Global Convergence 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

A general family of trust region algorithms for nonsmooth optimization is considered. Conditions for convergence are presented that allow a wide range of second derivative approximations. It si noted that the given theory applies to many known trust region methods.

Key words: Trust Region Algorithms, Nonsmooth Optimization,
Convergence.

Technical Report: DAMTP 1983/NA13

1. Introduction

The problem is to minimize a nonsmooth function which has the form $h(f(x))$, where $h(.)$ is a convex function from $\Re^m$ to $\Re$, and $f(x) = (f_1(x), \ldots, f_m(x))^T$ is a continuously differentiable function from $\Re^n$ to $\Re^m$. A trust region algorithm is used to solve this problem. The algorithm is iterative. At the beginning of the $k-$th iteration, x_k, Δ_k and B_k are available, where $x_k \in \Re^n$ is an estimate of the solution of the problem, $\Delta_k > 0$ is a step-bound, and B_k is an $n \times n$ symmetric matrix. A vector d_k is chosen, satisfying $||d_k|| \leq \Delta_k$, such that

$$\phi_k(d_k) \leq \phi_k(0) + c_1[\min_{||d|| \leq \Delta_k} \phi_k(d) - \phi_k(0)], \tag{1.1}$$

where $\phi_k(d)$ is defined as

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

and c_1 is a constant from $(0,1)$. We call the case when d_k is a local minimum of $\min \phi_k(d)$ Case A; otherwise we have Case B. Define

$$x_{k+1} = \begin{cases} x_k + d_k, & \text{if } h(f(x_k)) - h(f(x_k + d_k)) \geq c_2[\phi_k(0) - \phi_k(d_k)] \\ x_k, & \text{if } h(f(x_k)) \leq h(f(x_k + d_k) \end{cases} \tag{1.3}$$

and x_{k+1} is allowed to be either x_k or $x_k + d_k$ in the remaining case when

$$c_2[\phi_k(0) - \phi_k(d_k)] \geq h(f(x_k)) - h(f(x_k + d_k) > 0, \tag{1.4}$$

where $c_2 \in (0,1)$ is a constant. Let $c_3 \in (c_2, 1)$ be any constant. If

$$h(f(x_k)) - h(f(x_k + d_k) < c_3[\phi_k(0) - \phi_k(d_k)], \tag{1.5}$$

then Δ_{k+1} is chosen to satisfy

$$c_4||d_k|| \leq \Delta_{k+1} \leq c_5\Delta_k \quad ; \tag{1.6}$$

otherwise let

$$||d_k|| \leq \Delta_{k+1} \leq c_6\Delta_k \tag{1.7}$$

in Case A and let

$$\Delta_k \leq \Delta_{k+1} \leq c_6\Delta_k \tag{1.8}$$

in Case B, where $c_4 \leq c_5 < 1$ and $c_6 > 1$ are positive constants. We also require that

$$\Delta_k \leq \bar{\Delta} \quad , \tag{1.9}$$

for some positive constant $\bar{\Delta}$. Finally B_{k+1} is chosen for the next iteration.

This description is so general that many known trust region strategies are special cases of our algorithm (for example, Moré, 1982; Powell, 1975,1982,1983; Sorensen, 1982; and Yuan, 1983). When the objective function is smooth ($m = 1$, $h(f) = f$), Powell (1982) proves that the algorithm described in his paper has the following properties. If $\{B_k\}(k = 1, 2, \ldots)$ satisfies the inequality

$$||B_k|| \leq c_7 + c_8 k, \tag{1.10}$$

then "potential convergence" (Steihaug, 1982) is obtained in the sense that

$$\liminf_{k\to\infty} \psi(x_k) = 0, \tag{1.11}$$

where $\psi(x)$ is defined as

$$\psi(x) = h(f(x)) - \min_{||d||\leq 1} h(f(x) + \nabla^T f(x)d) \quad , \tag{1.12}$$

and where c_7 and c_8 are positive constants.

In the following section, it is shown that for our algorithm, if (1.10) is satisfied, then (1.11) holds. Our result is based on the assumption that $\{x_k\}(k = 1, 2, \ldots)$ is bounded. This is usually satisfied, especially when $\{x; h(f(x)) \leq h(f(x_1))\}$ is a bounded set.

By our assumption, there exists a bounded convex set $D \subset \Re^n$ such that $x_k \in D$ for all k. Hence $f(D)$ is a bounded closed set in $\Re^m$. Since $h(.)$ is convex and well defined, there exists a positive constant L such that

$$|h(f_1) - h(f_2)| \leq L||f_1 - f_2|| \quad , \tag{1.13}$$

for all $f_1, f_2 \in f(D)$ (Rockafellar, 1970, P237). By the continuity of $\nabla^T f$, there exists a constant $M > 0$ such that

$$||\nabla^T f(x)|| \leq M \quad , \tag{1.14}$$

for all $x \in D$.

2. The Result

We assume the norm $||.||$ is the 2-norm since any two norms in Euclidean space are equivalent. The following analysis may be generalized by introducing more constants. First we require some lemmas.

Lemma 2.1

$$h(f(x_k)) - \phi_k(d_k) \geq \frac{c_1}{2}\psi_k(x_k) \min\{1, \Delta_k, \psi(x_k)/||B_k||\} \quad , \tag{2.1}$$

where $\psi(x_k)$ is defined by (1.12).

Proof From lemma 6 of Powell (1983), we have that

$$h(f(x_k)) - \min_{||d||\leq\Delta_k} \phi_k(d) \geq \frac{1}{2}\psi_k(x_k) \min\{1, \Delta_k, \psi(x_k)/||B_k||\} \quad . \tag{2.2}$$

Our lemma follows from (2.2) and (1.1). □

Lemma 2.2 *Let δ be any positive constant. If*

$$\psi(x_k) \geq \delta, \tag{2.3}$$

for all k, then there exists a constant $c_9 > 0$ such that

$$h(f(x_k)) - \phi_k(d_k) \geq c_9 \min\{\Delta_k, 1/||B_k||\}. \tag{2.4}$$

Proof From (1.9) and Lemma 2.1, (2.4) holds for $c_9 = \frac{1}{2}\min\{1/\bar{\Delta}, 1, \delta\}\, c_1\delta$. □

Lemma 2.3 *If $||d_k|| < \Delta_k$ and d_k is a local minimum of $\{\phi_k(d); ||d|| \leq \Delta_k\}$, then*

$$||d_k|| \geq \frac{1}{2}\psi(x_k)\min\{1/LM, 1/(1+\bar{\Delta})||B_k||\}. \tag{2.5}$$

Proof Consider the function

$$\bar{\phi}_k(\beta) = \phi_k(d_k + \beta[\bar{d}_k - d_k]) \quad 0 \leq \beta \leq 1, \tag{2.6}$$

where d_k is defined in Section 1 and $\bar{d}_k$ satisfies

$$\psi(x_k) = h(f(x_k)) - h(f(x_k) + \nabla^T f(x_k)\bar{d}_k) \tag{2.7}$$

and $||\bar{d}_k|| \leq 1$. The definition (1.1) shows that $\bar{\phi}_k(\beta)$ is the sum of a term that depends on $h(.)$ and a term that depends on B_k. Using the convexity of $h(.)$, the definition of $\bar{d}_k$, and conditions (1.13) and (1.14), the first of these terms is bounded above by the expression

$$\begin{aligned}
&(1-\beta)h(f(x_k) + \nabla^T f(x_k)d_k) + \beta h(f(x_k) + \nabla^T f(x_k)\bar{d}_k) \\
&\quad = h(f(x_k) + \nabla^T f(x_k)d_k) + \beta[h(f(x_k)) - \psi(x_k) - h(f(x_k) + \nabla^T f(x_k)d_k)] \\
&\quad \leq h(f(x_k) + \nabla^T f(x_k)d_k) + \beta[-\psi(x_k) + LM||d_k||] \quad ,
\end{aligned} \tag{2.8}$$

and the other term satisfies

$$\begin{aligned}
&\frac{1}{2}(d_k + \beta[\bar{d}_k - d_k])^T B_k(d_k + \beta[\bar{d}_k - d_k]) \\
&\qquad \frac{1}{2}d_k^T B_k d_k + \beta||B_k||||d_k||(1+\bar{\Delta}) + \frac{\beta^2}{2}||B_k||(1+\bar{\Delta})^2 \quad .
\end{aligned} \tag{2.9}$$

Thus we deduce the relation

$$\begin{aligned}
\bar{\phi}_k(\beta) \quad &\leq \quad \bar{\phi}_k(0) + \beta[-\psi(x_k) + ||d_k||(LM + ||B_k||(1+\bar{\Delta}))] \\
&+ \quad \frac{\beta^2}{2}||B_k||(1+\bar{\Delta})^2 \quad .
\end{aligned} \tag{2.10}$$

Since $||d_k|| < \Delta_k$ and d_k is a local minimum, $\bar{\phi}_k(\beta)$ does not decrease initially when β is increased from zero. Hence the coefficient of β in (2.10) is non-negative, consequently

$$||d_k|| \geq \psi(x_k)/[LM + (1+\bar{\Delta})||B_k||] \quad . \tag{2.11}$$

Therefore the lemma is valid. $\square$

It is noted that the above lemma reduces to lemma 6 of Powell's (1983) if $||B_k||$ are uniformly bounded and $\psi(x_k)$ is bounded away from zero, and it should pointed out that the proof of the lemma is guided by Powell's lemma 6 (1983).

Lemma 2.4 *If $h(f(x))$ satisfies all the conditions stated in Section 1, and if (2.3) holds for all k, then there exists a positive number c_{10} such that*

$$\Delta_k \geq c_{10}/M_k \tag{2.12}$$

for all k, where M_k is defined by

$$M_k = 1 + \max_{1 \leq i \leq k} ||B_i|| \quad . \tag{2.13}$$

Proof Since $\nabla^T f(x)$ is continuous on D, there exists a $\eta > 0$ such that

$$||f(x) - f(x') - \nabla^T f(x')(x - x')|| \leq \frac{c_9(1-c_3)}{2L}||x - x'||, \quad ||x - x'|| \leq \eta, \tag{2.14}$$

holds for all $x, x' \in D$. We prove the lemma is true when c_{10} has the value

$$c_{10} = \min\{\Delta_1 M_1, c_4 \eta M_1 \eta M_1/2LM, \eta/2(1+\bar{\Delta}), c_4, c_4 c_9(1-c_3)\}. \tag{2.15}$$

Our proof is inductive.

By the definition of c_{10}, (2.12) holds for $k = 1$. We assume (2.12) is true for k and prove it is also true for $k + 1$.

If $||d_k|| \geq \eta$, then $\Delta_{k+1} \geq c_4||d_k|| \geq c_k \eta \geq c_{10}/M_1$, so (2.12) holds for $k + 1$. Therefore for the remainder of the proof we assume $||d_k|| \leq \eta$.

If (1.5) fails then in Case B we have $\Delta_{k+1} \geq \Delta_k \geq c_{10}/M_k \geq c_{10}/M_{k+1}$, so (2.12) holds for $k + 1$. In Case A when (1.5) fails lemma 2.3 gives

$$\begin{aligned} \Delta_{k+1} &\geq ||d_k|| \geq \min\{\Delta_k, \delta/2LM, \delta/2(1+\bar{\Delta})M_k\} \\ &\geq c_{10}/M_k \geq c_{10}/M_{k+1} \quad , \end{aligned} \tag{2.16}$$

so (2.12) holds for $k + 1$.

To complete our proof, we assume $||d_k|| < \eta$, and (1.5) is satisfied. From (2.14) and (1.13),

$$\begin{aligned} h(f(x_k + d_k)) &- h(f(x_k)) = h(f(x_k) + \nabla^T f(x_k)d_k) - h(f(x_k)) \\ &\quad + h(f(x_k + d_k)) - h(f(x_k) + \nabla^T f(x_k)d_k) \\ &\leq h(f(x_k) + \nabla^T f(x_k)d_k) - h(f(x_k)) \\ &\quad + L||f(x_k + d_k) - f(x_k) - \nabla^T f(x_k)d_k|| \\ &\leq h(f(x_k) + \nabla^T f(x_k)d_k) - h(f(x_k)) + \frac{1}{2}c_9(1-c_3)||d_k||. \end{aligned} \tag{2.17}$$

Hence, from (1.5) and (2.17), it follows that

$$(1-c_3)[h(f(x_k+d_k))-h(f(x_k))+\frac{1}{2}c_9||d_k||]\geq\frac{1}{2}c_3d_k^TB_kd_k \quad . \tag{2.18}$$

By adding $(1-c_3)$ times (2.4) and (2.18) and using (1.2), we deduce

$$||d_k||_2^2||B_k||\geq c_9(1-c_3)\min\{||d_k||,2/||B_k||-||d_k||\} \quad . \tag{2.19}$$

If $||d_k||\geq 2/||B_k||-||d_k||$ then $||d_k||\geq 1/||B_k||$, otherwise $||d_k||_2^2||B_k||\geq c_9(1-c_3)||d_k||$. Hence $||d_k||\geq \min\{1,c_9(1-c_3)\}/M_k$. Consequently $\Delta_{k+1}\geq c_4||d_k||\geq c_{10}/M_k\geq c_{10}/M_{k+1}$. This shows (2.12) holds for $k+1$. By induction, our lemma is true. □

From this lemma, we have the following theorem.

Theorem 2.5 *If $h(f(x))$ is bounded below, and satisfies all the conditions stated in Section 1, and if (2.3) holds for all k, then $\sum_{k=1}^{\infty}1/M_k$ is convergent.*

Proof The proof depends on the definition of Δ_{k+1} ((1.5)-1.9)), lemma 2.4 and the fact that $h(f(x))$ is bounded below. Because it is similar to the proof of Powell's (1982) theorem, the details are omitted. □

Corollary 2.6 *If $h(f(x))$ is bounded below, and satisfies all the conditions stated in Section 1, and if (1.10) holds for all k, then (1.11) holds.*

Proof If the corollary is not true, then (2.3) holds for some $\delta>0$, so the conditions of Theorem 2.5 are satisfied, therefore $\sum_{k=1}^{\infty}1/M_k$ is convergent, but this contradicts the bound (1.10). The contradiction proves our corollary. □

3. 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. The author also wish to thank him for many important corrections and many helpful suggestions which made the paper possible.

References

[1] J.J. Moré (1982), "Recent Developments in algorithms and software for trust region methods", ANL/MCS-TM-2. Argonne National Laboratory, Illinois.

[2] M.J.D. Powell (1975), "Convergence properties of a class of minimization algorithms", in: *Nonlinear Programming 2*, eds. O.L. Mangasarian, R.R. Meyer and S.M. Robinson.

[3] M.J.D. Powell (1982), "On the global convergence of trust region algorithms for unconstrained minimization", Report DAMTP 1982/NA7, University of Cambridge.

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

[5] R.T. Rockafellar (1970), *Convex Analysis*. Princeton University Press.

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

[7] T. Steihaug (1982), Discussions on Nonlinear Fitting, in: *Nonlinear Optimization 1981*, ed. M.J.D. Powell, Academic Press.

[8] Y. Yuan (1983), "Some properties of trust region algorithms for nonsmooth optimization", Report DAMTP 1983/NA4, University of Cambridge.