Journal of Computational Mathematics, Vol.18, No.6, 2000, 621–632.

TWO ALGORITHMS FOR LC^1 UNCONSTRAINED OPTIMIZATION[*1)]

Wen-yu Sun
(*School of Mathematics and Computer Science, Nanjing Normal University, Nanjing 210097, China*)

R.J.B.de Sampaio
(*Applied Informatics Program, Pontifical Catholic University of Parana, Curitiba, PR, 81611-970, Brazil*)

Jin-Yun Yuan
(*Department of Mathematics, Federal University of Parana, Curitiba, PR, 81531-990, Brazil*)

Abstract

In this paper we present two algorithms for LC^1 unconstrained optimization problems which use the second order Dini upper directional derivative. These methods are simple and easy to perform. We discuss the related properties of the iteration function, and establish the global and superlinear convergence of our methods.

Key words: Nonsmooth optimization, Directional derivative, Newton-like method, Convergence, Trust region method.

1. Introduction

The LC^1 optimization problems exist extensively in various optimization problems. For example, the problems from nonlinear complementarity, variational inequality and C^2 nonlinear programming can be formed as LC^1 optimization problems. In addition, LC^1 optimization problems also arise from the extended linear-quadratic programming problems, nonlinear minimax problems, stochastic optimization problems and some semi-infinite programs. See [6] [9] [11] [13] [14] [15] [17] [21].

In this class of problems, the objective function is LC^1 function, but not a C^2 function, i.e., it is continuously differentiable and its derivative is only locally Lipschitzian but not necessary F-differentiable.

The Lipschitz condition plays a vital role in developing generalized Newton methods for solving nonsmooth equations. The research of the Lipschitz condition stimulates us to consider LC^1 opitimization and develop some superlinearly convergent methods for LC^1 optimization problems. Now several authors investigate LC^1 optimization problems. In [6] Hiriat-Urruty etc. considered the second order optimality condition for

[*] Received January 23, 1998.

[1)] This work was supported by CNPq of Brazil and the National Natural Science Foundation of China.

LC^1 optimization. In [11] Qi presented a superlinearly convergent approximate Newton's method for LC^1 optimization. In [16] we discussed generalized Newton's method for LC^1 unconstrained optimization which uses Clarke's generalized Hessian matrix. In [17] we gave a quasi-Newton-SQP method for general LC^1 constrained optimization in which the global convergence and superlinear convergence are established. A convergence result of BFGS method for LC^1 linearly constrained optimization is presented by Chen [2].

The development in this paper is closely related to Qi [11], Sun [16] and Sun [17], in which a similar theory is established for solving LC^1 optimization. As a continuation of our works, in this paper we consider using the second order Dini upper directional derivative and the trust region technique to deal with LC^1 optimization problem and present two methods. These methods are simple and easy to perform.

The organization of this paper is as follows. In the next section, we give some preliminaries which will be used in the whole paper. In Section 3, we set up our direction-finding subproblem for LC^1 optimization, discuss some related properties of the iteration function and give Algorithm 3.3. In Section 4, we establish the global convergence and superlinear convergence of Algorithms 3.3 for LC^1 optimization. In Section 5, we describe an approach by combining the Dini upper directional derivative and trust region technique, and establish its convergence properties.

2. Preliminaries

We first give some definitions that will be used for the remainder of the paper. In [16] we have defined the second order generalized directional derivative of f at x in the direction d as

$$f^{\circ\circ}(x;d) = \limsup_{x'\to x, t\downarrow 0} \frac{f^{\circ}(x'+td;d) - f^{\circ}(x';d)}{t}, \tag{2.1}$$

where $f^{\circ}(x;d)$ denotes the generalized directional derivative of f at x in the direction d. We also discussed some basic properties on $f^{\circ\circ}(x;d)$ in [16]. Now we consider

$$\min f(x),\ x \in R^n,\ f \in LC^1, \tag{2.2}$$

that means the objective function f we want to minimize is differentiable and ∇f is locally Lipschitzian. Therefore, the second order Dini upper directional derivative of the function f at $x_k \in R^n$ in the direction $d \in R^n$ can be defined to be

$$f''_D(x_k;d) = \limsup_{\lambda\downarrow 0} \frac{[\nabla f(x_k+\lambda d) - \nabla f(x_k)]^T d}{\lambda}. \tag{2.3}$$

If ∇f is directionally differentiable at x_k, we have

$$f''_D(x_k;d) = f''(x_k;d) = \lim_{\lambda\downarrow 0} \frac{[\nabla f(x_k+\lambda d) - \nabla f(x_k)]^T d}{\lambda} \tag{2.4}$$

for all $d \in R^n$. From [3] and [16], $f''_D(x_k;d)$ possesses the following basic properties:

1.

$$f''_D(x_k;\lambda d) = \lambda^2 f''_D(x_k;d) \tag{2.5}$$

and

$$f''_D(x_k;d_1+d_2) \le 2[f''_D(x_k;d_1) + f''_D(x_k;d_2)]. \tag{2.6}$$

In addition,

$$|f''_D(x_k;d)| \le K\|d\|^2, \tag{2.7}$$

where K is some scalar. Note that (2.5) and (2.7) are immediate results from the definition. But (2.6) results from the fact that ∇f is Lipschitzian. In this case,

$$f''_D(x_k;d) = f^{\circ\circ}(x_k;d) = \limsup_{t\downarrow 0, x'\to x_k} \frac{[\nabla f(x'+td) - \nabla f(x')]^T d}{t}. \tag{2.8}$$

Therefore we have

$$\begin{aligned}
& f''_D(x_k;d_1+d_2) \\
= \; & \limsup_{t\downarrow 0} \frac{[\nabla f(x_k + t(d_1+d_2)) - \nabla f(x_k)]^T (d_1+d_2)}{t} \\
\le \; & \limsup_{t\downarrow 0} \frac{[\nabla f(x_k + td_1 + td_2) - \nabla f(x_k + td_2)]^T d_1}{t} \\
& + \limsup_{t\downarrow 0} \frac{[\nabla f(x_k + td_1 + td_2) - \nabla f(x_k + td_1)]^T d_2}{t} \\
& + \limsup_{t\downarrow 0} \frac{[\nabla f(x_k + td_1) - \nabla f(x_k)]^T d_2}{t} \\
& + \limsup_{t\downarrow 0} \frac{[\nabla f(x_k + td_2) - \nabla f(x_k)]^T d_1}{t} \\
\le \; & 2[f''_D(x_k;d_1) + f''_D(x_k;d_2)].
\end{aligned}$$

2. $f''_D(x_k;d)$ is upper semi-continuous with respect to $(x_k;d)$, i.e., if $x_i \to x_k$ and $d_i \to d$, then

$$\limsup_{i\to\infty} f''_D(x_i;d_i) \le f''_D(x_k;d). \tag{2.9}$$

3.

$$f''_D(x_k;d) = \max\{d^T V d,\ V \in \partial^2 f(x_k)\}. \tag{2.10}$$

According to Rademacher's thorem ∇f is F-differentiable almost everywhere. Let $D_{\nabla f}$ be the set of points where ∇f is differentiable. The generalized Hessian in the sense of Clarke [3] is defined to be

$$\partial^2 f(x_k) = \mathrm{co}\{\lim_{x_i\to x_k} \nabla^2 f(x_i) : \ x_i \in D_{\nabla f}\},$$

where co stands for the convex hull of all $n\times n$ matrices obtained as limit of sequence of Hessian matrices $\nabla^2 f(x_i)$, and $\partial^2 f(x_k)$ is a nonempty convex compact subset of $R^{n\times n}$. We also call

$$\partial^2_B f(x_k) = \{\lim_{x_i\to x_k} \nabla f(x_i) : \ x_i \in D_{\nabla f}\}$$

the second order B-subdifferential of f at x_k. From mean value theorem 2.6.5 of [3], we have

$$\nabla f(y) - \nabla f(x) \in \mathrm{co}\{\partial^2 f([y,x])(y-x)\}.$$

For the sake of convenience, we also give the following definitions and lemmas in which the definition and property of weak semi-smooth are new, others come from [11] [12] [16].

Definition 2.1. *A function $\nabla f : R^n \to R^n$ is said to be semi-smooth at x if ∇f is locally Lipschitzian at x and the limit*

$$\lim_{h \to d, \lambda \downarrow 0} \{Vh\},\ V \in \partial^2 f(x + \lambda h)$$

exists for any $d \in R^n$.

Definition 2.2. *A function $\nabla f : R^n \to R^n$ is said to be weak semi-smooth at x if ∇f is locally Lipschitzian at x and the limit*

$$\limsup_{h \to d, \lambda \downarrow 0} \{Vh\},\ V \in \partial^2 f(x + \lambda h)$$

exists for any $d \in R^n$.

From the definitions above, there is an obvious result as follows.

Proposition 2.3. *A function f is weak semi-smooth if f is semi-smooth. But, in general, the opposite is not true.*

Lemma 2.4. *If $\nabla f : R^n \to R^n$ is semi-smooth at x, then ∇f is directionally differentiable at x, and for any $V \in \partial^2 f(x + h), h \to 0$, we have*

$$Vh - (\nabla f)'(x; h) = o(\|h\|).$$

Similarly, we have that

$$h^T V h - f''(x; h) = o(\|h\|^2).$$

Furthermore, if $\nabla f : R^n \to R^n$ is weak semi-smooth at x, we have

$$Vh - (\nabla f)'_D(x; h) = o(\|h\|).$$

Lemma 2.5. *If $\nabla f : R^n \to R^n$ is a LC^1 function, then there exist $\lambda \in [0, 1]$ and $V \in \partial^2 f(x + \lambda(y - x))$ such that*

$$f(y) - f(x) - \nabla f(x)^T (y - x) = \frac{1}{2}(y - x)^T V (y - x).$$

In following two propositions we discuss the optimality conditions.

Proposition 2.6. *Let $f : D \subset R^n \to R$ be a LC^1 function on D, where $D \subset R^n$ is an open subset. If x is a solution of LC^1 optimization problem (2.2), then*

$$f'(x; d) = 0 \tag{2.11}$$

and

$$f''_D(x; d) \geq 0, \forall d \in R^n. \tag{2.12}$$

Proof. The result is obvious. In fact, if x is a solution of (2.2), then $f'(x; d) = \nabla f(x)^T d = 0$. Therefore

$$f'(x + td; d) \geq \frac{1}{t}[f(x + td) - f(x)] \geq 0.$$

From the above inequality we have immediately

$$f''_D(x; d) = \limsup_{t \downarrow 0} \frac{1}{t}[f'(x + td, d) - f'(x; d)] \geq 0.$$

Similarly, we can get the sufficiency condition.

Proposition 2.7. *Let $f : D \subset R^n \to R$ be a LC^1 function on D, where $D \subset R^n$ is an open subset. If x satisfies*

$$f'(x; d) = 0 \tag{2.13}$$

and

$$f''_D(x;d) > 0,\ \forall d \neq 0,\ d \in R^n, \tag{2.14}$$

then x is a strict local minimizer of (2.2).

Proof. Suppose that x is not a strict local minimizer of f. Then there exists a sequence $\{x_k\} \subset D, x_k \to x, \forall k$, such that $f(x_k) \leq f(x)$.

Let $x_k = x + t_k d$. From the definition of the second order Dini directional derivative we have

$$f(x_k) - f(x) - t_k \nabla f(x)^T d = \frac{1}{2} t_k^2 f''_D(x;d) + o(\|d\|^2).$$

Since $\nabla f(x) = 0$, it follows that $f''_D(x;d) \leq 0$ which contradicts (2.14).

3. An Algorithm and Related Properties

For solving LC^1 optimization problem (2.2), we present an algorithm which uses the second order Dini upper directional derivative. At the k-th iteration the direction-finding subproblem is

$$\min_{d \in R^n} \Phi_k(d) \equiv \nabla f(x_k)^T d + \frac{1}{2} f''_D(x_k;d), \tag{3.1}$$

where $f''_D(x_k;d)$ stands for the second order Dini upper directional derivative at x_k in the direction d. Since f is LC^1 function, $f''_D(x_k;d)$ always exists for all $d \in R^n$. If L is the Lipschitzian constant of f, it is also the Lipschitzian constant of ∇f. Here, in (3.1), $\Phi_k(\cdot)$ is called an iteration function. It is easy to see that $\Phi_k(0) = 0$ and $\Phi_k(\cdot)$ is Lipschitzian on R^n. The following lemma indicates that $\Phi_k(\cdot)$ is also coercive.

Lemma 3.1. *Assume that $V \in \partial^2 f(x_k)$ satisfies*

$$d^T V d \geq m\|d\|^2,\ m > 0.$$

Then the iteration function $\Phi_k(\cdot)$ is coercive.

Proof. From the assumption, there exists $M > 0$ such that

$$f''_D(x_k;d) = \max_{V \in \partial^2 f(x_k)} < Vd, d >= M\|d\|^2,$$

then

$$-L\|d\| + \frac{1}{2} M\|d\|^2 \leq \nabla f(x_k)^T d + \frac{1}{2} f''_D(x_k;d) \leq L\|d\| + \frac{1}{2} M\|d\|^2.$$

Therefore we immediately have

$$\lim_{\|d\| \to \infty} \Phi_k(d) = \infty.$$

This establishes the coercivity of $\Phi_k(\cdot)$.

The coercivity of $\Phi_k(\cdot)$ assures that the optimal solution of problem (3.1) exists. Besides it also means that d_k is a bounded sequence on R^n which will be shown in the following lemma.

Lemma 3.2. *The direction sequence $\{d_k\}$ is bounded.*

Proof. In fact, if it is not true, we could choose a subsequence $\{d_k :\ k \in K_1 \subset N\}$ with the property that $\|d_k\| > k\ \forall k \in K_1$, where K_1 is a subset of natural number set N. But, from the coercivity, it implies that $\Phi_k(d_k) \to \infty$, which contradicts $\Phi_k(d_k) \leq 0$ for all $k \in N$.

With the above lemmas, we may give the following algorithm for solving LC^1 optimization.

Algorithm 3.3. (Dini Algorithm for LC^1 Optimization)

Step 0. Given $x_0 \in R^n, \epsilon > 0, \rho, \sigma \in (0,1)$. Set $k = 0$.

Step 1. If $\|\nabla f(x_k)\| \leq \epsilon$, stop; otherwise, approximately solve subproblem (3.1) to obtain d_k;

Step 2. Take the smallest nonnegative integer m_k such that

$$f(x_k + \rho^{m_k} d_k) - f(x_k) \leq -\frac{1}{2}\sigma\rho^{m_k} f''_D(x_k; d_k); \tag{3.2}$$

Step 3. Set $x_{k+1} = x_k + \rho^{m_k} d_k$;

Step 4. $k := k + 1$ and go to Step 1.

Remarks.

1. In the algorithm we require $f''_D(x_k; d_k)$ satisfies
$$c_1\|d_k\|^2 \leq f''_D(x_k; d_k) \leq c_2\|d_k\|^2, \tag{3.3}$$
where c_1 and c_2 are the positive constants with $0 < c_1 \leq c_2$. If (3.3) is not satisfied, one replaces $f''_D(x_k; d_k)$ with B_k, a symmetric positive definite matrix satisfying (3.3), for example, $B_k = f''_D(x_k; d_k) + \mu I,\ \mu > 0$.

2. In Step 2, the alternative line-search rule is as follows:
$$f(x_k + \rho^{m_k} d_k) - f(x_k) \leq -\frac{1}{2}\sigma\rho^{m_k}\nabla f(x_k)^T d_k. \tag{3.4}$$
In this case, if $cos\theta_k \geq \beta$, where $\beta > 0$ and θ_k is an angle between $-\nabla f(x_k)$ and d_k, then we also can establish the global convergence of Algorithm 3.3.

3. In practice, we only need to approximately solve subproblem (3.1). For example, we may consider the following inexact version of subproblem (3.1):
$$\|\nabla f(x_k) + (\nabla f)'_D(x_k; d)\| \leq \eta_k\|\nabla f(x_k)\| \tag{3.5}$$
where $0 \leq \eta_k < \eta < 1$ for some number η (see Yuan and Sun [21]).

4. Convergence

In this section we establish the convergence results of Algorithm 3.3. We give the global convergence in Theorem 4.3, and then, in Theorem 4.4, establish the superlinear convergence under the conditions that ∇f is semi-smooth and weak semi-smooth respectively.

Proposition 4.1. *If $d_k \neq 0$ is a solution of (3.1), then for any $\sigma \in (0,1)$ there exists $\bar{\mu} > 0$ such that for all $\mu \in (0, \bar{\mu})$,*

$$f(x_k + \mu d_k) - f(x_k) \leq -\frac{1}{2}\sigma\mu f''_D(x_k; d_k). \tag{4.1}$$

Proof. Suppose by the contrary that there does not exist such a $\bar{\mu} > 0$ with the required property. Then there exists a sequence $\{\lambda_j\} \downarrow 0$ such that for each $j \in N$,

$$f(x_k + \lambda_j d_k) - f(x_k) > -\frac{1}{2}\sigma\lambda_j f''_D(x_k; d_k)$$

and then

$$\frac{f(x_k + \lambda_j d_k) - f(x_k)}{\lambda_j} > -\frac{1}{2}\sigma f''_D(x_k; d_k).$$

Hence we get

$$f'(x_k; d_k) \geq -\frac{1}{2}\sigma f''_D(x_k; d_k) > -\frac{1}{2} f''_D(x_k; d_k). \tag{4.2}$$

But, from the assumption, $d_k \neq 0$ is a solution of (3.1) and $\Phi_k(0) = 0$. Therefore

$$\Phi_k(d_k) = \nabla f(x_k)^T d_k + \frac{1}{2} f''_D(x_k; d_k) \leq 0$$

which means

$$f'(x_k; d_k) \leq -\frac{1}{2} f''_D(x_k; d_k). \tag{4.3}$$

The above inequality (4.3) contradicts (4.2).

Remarks.

1. This proposition indicates that the line-search rule (3.2) in Step 2 of Algorithm 3.3 holds, where $\mu = \rho^{m_k}$.

Lemma 4.2. *The following statements are equivalent:*

1. *$d = 0$ is a globally optimal solution of the problem (3.1);*
2. *0 is the optimum of objective function of problem (3.1);*
3. *the corresponding x_k is a stationary point of the objective function f, i.e., $\nabla f(x_k) = 0$.*

Proof. (1) $\Longrightarrow$ (2): It is obvious.

(2) $\Longrightarrow$ (3): Let 0 be a global optimal value of (3.1), then $\forall \lambda > 0$ and $\forall d \in R^n$,

$$0 \leq \Phi_k(\lambda d) = \lambda \nabla f(x_k)^T d + \frac{1}{2} f''_D(x_k; \lambda d).$$

Dividing both sides of the above inequality by λ, using (2.5) and letting $\lambda \downarrow 0$, we have

$$0 \leq \nabla f(x_k)^T d, \ \forall d$$

which means x_k is a stationary point of f.

(3) $\Longrightarrow$ (1): Let x_k be a stationary point of f, i.e.,

$$\nabla f(x_k)^T d \geq 0, \forall d \in R^n. \tag{4.4}$$

Suppose that $d \neq 0$ is the optimal solution of the problem (3.1). From the property of the iterative function,

$$\nabla f(x_k)^T d \leq -\frac{1}{2} f''_D(x_k; d)$$

which implies

$$\nabla f(x_k)^T d < 0. \tag{4.5}$$

The above two inequalities (4.5) and (4.4) are contradictory.

Now we are in the position to give the global convergence theorem. The proof of this theorem is different from that of the existing theorems of nonsmooth optimization.

Theorem 4.3. *Suppose that $f \in LC^1$ is bounded below on the level set*

$$L(x_0) = \{x \in R^n |\ f(x) \le f(x_0)\}. \tag{4.6}$$

Suppose that there exist constants $c_2 \ge c_1 > 0$ such that for all $d \in R^n$

$$c_1 d^T d \le f''_D(x; d) \le c_2 d^T d. \tag{4.7}$$

Then every accumulation point of the sequence $\{x_k\}$ generated from Algorithm 3.3 is a stationary point of problem (2.2).

Proof. First, from Algorithm 3.3, we know the sequence $\{x_k\}$ is strictly decreasing, i.e., $f(x_{k+1}) < f(x_k)$. In addition, the sequence $\{f(x_k)\}$ is bounded below from the assumption. Therefore $\{f(x_k)\}$ is convergent and we have

$$f(x_{k+1}) - f(x_k) \to 0.$$

Second, from the assumption (4.7) and the mean value theorem, we have that, for all $x \in L(x_0)$, there exists $t \in (0,1)$ such that

$$\begin{aligned} f(x) - f(x_0) &= \nabla f(x_0)^T (x - x_0) + \frac{1}{2} f''_D(x_0 + t(x - x_0); x - x_0) \\ &\ge -\|\nabla f(x_0)\| \|x - x_0\| + \frac{1}{2} c_1 \|x - x_0\|^2. \end{aligned} \tag{4.8}$$

This implies that

$$\|x - x_0\| \le \frac{2}{c_1} \|\nabla f(x_0)\|, \tag{4.9}$$

i.e., the level set $L(x_0)$ is bounded. Since $x_k \in L(x_0)$ from Algorithm 3.3, it follows that the sequence $\{x_k\}$ is bounded. Therefore there exist accumulation points of the sequence $\{x_k\}$.

Finally, let $\bar{x}$ be an arbitrary accumulation point of $\{x_k\}$, but not a stationary point of f. Then, from Lemma 4.2, the corresponding direction vector $\bar{d} \neq 0$. From line-search Armijo rule, m_k is the smallest nonnegative integer such that Armijo rule (3.2) holds. Then the number $m_k - 1$ satisfies the following inequality

$$f(x_k) - f(x_k + \rho^{m_k - 1} d_k) < \frac{1}{2} \sigma \rho^{m_k - 1} f''_D(x_k; d_k). \tag{4.10}$$

Set $\tau_k = \rho^{m_k - 1}$. Dividing both sides in the expression (4.10) by τ_k, passing to limit $k \to \infty$ and using $\lim_{k\to\infty} \tau_k = \lim_{k\to\infty} \rho^{m_k - 1} = 0$, we get

$$\nabla f(\bar{x})^T \bar{d} > -\frac{1}{2} \sigma f''_D(\bar{x}; \bar{d}) > -\frac{1}{2} f''_D(\bar{x}; \bar{d}). \tag{4.11}$$

But, from the property of the iterative function $\Phi_k(\cdot)$, we have

$$\nabla f(\bar{x})^T \bar{d} \le -\frac{1}{2} f''_D(\bar{x}; \bar{d}). \tag{4.12}$$

Therefore we get a contradiction.

In the following, we establish the superlinear convergence results. The following theorem is obtained under the conditions that ∇f is weak semi-smooth and BD-regular at the minimal point x^*, the latter means that all elements in $\partial_B \nabla f(x^*)$ are nonsingular.

Theorem 4.4. *Let $f : D \subset R^n \to R$ be LC^1 function and D an open convex set. Assume that ∇f is weak semi-smooth and BD-regular at $x^* \in D$. Let the sequence $\{x_k\}$ be generated by Algorithm 3.3. Then $\{x_k\}$ converges Q-superlinearly to x^*.*

Proof. Let $e_k = x_k - x^*$ and $d_k = x_{k+1} - x_k$. Then both sequences $\{e_k\}$ and $\{d_k\}$ converge to zero.

From subproblem (3.1), Algorithm 3.3 and Lemma 2.4, we have

$$\nabla f(x_k) + (\nabla f)'_D(x_k; d_k) + o(\|d_k\|) = 0 \tag{4.13}$$

which means

$$\limsup_{k\to\infty} \frac{\|\nabla f(x_k) + (\nabla f)'_D(x_k; d_k)\|}{\|d_k\|} = 0. \tag{4.14}$$

Also,

$$\begin{aligned} & \nabla f(x^*) \\ = \; & [\nabla f(x_k) + (\nabla f)'_D(x_k; d_k)] - [\nabla f(x_k) - \nabla f(x^*) - (\nabla f)'_D(x_k; e_k)] \\ & -(\nabla f)'_D(x_k; e_{k+1}) + o(\|d_k\|) + o(\|e_k\|) + o(\|e_{k+1}\|). \end{aligned} \tag{4.15}$$

The weak semi-smoothness of ∇f and (4.14) imply that the first two square brackets approach to zero as $k \to \infty$. Note also that $(\nabla f)'_D(x_k; e_{k+1}) \to 0$. Therefore, by taking the limsup in both sides of (4.15), we have $\nabla f(x^*) = 0$.

Using the BD-regularity and (4.15), it follows that

$$\lim_{k\to\infty} \frac{\|e_{k+1}\|}{\|e_k\|} = 0 \tag{4.16}$$

which establishes the Q-superlinear convergence of our algorithm.

Immediately, we have the following corollary for semi-smooth function f.

Corollary 4.5. *Let $f : D \subset R^n \to R$ be LC^1 function and D an open convex set. Assume that ∇f is semi-smooth and BD-regular at $x^* \in D$. Let the sequence $\{x_k\}$ be generated by Algorithm 3.3. Then $\{x_k\}$ converges Q-superlinearly to x^*.*

5. Another Algorithm and Its Convergence

Trust region methods are an important class of iterative methods for solving optimization problems. They have been proved to be efficient and robust not only for unconstrained and constrained optimization [10] [1], but also for nonsmooth optimization [22] [21]. It is worth mention that the works of Fletcher [5] and Yuan [22] built an important base for solving nonsmooth optimization by means of the trust region methods. In this section, following this line, we give an approach that combines the trust region technique with subproblem (3.1). At the k-th iteration the direction-finding subproblem now is

$$\begin{aligned} \min \quad & \Phi_k(d) \equiv \nabla f(x_k)^T d + \frac{1}{2} f''_D(x_k; d) \\ \text{s.t.} \quad & \|d\| \le \Delta_k \end{aligned} \tag{5.1}$$

for some bound Δ_k. This is a typical subproblem which the trust region method will solve repeatedly for different pairs (x_k, Δ_k). The norm $\|\cdot\|$ is arbitrary but it is usually chosen as the l_2 norm. About the basic theory and the background of trust region methods, please consult [21].

Now we state this algorithm in which we only solve the subproblem (5.1) approximately. Assume d_k^* and d_k are the exact and an inexact solution of (5.1) respectively. We require that the decrease in $\Phi_k(d_k)$ must be at least a fraction of the optimal decrease in Φ_k, i.e.,

$$f(x_k) - \Phi_k(d_k) \geq \beta_0[f(x_k) - \Phi_k(d_k^*)], \tag{5.2}$$

where $0 < \beta_0 \leq 1$. In the following algorithm, $Ared_k = f(x_k) - f(x_k + d_k)$ and $Pred_k = f(x_k) - \Phi_k(d_k)$ stand for the actual reduction and the predicted reduction in the objective function respectively.

Algorithm 5.1. (Dini-TR Algorithm for LC^1 Optimization)

Step 0. Given $x_0 \in R^n, \epsilon > 0, \Delta_0 > 0, \beta_1 > 1 > \beta_2 > 0, 0 < \beta_3 < \beta_4 < 1, k := 0$.

Step 1. If $\|\nabla f(x_k)\| \leq \epsilon$, stop; otherwise, determine an approximate solution d_k to subproblem (5.1).

Step 2. Compute $r_k = Ared_k / Pred_k$.
Set

$$x_{k+1} = \begin{cases} x_k + d_k, & \text{if } r_k > 0 \\ x_k, & \text{otherwise.} \end{cases} \tag{5.3}$$

Step 3. Update Δ_k and the model Φ_k.

$$\Delta_{k+1} \in \begin{cases} [\beta_3, \beta_4]\Delta_k & \text{if } r_k \leq \beta_2 \\ [1, \beta_1]\Delta_k & \text{if } r_k > \beta_2. \end{cases} \tag{5.4}$$

Step 4. $k := k + 1$, go to Step 1.

The following lemma is necessary to prove the global convergence.

Lemma 5.2.

$$\begin{aligned} pred_k &= \Phi_k(0) - \Phi_k(d_k) \\ &\geq \frac{\beta_0}{2}\|\nabla f(x_k)\| \min\{\Delta_k, \frac{\|\nabla f(x_k)\|}{c_2}\}, \end{aligned} \tag{5.5}$$

where β_0 and c_2 satisfy (5.2) and (3.3).

Proof. From the definition of d_k, for any $\alpha \in [0, 1]$, we have

$$\begin{aligned} & \Phi_k(0) - \Phi_k(d_k) \\ \geq\ & \beta_0[\Phi_k(0) - \Phi_k(d_k^*)] \\ \geq\ & \beta_0 \left[\Phi_k(0) - \Phi_k(-\alpha \frac{\Delta_k}{\|\nabla f(x_k)\|} \nabla f(x_k))\right] \\ =\ & \beta_0[\alpha \Delta_k \|\nabla f(x_k)\| - \frac{1}{2}\alpha^2 \Delta_k^2 f_D''(x_k; d_k)/\|\nabla f(x_k)\|^2] \\ \geq\ & \beta_0[\alpha \Delta_k \|\nabla f(x_k)\| - \frac{1}{2}\alpha^2 \Delta_k^2 c_2], \end{aligned} \tag{5.6}$$

where c_2 is defined in (3.3). Therefore we have

$$\begin{aligned} Pred_k &\geq \beta_0 \max_{0 \leq \alpha \leq 1} [\alpha \Delta_k \|\nabla f(x_k)\| - \frac{1}{2}\alpha^2 \Delta_k^2 c_2] \\ &\geq \frac{\beta_0}{2}\|\nabla f(x_k)\| \min\{\Delta_k, \|\nabla f(x_k)\|/c_2\}. \end{aligned} \tag{5.7}$$

In the following, we establish the global convergence of Algorithm 5.1. Our goal is to prove the sequence $\{x_k\}$ generated from Algorithm 5.1 satisfies $\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0$. Since our optimization problem is LC^1 problem, the conclusion is positive. The following proof is typical and essentially similar to [10] [7] and [21].

Theorem 5.3. *Let $f \in LC^1$ and f be bounded below on R^n. Assume the condition (3.3) holds. Then the sequence $\{x_k\}$ generated from Algorithm 5.1 satisfies*

$$\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0. \tag{5.8}$$

Proof. Assume that the conclusion is not true. Then there exists an $\epsilon > 0$ such that $\nabla f(x_k) \geq \epsilon$ for all k sufficiently large. Let

$$K = \{k : \ r_k \geq \beta_2\}$$

be the set of successful iterations. From Algorithm 5.1 and Lemma 5.2, we have obviously that

$$\begin{aligned} f(x_k) - f(x_k + d_k) &\geq \beta_2[\Phi_k(0) - \Phi_k(d_k)] \\ &\geq \frac{1}{2}\beta_0\beta_2\|\nabla f(x_k)\| \min\{\Delta_k, \|\nabla f(x_k)\|/c_2\}. \end{aligned} \tag{5.9}$$

Since f is bounded below on R^n, we have

$$\sum_{k\in K} \Delta_k < \infty. \tag{5.10}$$

From the above inequality, the updating rules of the trust region in Algorithm 5.1 and (2.1) in Powell [10], we have

$$\sum_{k=1}^{\infty} \Delta_k \leq (1 + \frac{\beta_1}{1-\beta_4}) \sum_{k\in K} \Delta_k < \infty. \tag{5.11}$$

This means

$$\Delta_k \to 0, \forall k. \tag{5.12}$$

But, on the other hand, Lemma 5.2 shows

$$\Phi_k(0) - \Phi_k(d_k) \geq \frac{1}{2}\beta_0 \epsilon \Delta_k. \tag{5.13}$$

Using $f \in LC^1$ and (3.1), we know that there exists a sequence $\{\epsilon_k\}$ converging to zero such that

$$|[f(x_k) - f(x_k + d_k)] - [\Phi_k(0) - \Phi_k(d_k)]| \leq \epsilon_k \Delta_k. \tag{5.14}$$

Dividing the both sides of (5.14) by $\Phi_k(0) - \Phi_k(d_k)$ and passing to the limit yield

$$r_k \to 1 \tag{5.15}$$

which means from Algorithm 5.1 that

$$\Delta_{k+1} \geq \Delta_k, \ \forall k \text{ sufficiently large.} \tag{5.16}$$

Therefore we get a contradiction from (5.16) and (5.12). The global convergence is proved.

Besides, from Theorem 4.4, it is easy to know that Algorithm 5.1 is also superlinearly convergent.

References

[1] R.H. Byrd, R.B. Schnabel, G.A. Shultz, A trust region algorithm for nonlinear constrained optimization, *SIAM J. Numer.Anal.,* **24** (1987), 1152-1170.

[2] X. Chen, Convergence of the BFGS method for LC^1 convex constrained optimization, *SIAM J. Control and Optimization,* **34** (1996), 2051-2063.

[3] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley, New York, 1983.

[4] S. Di, W. Sun, Trust region method for conic model to solve unconstrained optimization, *Optimization Methods and Software,* **6** (1996), 237-263.

[5] R. Fletcher, Practical Methods of Optimization, Second Edition, John Wiley & Sons, 1987.

[6] J.B. Hiriat-Urruty, J.J. Strodoit, V.H. Nguyen, Generalized Hessian matrix and second-order optimality conditions for problems with $C^{1,1}$ data, *Applied Mathematics and Optimization,* **11** (1984), 43-56.

[7] J.J. More, Recent development in algorithms and software for trust region methods, in Mathematical Programming, The State of the Art, Bonn 1982, A.Bachem, M.Grötschel, B.Korte eds., Springer-Verlag, Berlin, 1983, 258-287.

[8] J.S. Pang, S.P. Han, N. Rangaraj, Minimization of locally Lipschitzian functions, *SIAM J. Optimization,* **1** (1991), 57-82.

[9] J.S. Pang, L. Qi, Nonsmooth equations: motivation and algorithms, *SIAM J. Optimization,* **3** (1993), 443-465.

[10] M.J.D.Powell, Convergence properties of a class of minimization algorithms, in Nonlinear Programming 2, O.L.Mangasarian, R.R.Meyer and S.M.Robinson eds., Academic Press, New York, 1975.

[11] L. Qi, Superlinear convergent approximate Newton method for LC^1 optimization problems, *Mathematical Programming,* **64** (1994), 277-294.

[12] L. Qi, J. Sun, A nonsmooth version of Newton method, *Mathematical Programming,* **58** (1993), 353-367.

[13] L. Qi, W. Sun, An iterative method for the minimax problem, in D.Z.Du and P.M.Pardalos eds., Minimax and Applications, Kluwer Academic Publisher, Boston, USA, 1995, 55-67.

[14] R.T. Rockfellar, R.J.B. Wets, Generalized linear-quadratic problems of deterministic and stochastic optimal control in discrete time, *SIAM J. Control and Optimization,* **28** (1990), 810-822.

[15] R.J.B. Sampaio, W. Sun, J. Yuan, On trust region algorithm for nonsmooth optimization, *Applied Mathematics and Computation,* **85** (1997), 109-116.

[16] W. Sun, Generalized Newton's method for LC^1 unconstrained optimization, *Journal of Computational Mathematics*, **13** (1995), 250-258.

[17] W. Sun, Newton's method and quasi-Newton-SQP method for general LC^1 constrained optimization, *Applied Mathematics and Computation,* **92** (1998), 69-84.

[18] W. Sun, On non-quadratic model methods for optimization, *Asia and Pacific Journal of Operations Research,* **13** (1996), 43-63.

[19] W.Sun, On convergence of an iterative method for minimax problem, *Journal of Australian Mathematics Society, Series B,* **39** (1997), 280-292.

[20] W. Sun, R.J.B. Sampaio, J. Yuan, Trust region method with quasi-Newton update for non-smooth least-squares problems, *Applied Mathematics and Computation,* **105** (1999), 183-194.

[21] Y. Yuan, W. Sun, Optimization Theory and Methods, Science Press, Beijing, 1997.

[22] Y. Yuan, Conditions for convergence of trust region algorithms for nonsmooth optimization, *Mathematical Programming,* **31** (1985), 220-228.