because inequality (3.1) and lemma 3.1 imply that the conditions of corollary 2.3 hold, we have the following result.

Corollary 3.4 *Under the conditions of Theorem 3.4,* $||B_k||(k = 1, 2, ...)$ *are uniformly bounded and* $\{x_k\}$ *converges to* x^*.

Proof This follows directly from theorem 3.4, corollary 2.3 and $\Delta_i \leq c_1 ||d_{i-1}||$ for $i > 1$. □

4. Minimax Problem

In this section and the next one, we consider minimax and L_1 problems respectively. In each case it is proved that the two conditions given at the beginning of Section 3 are satisfied under strict complementarity and second order sufficiency. Hence convergence results follow directly.

Throughout this section, we consider the case when

$$h(f(x)) = ||f(x)||_{\infty}. \tag{4.1}$$

When analyzing the minimax problem, strict complementarity and second order sufficiency conditions are often needed, for example, see Han (1978) and Powell (1983). We also assume these conditions and that $f_i(x)(i = 1, 2, ..., m)$ are all twice continuously differentiable, that x^* is the solution of (1.1), and (without loss of generality) that

$$f_i(x^*) = ||f(x^*)||_{\infty} > 0. \tag{4.2}$$

Therefore there exist unique positive Lagrange multipliers λ_i^* $(i = 1, 2, ..., m)$ such that

$$\sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) = 0 \tag{4.3}$$

$$\sum_{i=1}^{m} \lambda_i^* = 1 \tag{4.4}$$

and, if d is any nonzero vector such that

$$(\nabla f(x^*)^T)^T d = 0 \tag{4.5}$$

then

$$d^T W_{\infty} d > 0 \tag{4.6}$$

where

$$\nabla f(x^*)^T = (\nabla f_1(x^*) \dots \nabla f_m(x^*)) \tag{4.7}$$

and

$$G_\infty = \sum_{i=1}^{m} \lambda_i^* \nabla^2 f_i(x^*). \tag{4.8}$$

Throughout this section, we assume that all these conditions hold. First we have the following lemma

Lemma 4.1 *The inequality*

$$\max_{1 \le i \le m} (\nabla f_i(x^*))^T d \le 0 \tag{4.9}$$

is equivalent to

$$(\nabla f(x^*)^T)^T d = 0. \tag{4.10}$$

Proof Obviously (4.10) implies (4.9). Assume (4.9) holds. If (4.10) fails, then

$$\min_{1 \le i \le m} (\nabla f_i(x^*))^T d < 0. \tag{4.11}$$

Hence, since $\lambda_i^* > 0$ for $i = 1, 2, ..., m$,

$$0 = \left(\sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) \right)^T d = \sum_{i=1}^{m} \lambda_i^* (\nabla f_i(x^*))^T d < 0. \tag{4.12}$$

This is a contradiction. Thus (4.10) holds. □

Lemma 4.2 *$h(f)$ satisfies Condition (I) at x^*.*

Proof If the lemma is invalid, then there exists $\hat{x}_k$, $(k = 1, 2, ...)$ such that

$$\hat{x}_k \to x^* \tag{4.13}$$

and

$$\lim_{k \to \infty} [h(f(\hat{x}_k)) - h(f(x^*))]/||\hat{x}_k - x^*||^2 = 0. \tag{4.14}$$

Without loss of generality, we assume

$$\lim_{k \to \infty} (\hat{x}_k - x^*)/||\hat{x}_k - x^*|| = \hat{d}_0. \tag{4.15}$$

It is elementary that

$$\lim_{k \to \infty} [h(f(\hat{x}_k)) - h(f(x^*))]/||\hat{x}_k - x^*||^2 = \max_{1 \le i \le m} (\nabla f_i(x^*))^T \hat{d}_0. \tag{4.16}$$

By lemma 4.1, (4.16), we have

$$(\nabla f(x^*)^T)^T \hat{d}_0 = 0. \tag{4.17}$$

Since the following inequality

$$\begin{aligned}
h(f(x)) \quad - \quad & h(f(x^*)) \geq \sum_{i=1}^{m} \lambda_i^* [f_i(x) - f_i(x^*)] \\
= \quad & \sum_{i=1}^{m} \lambda_i^* [(x-x^*)^T \nabla f_i(x^*) + \frac{1}{2}(x-x^*)^T \nabla^2 f_i(x^*)(x-x^*)] + o(||x-x^*||^2) \\
= \quad & \frac{1}{2}(x-x^*)^T G_\infty (x-x^*) + o(||x-x^*||^2) \qquad (4.18)
\end{aligned}$$

holds when x close to x^*, we have, by (4.14) and (4.15),

$$\hat{d}_0^T G_\infty \hat{d}_0 \leq 0. \qquad (4.19)$$

This is a contradiction to second order sufficiency because of (4.17). Hence the lemma is valid. □

Lemma 4.3 *$h(f)$ satisfies Condition (II) at x^* with respect to any $\alpha_0 \in (0,1)$.*

Proof If the lemma is invalid, then there exist $\alpha_1 \in (0,1)$ and $\tilde{x}_k (k = 1, 2, ...)$ such that

$$\tilde{x}_k \to x^* \qquad (4.20)$$

and

$$Max_{||\tilde{x}_k - x^*||}(\tilde{x}_k) < \alpha_1 [h(f(\tilde{x}_k)) - h(f(x^*))]. \qquad (4.21)$$

Without loss of generality, we assume $||.||$ is the 2-norm and

$$\lim_{k\to\infty} (\tilde{x}_k - x^*)/||\tilde{x}_k - x^*|| = \tilde{d}_\infty. \qquad (4.22)$$

Since

$$Max_{||x-x^*||}(x) \geq [h(f(x)) - h(f(x^*))] + O(||x-x^*||^2), \qquad (4.23)$$

it follows from (4.21) that

$$[h(f(\tilde{x}_k)) - h(f(x^*))] = O(||\tilde{x}_k - x^*||^2). \qquad (4.24)$$

Thus

$$\limsup_{k\to\infty} [f_i(\tilde{x}_k) - f_i(x^*)]/||\tilde{x}_k - x^*||^2 < +\infty \qquad (4.25)$$

for all $i = 1, 2, ..., m$. Noticing

$$\begin{aligned}
\sum_{i=1}^{m} \lambda_i^* [f_i(x) - f_i(x^*)]/||x-x^*||^2 \quad & = \quad \frac{1}{2}(x-x^*)^T G_\infty (x-x^*)/||x-x^*||^2 \\
& \to \quad O(1) \qquad (4.26)
\end{aligned}$$

and that $\lambda_i^* > 0$ for all i, using (4.25), we have

$$\liminf_{k\to\infty}[f_i(\tilde{x}_k) - f_i(x^*)]/||\tilde{x}_k - x^*||^2 > -\infty \tag{4.27}$$

for all $i = 1, 2, ..., m$. So, by replacing $\{\tilde{x}_k\}$ by a subsequence if necessary, we assume

$$\lim_{k\to\infty}[f_i(\tilde{x}_k) - f_i(x^*)]/||\tilde{x}_k - x^*||^2 = a_i \tag{4.28}$$

exist for all $i = 1, 2, ..., m$. Since

$$f_i(x) - f_i(x^*) = (\nabla f_i(x^*))^T(x - x^*) + \frac{1}{2}(x - x^*)^T\nabla^2 f_i(x^*)(x - x^*) + o(||x - x^*||^2), \tag{4.29}$$

we have

$$\lim_{k\to\infty}(\nabla f_i(x^*))^T(\tilde{x}_k - x^*)/||x - x^*||^2 = a_i - \frac{1}{2}\tilde{d}_\infty^T\nabla^2 f_i(x^*)\tilde{d}_\infty. \tag{4.30}$$

Consequently,

$$(\nabla f(x^*))^T(\tilde{x}_k - x^*) = O(||\tilde{x}_k - x^*||^2). \tag{4.31}$$

Noticing that

$$f_i(x) - f_i(x^*) = (\nabla f_i(x^*))^T\frac{x - x^*}{2} + (\nabla f_i(x))^T\frac{x - x^*}{2} + o(||x - x^*||^2), \tag{4.32}$$

we have

$$\begin{aligned} Max_{||\tilde{x}_k - x^*||}(\tilde{x}_k) &= h(f(\tilde{x}_k)) - \min_{||d||\le||\tilde{x}_k - x^*||} h(f(\tilde{x}_k) + (\nabla f(\tilde{x}_k))^T d) \\ &= h(f(\tilde{x}_k)) - \min_{||d||\le||\tilde{x}_k - x^*||} h(f(\tilde{x}_k) + (\nabla f(x^*))^T\frac{\tilde{x}_k - x^*}{2} \\ &\quad +(\nabla f(\tilde{x}_k))^T[d + \frac{x_k - x^*}{2}]) + o(||\tilde{x}_k - x^*||^2) \\ &= h(f(\tilde{x}_k)) - \min_{||d||\le||\tilde{x}_k - x^*||/2} h(f(x^*) + (\nabla f(x^*))^T\frac{x_k - x^*}{2} \\ &\quad -(\nabla f(\tilde{x}_k))^T d) + o(||\tilde{x}_k - x^*||^2) \\ &= h(f(\tilde{x}_k)) - \min_{||d||\le||\tilde{x}_k - x^*||} h(f(x^*) + \frac{1}{2}(\nabla f(x^*))^T(\tilde{x}_k - x^*) \\ &\quad -(\nabla f(\tilde{x}_k))^T d) + o(||\tilde{x}_k - x^*||^2). \end{aligned} \tag{4.33}$$

Let u_k be the vector defined by

$$u_k = ((\nabla f(x^*))^T)^+(\nabla f(x^*))^T(\tilde{x}_k - x^*), \tag{4.34}$$

where $((\nabla f(x^*))^T)^+$ is the Moore-Penrose generalized inverse of $(\nabla f(x^*))^T$ (see Stewart (1973)). Equation (4.31) and (4.34) imply

$$||u_k|| = O(||\tilde{x}_k - x^*||^2) \tag{4.35}$$

and

$$(\nabla f(x^*))^T u_k = (\nabla f(x^*))^T (\tilde{x}_k - x^*), \tag{4.36}$$

hence we have

$$(\nabla f(\tilde{x}_k))^T u_k - (\nabla f(x^*))^T (\tilde{x}_k - x^*) = o(||\tilde{x}_k - x^*||^2). \tag{4.37}$$

Therefore, because (4.35) allows $d = u_k$ in (4.33) for large k, it follows that

$$Max_{||\tilde{x}_k - x^*||} \geq h(f(\tilde{x}_k)) - h(f(x^*)) + o(||\tilde{x}_k - x^*||^2). \tag{4.38}$$

Hence (4.21) contradicts lemma 4.2, which verifies our lemma. □

Theorem 4.4 *If f satisfies all the conditions stated in the beginning of the section, and if $B_k (k = 1, 2, ...)$ satisfies (1.11), then there exists a neighbourhood of x^* such that if $\{x_k\}$ are calculated by the methods stated in section 1, and if x_k is in the neighbourhood of x^* for sufficiently large k, then*

$$\sum_{k=1}^{\infty} ||x_k - x^*|| \tag{4.39}$$

is convergence. Also $\{x_k\}$ converges to x^*, and $||B_k||$ is bounded uniformly.

Proof This follows directly from Lemmas 4.2 and 4.3, Theorem 3.4 and Corollary 3.5. □

5. L_1 Problem

In this section, we consider the case when

$$h(f(x)) = ||f(x)||_1. \tag{5.1}$$

We assume that $f_i(x)(i = 1, 2, ..., m)$ are all twice continuously differentiable, that x^* is the solution of (1.1) and that

$$\begin{aligned} f_i(x^*) &= 0 \qquad i \in I_1 \subset \{1, ..., m\} \\ f_i(x^*) &\neq 0 \qquad i \in \{1, ..., m\} \backslash I_1. \end{aligned} \tag{5.2}$$

As in the minimax problem, we assume the strict complementarity and second order sufficiency conditions. Therefore, there exist unique Lagrange multipliers $\{u_i^*; i = 1, ..., m\}$ such that

$$\begin{gathered} u_i^* = sign(f_i(x^*)), \quad i \notin I_1 \\ \sum_{i=1}^{m} u_i^* \nabla f_i(x^*) = 0 \end{gathered} \tag{5.3}$$

and

$$-1 < u_i^* < 1, \quad i \in I_1 \tag{5.4}$$

and, if d is any nonzero vector such that

$$(\nabla f_i(x^*)^T d = 0 \quad \forall i \in I_1, \tag{5.5}$$

then

$$d^T G_1 d > 0 \tag{5.6}$$

where

$$G_1 = \sum_{i=1}^{m} u_i^* \nabla^2 f_i(x^*). \tag{5.7}$$

Throughout this section, we assume that all these conditions hold. For convenience of notation, without loss of generality, we assume that

$$f_i(x^*) > 0 \quad \forall i \notin I_1. \tag{5.8}$$

Lemma 5.1 *$h(f(x))$ satisfies Condition (I) at x^*.*

Proof If the lemma is invalid, then there exists $\bar{x}_k (k = 1, 2, ...)$ such that

$$\bar{x}_k \to x^* \tag{5.9}$$

$$\lim_{k\to\infty} [h(f(\bar{x}_k)) - h(f(x^*))]/||\bar{x}_k - x^*||^2 = 0 \tag{5.10}$$

and

$$\lim_{k\to\infty} (\bar{x}_k - x^*)/||\bar{x}_k - x^*|| = d_0'. \tag{5.11}$$

By (5.8) and (5.10), we have

$$\sum_{i\in I_1} |(\nabla f_i(x^*))^T d_0'| + \sum_{i\notin I_1} (\nabla f_i(x^*))^T d_0' = 0, \tag{5.12}$$

and by (5.3) we have

$$\sum_{i=1}^{m} u_i^* (\nabla f_i(x^*))^T d_0' = 0, \tag{5.13}$$

which gives the equation

$$\sum_{i\in I_1} |(\nabla f_i(x^*))^T d_0'| [1 - sign((\nabla f_i(x^*))^T d_0') u_i^*] = 0. \tag{5.14}$$

Thus, since $|u_i^*| < 1$ for all $i \in I_1$, we have

$$(\nabla f_i(x^*))^T d_0' = 0 \quad \forall i \in I_1. \tag{5.15}$$

Now the inequality

$$h(f(x)) - h(f(x^*)) \geq \sum_{i=1}^{m} u_i^*[f_i(x) - f_i(x^*)]$$
$$= \frac{1}{2}(x-x^*)^T G_1 (x-x^*) + o(||x-x^*||^2) \quad (5.16)$$

holds when x close to x^*. Therefore (5.10) and (5.11) imply

$$(d_0')^T G_1 d_0' \leq 0 \quad (5.17)$$

which, because of (5.15), contradicts (5.6). Hence the lemma is valid. □

Lemma 5.2 *$h(f(x))$ satisfies Condition (II) at x^* with respect to any $\alpha_0 \in (0,1)$.*

Proof If the lemma is invalid, then there exist $\alpha_2 \in (0,1)$ and x_k' $(k = 1, 2, ...)$ such that

$$x_k' \to x^* \quad (5.18)$$

and

$$Max_{||x_k' - x^*||}(x_k') < \alpha_2[h(f(x_k')) - h(f(x^*))]. \quad (5.19)$$

Without loss of generality, we assume $||.||$ is the 2-norm and

$$\lim_{k\to\infty}(x_k' - x^*)/||x_k' - x^*|| = \tilde{d}_1. \quad (5.20)$$

As in (4.24), we have

$$h(f(x_k')) - h(f(x^*)) = O(||x_k' - x^*||^2), \quad (5.21)$$

and as in (5.15) we have

$$(\nabla f_i(x^*))^T \tilde{d}_1 = 0 \quad \forall i \in I_1, \quad (5.22)$$

which implies

$$\sum_{i \notin I_1} (\nabla f_i(x^*))^T \tilde{d}_1 = 0. \quad (5.23)$$

By our assumptions, there exist k_4 such that, for $k > k_4$,

$$h(f(x_k')) - h(f(x^*)) = \sum_{i\notin I_1}[f_i(x_k') - f_i(x^*)]$$
$$+ \sum_{i \in I_1} |f_i(x_k')|. \quad (5.24)$$

Moreover, some $u_i^* = 1$ for $i \notin I_1$, (5.3) implies

$$\lim[\sum_{i \notin I_1}(f_i(x_k') - f_i(x^*)) + \sum_{i \in I_1} u_i^* f_i(x^*)]/||x_k' - x^*||^2$$

$$= \frac{1}{2}\tilde{d}_1^T G_1 \tilde{d}_1. \tag{5.25}$$

Equations (5.21), (5.24) and (5.25) and strict complementarity imply the conditions

$$f_i(x_k') = O(||x_k' - x^*||^2) \quad \forall i \in I_1, \tag{5.26}$$

and

$$\sum_{i \notin I_1}(f_i(x_k') - f_i(x^*)) = O(||x_k' - x^*||^2). \tag{5.27}$$

Define

$$\phi(x) = \sum_{i \notin I_1} f_i(x). \tag{5.28}$$

As in the proof of lemma 4.3, we have

$$(\nabla f_i(x^*))^T(x_k' - x^*) = O(||x_k' - x^*||^2) \quad \forall i \in I_1, \tag{5.29}$$

and

$$(\nabla \phi(x^*))^T(x_k' - x^*) = O(||x_k' - x^*||^2). \tag{5.30}$$

Hence there exists $u_k' \in \Re^n$ satisfying

$$u_k' = O(||x_k' - x^*||^2 \tag{5.31}$$

and

$$(\nabla \tilde{f}(x^*))^T(x_k' - x^*) = (\nabla \tilde{f}(x^*))^T u_k', \tag{5.32}$$

where $\tilde{f}(x)$ is a map from $\Re^n$ to $\Re^{|I_1|+1}$, whose components are $f_i(x)$ $(i \in I_1)$ and $\phi(x)$. Therefore as in (4.38), we have

$$Max_{||x_k' - x^*||}(x_k') \geq h(\tilde{f}(x_k')) - h(f(x^*)) + o(||x_k' - x^*||^2). \tag{5.33}$$

Since $h(\tilde{f}(x_k')) = h(f(x_k'))$ for large k, it follows from (5.19) that

$$h(f(x_k')) - h(f(x^*)) = o(||x_k' - x^*||), \tag{5.34}$$

which contradicts lemma 5.1. Hence the lemma is true. □

Theorem 5.3 *If $f(x)$ satisfies all the conditions stated in the beginning of the section, if B_k $(k = 1, 2, ...)$ satisfy (1.11), then there exists a neighbourhood of x^* such that, if $\{x_k\}$ are calculated by the methods stated in Section 1, and if x_k is in this neighbourhood of x^* for sufficiently large k, then*

$$\sum_{k=1}^{\infty} ||x_k - x^*|| \tag{5.35}$$

is convergent. Also, $\{x_k\}$ *converges to* x^**, and* $||B_k||$ $(k = 1, 2, ...)$ *are uniformly bounded*

Proof This follows directly from Lemmas 5.1 and 5.2, Theorem 3.4 and Corollary 3.5. □

6. Discussion

Since minimizing a smooth function $f(x)$ from $\Re^n$ to $\Re$ is the same as minimizing $||F(x)+c||$ for some constant c if $F(x)$ is bounded below, our results are applicable to the smooth case. Indeed it is elementary that (3.1) and (3.2) hold in this case if $\nabla F(\bar{x}) = 0$ and $\nabla^2 F(\bar{x})$ is positive definite, where $h(.) = ||.||$ and $f(.) = F(.) + c$. Hence our results are a generalization of Powell's results (1975). Specific updating schemes for the matrices B_k are available such that (1.11) holds (see, Powell, 1975, for example), and a fast rate of convergence is expected. However, it follows that a general superlinear convergence result can not be proved for nonsmooth $h(.)$ without additional conditions. An extension of our work to questions of superlinear convergence will be the subject of another paper.

7. Acknowledgement

The author is greatly indebted to Professor M.J.D. Powell for his constant help and encouragement, and for his studying the early drafts of this paper. He made many important corrections and many helpful suggestions which improved the paper greatly.

References

[1] Han, S.P. (1978) "Superlinear convergence of a minimax method", Report TR78-336, Computer Sci Dept, Cornell University, USA

[2] Moré, J.J. (1982) "Recent developments in algorithms and softwares for trust region methods", Report AML/MCS-TM-2, 1982.

[3] Powell, M.J.D. (1970) "A new algorithm for unconstrained optimization", in: *Nonlinear Programming*, eds., J.D. Rosen, O.L. Mangasarian and K. Ritter, Academic Press.

[4] Powell, M.J.D. (1975) "Convergence properties of a class of minimization algorithms", in: *Nonlinear Programming II*, eds., O.L. Mangasarian R.R. Meyer and S.M. Robinson, Academic Press.

[5] Powell, M.J.D. (1983) "General algorithm for discrete nonlinear approximation calculation", Report DAMTP/NA2, University of Cambridge.

[6] Sorensen (1982) "Trust region methods for unconstrained optimization", in: *Nonlinear Optimization 1981*, ed., M.J.D. Powell, Academic Press.