Mathematical Programming 31 (1985) 220-228
North-Holland

CONDITIONS FOR 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, UK

Received 3 December 1983
Revised manuscript received 4 June 1984

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 $h(\cdot)$ is convex and f is a continuously differentiable mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$. Bounds for the second order derivative approximation matrices are discussed. It is shown that Powell's [7, 8] results hold for nonsmooth optimization.

Key words: Trust Region Algorithms, Nonsmooth Optimization, Stationary Points.

1. Introduction

Many papers have been published on trust region algorithms, but most attention has been given to the smooth case, for example see Fletcher [2], Móre [5], Powell [6, 7, 8], Sorensen [12], Steihaug [13] and Toint [14]. Trust region algorithms for nonsmooth optimization are studied by Fletcher [1, 3, 4], and Powell [9]. The problem we want to solve is

$$\min_{x \in \mathbb{R}^n} \quad F(x) \equiv h(f(x)), \tag{1.1}$$

where $h(\cdot)$ is a convex function defined on $\mathbb{R}^n$ and is bounded below; $f(x) = (f_1(x), f_2(x), \ldots, f_m(x))^{\mathrm{T}}$ is a map from $\mathbb{R}^n$ to $\mathbb{R}^m$ and $f_i(x)$ $(i = 1, \ldots, m)$ are all continuously differentiable functions on $\mathbb{R}^n$.

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

$$\min \phi_k(d) \equiv h(f(x) + \nabla^{\mathrm{T}} f(x) d) + \tfrac{1}{2} d^{\mathrm{T}} B_k d \tag{1.2}$$

subject to

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

Here $\|\cdot\|$ may be any norm in $\mathbb{R}^n$ space. Since any two norms in a Euclidean space are equivalent, without loss of generality throughout this paper we assume that $\|\cdot\|$

is $\|\cdot\|_2$. Let

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

It is noted that our choice of x_{k+1} is different from Powell's [9]. Because our condition for letting $x_{k+1} = x_k + d_k$ is weaker, our algorithms let $x_{k+1} = x_k + d_k$ more often, so we have the desirable property of accepting any trial vector of variables that reduces the objective function.

Let Δ_{k+1} satisfy

$$\|d_k\| \leqslant \Delta_{k+1} \leqslant \min\{c_1\Delta_k, \bar{\Delta}\} \tag{1.5}$$

if

$$F(x_k) - F(x_{k+1}) \geqslant c_2[F(x_k) - \phi_k(d_k)], \tag{1.6}$$

otherwise let

$$c_3\|d_k\| \leqslant \Delta_{k+1} \leqslant c_4\Delta_k, \tag{1.7}$$

where c_i $(i = 1, 2, 3, 4)$ are positive constants satisfying $c_1 \geqslant 1$, $c_2 < 1$ and $c_3 \leqslant c_4 < 1$, and where $\bar{\Delta}$ is a positive constant which can be taken equal to the diameter of D (D will be defined below). Our theory applies to several techniques for generating $\{B_k\}$.

Fletcher [3] proves that if x_k $(k = 1, 2, \ldots)$ are all in a bounded set and if B_k is the Hessian of the Lagrange function at the kth iteration, then there exists an accumulation point x^* of the algorithm at which first order condition holds, which means,

$$\max_{\lambda \in \partial h^*} \lambda^{\mathrm{T}} \nabla^{\mathrm{T}} f(x^*) d \geqslant 0 \quad \text{for all } d \in \mathbb{R}^n. \tag{1.8}$$

He also points out that the above result holds for a quasi-Newton method as long as $\|B_k\|$ is bounded above. However, for many updating methods one can easily prove that

$$\|B_k\| \leqslant c_5 + c_6 \sum_{i=1}^{k} \Delta_i \tag{1.9}$$

(see Powell [7]), or that

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

yet the boundedness of $\{\|B_k\|\}$ is not explicit. Our main result is to show that Fletcher's result (1.8) holds if (1.9) or (1.10) is satisfied for all k.

Throughout this paper we assume that $\{x_k\}$ $(k = 1, 2, \ldots)$ is bounded, which is usually satisfied, especially when $\{x;\, h(f(x)) \leqslant h(f(x_1))\}$ is a bounded set. Hence there exists a compact convex closed set $D \subset \mathbb{R}^n$ such that $x_k \in D$ for all k. Since

$h(\cdot)$ is convex and well defined, there exists a positive constant L such that

$$|h(f_1)-h(f_2)| \leq L\|f_1-f_2\|, \tag{1.11}$$

for all $f_1, f_2 \in f(D)$ (Rockafellar, [10, p. 237]). By the continuity of $\nabla^{\mathrm{T}} f$, there exists a constant $M>0$ such that

$$\|\nabla^{\mathrm{T}} f(x)\| \leq M, \tag{1.12}$$

for all $x \in D$.

2. Stationary points

For the simplification of notation, we denote

$$\begin{aligned} &\chi(x; d)=h(f(x))-h(f(x)+\nabla^{\mathrm{T}} f(x) d), \\ &\psi_r(x)=\max_d\{\chi(x; d) \mid \|d\| \leq r\}, \\ &\mathrm{DF}(x; d)=\sup_\lambda\{\lambda^{\mathrm{T}} \nabla^{\mathrm{T}} f(x) d \mid \lambda \in \partial h(f(x))\}, \end{aligned} \tag{2.1}$$

where $\partial h(f(x))$ be the subgradient of $h(\cdot)$, evaluated at $f(x)$. x^* is called a stationary point of $h(f(x))$ if

$$\mathrm{DF}(x^*; d) \geq 0 \quad \text{for all } d \in \mathbb{R}^n, \tag{2.2}$$

which is the same as the first order condition of Fletcher [3]. The following results are elementary results in convex analysis (see Rockafellar [10, 11]).

Lemma 2.1. (i) $\mathrm{DF}(x; d)$ *exists for all* x *and* d;

(ii) $\chi(x; \cdot)$ *is convex, given* $d \in \mathbb{R}^n$, *its directional derivative in the direction* d, *evaluated at* $d^*=0$, *is* $\mathrm{DF}(x; d)$;

(iii) $\psi_r(x) \geq 0$ *for any* $r>0$, *and* $\psi_r(x)=0$ *if and only if* x *is a stationary point of* $h(f(x))$;

(iv) $\psi_r(x)$ *is concave in* r;

(v) $\psi_r(\cdot)$ *is continuous for any given* $r \geq 0$.

By using the above results, one can prove that the condition that there exists an accumulation point x^* of the algorithm at which the first order condition holds is equivalent to the limit

$$\liminf_{k \to \infty} \psi_1(x_k)=0. \tag{2.3}$$

And we also have the following lemma:

Lemma 2.2

$$F(x_k)-\phi_k(d_k) \geq \tfrac{1}{2} \psi_{\Delta_k}(x_k) \min\{1, \psi_{\Delta_k}(x_k) / \|B_k\| \Delta_k^2\}. \tag{2.4}$$

Proof. By the definition (1.2) of d_k, we have

$$F(x_k)-\phi_k(d_k)\geq F(x_k)-\phi_k(d) \quad \text{for all } \|d\|\leq \Delta_k. \tag{2.5}$$

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

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

Then by using the convexity of $h(\cdot)$, remembering that $\|\cdot\|$ is the 2-norm, we have that for all $\alpha\in[0,1]$,

$$F(x_k)-\phi_k(d_k)\geq F(x_k)-\phi_k(\alpha\bar{d}_k)\geq \alpha\psi_{\Delta_k}(x_k)-\tfrac{1}{2}\alpha^2\bar{d}_k^{\mathrm{T}}B_k\bar{d}_k$$
$$\geq \alpha\psi_{\Delta_k}(x_k)-\tfrac{1}{2}\|B_k\|\Delta_k^2\alpha^2.$$

Therefore

$$F(x_k)-\phi_k(d_k)\geq \max_{0\leq\alpha\leq 1}\{\alpha\psi_{\Delta_k}(x_k)-\tfrac{1}{2}\|B_k\|\Delta_k^2\alpha^2\}$$
$$\geq \tfrac{1}{2}\min\{\psi_{\Delta_k}(x_k),[\psi_{\Delta_k}(x_k)]^2/\|B_k\|\Delta_k^2\},$$

which ensures (2.4). □

3. Bounds for B_k

In this section it is shown that the result (1.8) holds if B_k satisfy (1.9), then we establish that the result remains valid if (1.9) is replaced by (1.10). Though the latter result is stronger than the previous one, we still prove both, because the proofs are different.

Theorem 3.1. *If $h(f(x))$ satisfies all the conditions in Section 1, if $\{x_k\}$, generated by the algorithms stated in Section 1, is in a bounded set D, and if all matrices B_k satisfy* (1.9), *then* (1.8) *holds, or in other words, $\{x_k\}$ is not bounded away from stationary points of $h(f(x))$.*

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

$$\psi_1(x_k)>\delta \tag{3.1}$$

for all k. From (iv) of Lemma 2.1, Lemma 2.2, the above inequality and the fact that Δ_k is bounded, we can show that the inequality

$$F(x_k)-\phi_k(d_k)\geq c_9\min\{\Delta_k, 1/\|B_k\|\}, \tag{3.2}$$

holds for some positive constant c_9. Let $\sum'$ denote the sum over the iterations on which (1.6) holds. Then by the fact that $h(\cdot)$ is bounded below, we have

$$\sum_k{}'[F(x_k)-\phi_k(d_k)] \tag{3.3}$$

is convergent. From (3.2) we have

$$\sum_k{}' \Delta_k \Big/ \left(c_5 + c_6 \sum_{i=1}^{k} \Delta_i\right) \tag{3.4}$$

is convergent. By the definition of Δ_k, we have, due to Powell [7],

$$\sum_{i=1}^{k} \Delta_i \leq (1 + c_1/(1-c_4)) \left[\Delta_1 + \sum_{i=1}^{k}{}' \Delta_i\right]. \tag{3.5}$$

Therefore

$$\sum_k{}' \Delta_k \Big/ \left[c_5 + c_6\left(1 + \frac{c_1}{1-c_4}\right) \sum_{i=1}^{k}{}' \Delta_i\right]$$

is also convergent. Hence, we can show that $\sum_k' \Delta_k$ is convergent. Notice that by (3.5), we have that $\sum_{k=1}^{\infty} \Delta_k$ is finite. Consequently, $\|B_k\|$ is uniformly bounded. Then from Fletcher's results [3], (3.1) cannot be satisfied for all k. This is a contradiction, which shows that the theorem is true. □

To prove that the above theorem is still true if (1.9) is replaced by (1.10), we need the following lemmas.

Lemma 3.2. *If* $\|d_k\| < \Delta_k$, *then*

$$\|d_k\| \geq \tfrac{1}{2}\psi_1(x_k) \min\{1/LM, 1/(1+\bar{\Delta})\|B_k\|\}. \tag{3.6}$$

Proof. Consider the function

$$\tilde{\phi}_k(\beta) = \phi_k(d_k + \beta[\bar{d}_k - d_k]), \quad 0 \leq \beta \leq 1, \tag{3.7}$$

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

$$\chi(x_k; \bar{d}_k) = \psi_1(x_k)$$

and $\|\bar{d}_k\| \leq 1$. The definition (1.2) shows that $\tilde{\phi}_k(\beta)$ is the sum of a term that depends on $h(\cdot)$ and a term that depends on B_k. Using the convexity of $h(\cdot)$, the definition of $\bar{d}_k$, and conditions (1.11) and (1.12), the first of these terms is bounded above by the expression

$$\begin{aligned}
&(1-\beta)h(f(x_k) + \nabla^{\mathrm{T}} f(x_k) d_k) + \beta h(f(x_k) + \nabla^{\mathrm{T}} f(x_k)\bar{d}_k) \\
&\quad = h(f(x_k) + \nabla^{\mathrm{T}} f(x_k) d_k) + \beta[h(f(x_k)) - \psi_1(x_k) - h(f(x_k) + \nabla^{\mathrm{T}} f(x_k) d_k)] \\
&\quad \leq h(f(x_k) + \nabla^{\mathrm{T}} f(x_k) d_k) + \beta[-\psi_1(x_k) + LM\|d_k\|],
\end{aligned}$$

and the other term satisfies

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

Thus we deduce the relation

$$\tilde{\phi}_k(\beta) \leq \tilde{\phi}_k(0) + \beta[-\psi_1(x_k) + \|d_k\|(LM + \|B_k\|(1+\bar{\Delta}))] + \frac{\beta^2}{2}\|B_k\|(1+\bar{\Delta})^2.$$

Since $\|d_k\| < \Delta_k$, $\tilde{\phi}_k(\beta)$ does not decrease initially when β is increased from zero. Hence the coefficient of β in (3.7) is nonnegative, consequently

$$\|d_k\| \geq \psi_1(x_k)/[LM + (1+\bar{\Delta})\|B_k\|].$$

Therefore the lemma is valid. □

It is noted that the above lemma reduces to lemma 6 of Powell's [9] if $\|B_k\|$ are uniformly bounded and $\psi_1(x_k)$ is bounded away from zero, and it should be pointed out that the proof of the lemma is guided by that of Powell's lemma 6 [9].

Lemma 3.3. *If $h(f(x))$ satisfies all the conditions stated in Section 1, and if (3.1) holds for all k, then there exists a positive number c_{10} such that*

$$\Delta_k \geq c_{10}/M_k \tag{3.8}$$

for all k, where M_k is defined by

$$M_k = \max_{i \leq k}\{\|B_i\|\} + 1. \tag{3.9}$$

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

$$\|f(x) - f(x') - \nabla^T f(x')(x - x')\| \leq \frac{c_9(1-c_2)}{2L}\|x - x'\| \tag{3.10}$$

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

$$c_{10} = \min\{\Delta_1 M_1, c_4\eta M_1, \delta M_1/2LM, \delta/2(1+\bar{\Delta}), c_4, c_4c_9(1-c_2)\}.$$

Our proof is inductive.

By the definition of c_{10}, (3.8) holds for $k = 1$. We assume (3.8) is true for k, and prove it is also true for $k+1$.

If $\|d_k\| \geq \eta$, then $\Delta_{k+1} \geq c_4\|d_k\| \geq c_4\eta \geq c_{10}/M_1$, so (3.8) holds for $k+1$, since the definition (3.9) indicates that $M_{k+1} \geq M_k$ for all k. Therefore for the remainder of the proof we assume $\|d_k\| < \eta$.

If (1.6) is satisfied, Lemma 3.2 gives

$$\Delta_{k+1} \geq \|d_k\| \geq \tfrac{1}{2}\delta \min\{1/LM, 1/(1+\bar{\Delta})M_k\} \geq c_{10}/M_k \geq c_{10}/M_{k+1},$$

so (3.7) holds for $k+1$.

To complete our proof, we assume $\|d_k\| < \eta$, and (1.6) fails. From (3.10) and (1.11), it follows that

$$\begin{aligned} F(x_k + d_k) - F(x_k) &= h(f(x_k + d_k)) - h(f(x_k) + \nabla^{\mathrm{T}} f(x_k) d_k) - \chi(x_k; d_k) \\ &\leqslant L\|f(x_k + d_k) - f(x_k) - \nabla f(x_k) d_k\| - \chi(x_k; d_k) \\ &\leqslant \tfrac{1}{2} c_9 (1 - c_2)\|d_k\| - \chi(x_k; d_k). \end{aligned}$$

Remembering that (1.6) fails, from the above inequality we can show that

$$(1 - c_2)[\tfrac{1}{2} c_9 \|d_k\| - \chi(x_k; d_k)] \geqslant \tfrac{1}{2} c_2 d_k^{\mathrm{T}} B_k d_k. \tag{3.11}$$

By adding (3.11) and $(1 - c_2)$ times (3.2) and using (1.2) and (2.1), we deduce

$$\|d_k\|^2 \|B_k\| \geqslant c_9 (1 - c_2) \min\{\|d_k\|, 2/\|B_k\| - \|d_k\|\}.$$

If $\|d_k\| \geqslant 2/\|B_k\| - \|d_k\|$ then $\|d_k\| \geqslant 1/\|B_k\|$, otherwise $\|d_k\|^2\|B_k\| \geqslant c_9(1-c_2)\|d_k\|$. Hence $\|d_k\| \geqslant \min\{1, c_9(1-c_2)\}/M_k$. Consequently $\Delta_{k+1} \geqslant c_4\|d_k\| \geqslant c_{10}/M_k \geqslant c_{10}/M_{k+1}$. This shows (3.8) holds for $k+1$. By induction, our lemma is true. □

From this lemma, we have the following result, which and whose proof are due to Powell [8].

Lemma 3.4 (Powell, 1982). *Let $\{\Delta_k\}$ and $\{M_k\}$ be two sequences such that $\Delta_k \geqslant c_{10}/M_k > 0$ for all k, where $c_{10} > 0$ is a positive constant. Let I be a subset of $\{1, 2, \ldots\}$. Assume*

$$\Delta_{k+1} \leqslant c_1 \Delta_k, \quad k \in I,$$

$$\Delta_{k+1} \leqslant c_4 \Delta_k, \quad k \notin I,$$

$$M_{k+1} \geqslant M_k \quad \text{for all } k, \qquad \sum_I \min(\Delta_k, 1/M_k) < \infty,$$

where $c_1 > 1$, $c_4 < 1$ are positive constants. Then the sum

$$\sum_{k=1}^{\infty} 1/M_k < \infty.$$

Proof. Let p be a positive integer such that $c_1 c_4^{p-1}$ 1. Denote $I_k = I \cap \{1, 2, \ldots, k\}$ and $q(k)$ be the number of the elements of I_k. Let $J = \{k; k \leqslant pq(k)\}$ and $J_k = J \cap \{1, 2, \ldots, k\}$. Since M_k does not decrease as k increases, we have that (for details, see Powell [8])

$$\sum_{J_k} 1/M_k \leqslant p \sum_{I_k} 1/M_k,$$

which shows that $\sum_J 1/M_k$ is finite. By the definition of J, we have the inequality

$$\begin{aligned} c_{10}/M_k \leqslant \Delta_k \leqslant \Delta_1 c_1^{q(k-1)} c_4^{(k-1)-q(k-1)} = \Delta_1 c_4^{k-1} \left(\frac{c_1}{c_4}\right)^{q(k-1)} &\leqslant \frac{\Delta_1}{c_4} \left(\frac{c_1}{c_4}\right)^{q(k)} c_4^k \\ = \frac{\Delta_1}{c_4} (c_1 c_4^{p-1})^{k/p} \quad \text{for all } k \notin J, \end{aligned}$$

which shows that the sum $\sum_{k\notin J} 1/M_k$ is also finite. This completes the proof. □

From the above lemmas, it can be shown that

Theorem 3.5. *Theorem* 3.1 *still holds if* (1.9) *is replaced by* (1.10).

Proof. If the theorem is not true, then (3.1) holds for some $\delta > 0$, consequently (3.2) and (3.3) hold. Let I be the set of those k such that (1.6) holds. Then from (3.2), (3.3), Lemma 3.3 and 3.4, it follows that

$$\sum_{k=1}^{\infty} 1/M_k < \infty,$$

which contradicts (1.10). Therefore the theorem is true. □

4. Discussions

The main interest of this paper is investigating bounds on B_k to ensure global convergence (Fletcher [3]). Global convergence result holds if the sum

$$\sum_{k=1}^{\infty} 1/M_k$$

is infinite, where M_k is defined by (3.9), and the condition could not be strengthened (see Powell [8]). Hence our results are a generalization of Powell's results [8].

It would be interesting to investigate relations between the boundedness of $\|B_k\|$ and convergence of the algorithms, since one might ask whether or not the boundedness of $\|B_k\|$ is a technical step towards the more interesting result of superlinear convergence. But, one can easily show that the boundedness of $\|B_k\|$ is not necessary for convergence (not even for superlinear convergence). However, the superlinear convergence ensures that $|d_k^T B_k d_k|/\|d_k\|^2$ is bounded (see Powell [9]).

Updating schemes for the matrices B_k can be obtained by applying updating formulas for smooth optimization (see [7] for example]. The only change we need to make is replacing the gradient of the objective function by that of the approximate Lagrange function. If the approximate Lagrangian multipliers are sufficiently accurate, $\{B_k\}$ can be updated such that (1.9) holds, and a fast rate of convergence is expected. However, due to the Maratos effect, it seems that a general superlinear convergence result cannot be proved for nonsmooth $h(\cdot)$ without other additional conditions. Yuan [15] gives examples of only linearly convergence of trust region algorithms for nonsmooth optimization, and the author believes that second order information should be considered to construct superlinear convergence algorithms. Second order algorithms have been studied by Fletcher [4] and Yuan [16], and the author thinks Fletcher's conjecture [4] that his second order algorithm [4] ensures superlinear convergence is true.

5. Acknowledgements

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 made the paper possible. The author also wishes to thank the referees for their helpful corrections and suggestions.

References

[1] R. Fletcher, "A model algorithm for composite NDO problems", *Mathematical Programming Studies* 17 (1982) 67–76.

[2] R. Fletcher, *Practical methods of optimization, Vol. 1, Unconstrained optimization* (John Wiley & Sons, New York, 1980).

[3] R. Fletcher, *Practical methods of optimization, Vol. 2, Constrained optimization* (John Wiley & Sons, New York, 1981).

[4] R. Fletcher, "Second order correction for nondifferentiable optimization", in: G.A. Watson, ed., *Numerical analysis* (Springer-Verlag, Berlin, 1982).

[5] J.J. Moré, "Recent developments in algorithms and software for trust region methods", in: A. Bachem, M. Grötschel and B. Korte, eds., *Mathematical programming, The state of the art* (Springer-Verlag, Berlin–Heidelberg, 1983).

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

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

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

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

[10] R.T. Rockafellar, *Convex analysis* (Princeton University Press, Princeton, New Jersey, 1970).

[11] R.T. Rockafellar, *The theory of subgradient and its application to problems of optimization: Convex and not convex functions* (Heldermann-Verlag, West Berlin, 1981).

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

[13] T. Steihaug, "The conjugate gradient method and trust regions in large scale optimization", *SIAM Journal of Numerical Analysis* 20 (1983) 626–637.

[14] Ph.L. Toint, "On the superlinear convergence of an algorithm for a sparse minimization problem", *SIAM Journal of Numerical Analysis* 16 (1979) 1036–1045.

[15] Y. Yuan, "An example of only linearly convergence of the trust region algorithm for nonsmooth optimization", Report DAMTP 1983/NA15, University of Cambridge.

[16] Y. Yuan, "On the superlinear convergence of a trust region algorithm for nonsmooth optimization", Report DAMTP 1983/NA16, University of Cambridge.