A MULTI-STEP CURVE SEARCH ALGORITHM IN NONLINEAR OPTIMIZATION

Nada I. ĐURANOVIĆ-MILIČIĆ

Department of Mathematics, Faculty of Technology and Metallurgy
University of Belgrade, Belgrade, Serbia
nmilicic@imf.bg.ac.yu

Received: November 2006 / Accepted: April 2008

Abstract: In this paper a multi-step algorithm for LC^1 unconstrained optimization problems is presented. This method uses previous multi-step iterative information and curve search to generate new iterative points. A convergence proof is given, as well as an estimate of the rate of convergence.

Keywords: Unconstrained optimization, multi-step curve search, convergence.

1. INTRODUCTION

We shall consider the following LC^1 problem of unconstrained optimization

\[ \min \left\{ f(x) \mid x \in D \subset \mathbb{R}^n \right\}, \]  

where \( f : D \subset \mathbb{R}^n \to \mathbb{R} \) is a LC^1 function on the open convex set \( D \), that means the objective function we want to minimize is continuously differentiable and its gradient is locally Lipschitzian, i.e.

\[ \left| g(y) - g(x) \right| \leq L \| y - x \| \quad \text{for} \ x, y \in D \]

for some \( L > 0 \), where the gradient computed at \( x \) is denoted by \( g(x) \).

We shall present an iterative multi-step algorithm which is based on the algorithms from [1] and [4] for finding an optimal solution to problem (1) generating the sequence of points \( \{x_k\} \) of the following form:

\[ x_{k+1} = x_k + \alpha_k s_k + \alpha_k^2 d_k, \quad k = 0,1,...,s_k \neq 0,d_k \neq 0 \]

where the step-size \( \alpha_k \) and the directional vectors \( s_k \) and \( d_k \) are defined by the particular algorithms.
2. PRELIMINARIES

We shall give some preliminaries that will be used for the remainder of the paper.

**Definition** (see [5]) The second order Dini upper directional derivative of the function $f \in LC^1$ at $x_k \in R^n$ in the direction $d \in R^n$ is defined to be

$$f^+_D(x;d) = \limsup_{\lambda \downarrow 0} \frac{g(x + \lambda d) - g(x)}{\lambda}$$

If $g$ is directionally differentiable at $x_k$, we have

$$f^+_D(x_k;d) = f^+(x_k;d) = \lim_{\lambda \downarrow 0} \frac{g(x + \lambda d) - g(x)}{\lambda}$$

for all $d \in R^n$.

**Lemma 1** (See [5]) Let $f : D \subset R^n \rightarrow R$ 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 (1), then:

$$f'(x;d) = 0$$

and $f^+_D(x;d) \geq 0, \forall d \in R^n$.

**Lemma 2** (See [5]) Let $f : D \subset R^n \rightarrow R$ be a LC$^1$ function on $D$, where $D \subset R^n$ is an open subset. If $x$ satisfies

$$f'(x;d) = 0$$

and $f^+_D(x;d) > 0, \forall d \neq 0, d \in R^n$, then $x$ is a strict local minimizer of (1).

3. THE OPTIMIZATION ALGORITHM

**Algorithm:** $0 < \sigma < 1, 0 < \rho < 1, x_i \in D$, $m$ is a positive integer, $k := 1$.

Step 1. If $\|x_i\| = 0$ then STOP; else go to step 2.

Step 2. $x_{i+1} = x_i + \alpha_i s_i(\alpha_i) + \alpha_i^2 d_i(\alpha_i)$, where $\alpha_i$ is selected by the curve search rule, and $s_i(\alpha_i)$ and $d_i(\alpha_i)$ are computed by the direction vector rules 1 and 2. For simplicity, we denote $s_i(\alpha_i)$ by $s_i$, $d_i(\alpha_i)$ by $d_i$ and $g(x_i)$ by $g_i$.

**Curve search rule:** Choose $\alpha_i = \hat{q}^{(i)}$, $0 < q < 1$, where $i(k)$ is the smallest integer from $i = 0, 1, \ldots$ such that
\[ x_{k+1} = x_k + q^{(k)} s_k + q^{2(k)} d_k \in D \]

and

\[ f(x_k) - f(x_k + q^{(k)} s_k + q^{2(k)} d_k) \geq \sigma \left[ -q^{(k)} g_k s_k + \frac{1}{2} q^{(k)} f_D(x_k; d_k) \right] \]  

(3)

**Direction vector rule 1**

\[ s_k(\alpha) = \begin{cases} s_k^* , & k \leq m-1 \\ - \left( 1 - \sum_{j=1}^{\infty} \alpha_i p_i \right) g_k + \sum_{i=m+1}^{m} \alpha_i p_i s_{k-i+1} \end{cases}, \quad k \geq m, \]

where

\[ p_i = \frac{\rho \|g_i\|^2}{(m-1) \left( \|g_i\|^2 + \|g^*_i s_{k-i+1}\|^2 \right)} , \quad i = 2, 3, ..., m, \]

and \( s_k^* \neq 0, \ k \leq m-1 \) is any vector satisfying the descent property \( g_k^T s_k^* \leq 0 \).

**Direction vector rule 2.** The direction vector \( d_k^* , \ k \leq m-1, \) presents a solution of the problem

\[ \min \left\{ \Phi_k (d) | d \in R^m \right\}, \]

(4)

where

\[ \Phi_k (d) = g_k^T d + \frac{1}{2} f_D (x_k; d), \]

and

\[ d_k(\alpha) = \begin{cases} d_k^* , & k \leq m-1 \\ \sum_{i=2}^{m} \alpha^{i-1} d_{k-i+1}^* , & k \geq m. \end{cases} \]

Step 3. \( k := k+1, \) go to step 1.

We make the following assumptions.

A1. We suppose that there exist constants \( c_2 \geq c_1 > 0 \) such that

\[ c_1 \|d\| \leq f_D(x; d) \leq c_2 \|d\| \]

(5)

for every \( d \in R^m \).

A2. \( \|d_k\| = 1 \) and \( \|s_k\| = 1, \ k = 0, 1, ... \)

It follows from Lemma 3.1 in [5] that under the assumption A1 the optimal solution of the problem (4) exists.
Proposition: If the function \( f \in LC^1 \) satisfies the condition (5), then: 1) the function \( f \) is uniformly and, hence, strictly convex, and, consequently; 2) the level set \( L(x_0) = \{x \in D : f(x) \leq f(x_0)\} \) is a compact convex set; 3) there exists a unique point \( x^* \) such that \( f(x^*) = \min_{x \in L(x_0)} f(x) \).

Proof: 1) From the assumption (5) and the mean value theorem it follows that for all \( x \in L(x_0) \) there exists \( \theta \in (0,1) \) such that
\[
 f(x) - f(x_0) = g(x_0)^T (x-x_0) + \frac{1}{2} f_0'[x_0 + \theta(x-x_0); x-x_0] \\
 \geq g(x_0)^T (x-x_0) + \frac{1}{2} c_1 \|x-x_0\| > g(x_0)^T (x-x_0),
\]
that is, \( f \) is uniformly and consequently strictly convex on \( L(x_0) \).

2) From [3] it follows that the level set \( L(x_0) \) is bounded. The set \( L(x_0) \) is closed because of the continuity of the function \( f \); hence, \( L(x_0) \) is a compact set. \( L(x_0) \) is also (see [6]) a convex set.

3) The existence of \( x^* \) follows from the continuity of the function \( f \) on the bounded set \( L(x_0) \). From the definition of the level set it follows that
\[
 f(x^*) = \min_{x \in L(x_0)} f(x) = \min_{x \in D} f(x)
\]
Since \( f \) is strictly convex it follows from [6] that \( x^* \) is a unique minimizer.

Lemma 3 (See [5]) The following statements are equivalent:
1. \( d = 0 \) is a globally optimal solution of the problem (4);
2. \( 0 \) is the optimum of the objective function of the problem (4);
3. the corresponding \( x_k \) is a stationary point of the function \( f \).

Lemma 4: For \( \alpha \in [0,1] \) and all \( k \geq m \), we have
\[
 g_k^T s_k(\alpha) \leq -(1-\rho) \|s_k\|^2.
\]
Proof is analogous to the proof of Lemma 2.1 in [4].

Convergence theorem. Suppose that \( f \in LC^1 \) and that the assumptions A1 and A2 hold. Then for any initial point \( x_0 \in D \), \( x_k \rightarrow \bar{x} \), as \( k \rightarrow \infty \), where \( \bar{x} \) is a unique minimal point.

Proof: If \( d_k^* \neq 0 \) is a solution of (3), it follows that \( \Phi_k(d_k^*) \leq 0 = \Phi_k(0) \). Consequently, we have by (5) that
\[ g(x_k)^T d_k \leq -\frac{1}{2} f'_D(x_k; d_k) \leq -\frac{1}{2} c_1 \|d_k\| < 0, \quad \text{i.e.} \quad (6) \]

\( d_k \) is a descent direction at \( x_k \). From (3), (5) and Lemma 4 it follows that
\[
f(x_k) - f(x_{k+1}) \geq \sigma \left[ -q^{(i(k))} g_{s_k}^T s_k + \frac{1}{2} q^{(i(k))} f'_D(x_k; d_k) \right] \geq q^{(i(k))} (1 - \rho) \|g_{s_k}\| + \frac{\sigma}{2} q^{(i(k))} c_1, \quad (7) \]

Hence \( \{f(x_k)\} \) is a decreasing sequence and consequently \( \{x_k\} \subset L(x_0) \). Since \( L(x_0) \) is by Proposition a compact convex set, it follows that the sequence \( \{x_k\} \) is bounded. Therefore there exist accumulation points of \( \{x_k\} \). Since the gradient \( g \) is by assumption continuous, then, if \( g(x_k) \to 0 \) as \( k \to \infty \), it follows that every accumulation point \( \bar{x} \) of the sequence \( \{x_k\} \) satisfies \( g(\bar{x}) = 0 \) . Since \( f \) is by the Proposition strictly convex, it follows that there exists a unique point \( \bar{x} \in L(x_0) \) such that \( g(\bar{x}) = 0 \). Hence, \( \{x_k\} \) has a unique limit point \( \bar{x} \) and it is a global minimizer.

Therefore we have to prove that \( g(x_k) \to 0, k \to \infty \). There are two cases to consider.

a) The set of indices \( \{i(k)\} \) for \( k \in K_i \) , is uniformly bounded above by a number \( I \), i.e.
\( i(k) \leq I < \infty \) for \( k \in K_i \). Consequently, from (3) and (7) it follows that
\[
f(x_k) - f(x_{k+1}) \geq \sigma \left[ -q^{(i(k))} g_{s_k}^T s_k + \frac{1}{2} q^{(i(k))} f'_D(x_k; d_k) \right] \geq q^{(i(k))} (1 - \rho) \|g_{s_k}\| + \frac{\sigma}{2} q^{(i(k))} c_1, \quad (8)
\]

(since \( g(x_k)^T s_k \leq 0 \) and \( f'_D(x_k; d_k) > 0 \) \( \geq q^{(i(k))} (1 - \rho) \|g_{s_k}\| + \frac{\sigma}{2} q^{(i(k))} c_1 \). Since \( \{f(x_k)\} \) is bounded below (on the compact set \( L(x_0) \) ) and monotone (by (7)), it follows that \( f(x_{k+1}) - f(x_k) \to 0 \) as \( k \to \infty, k \in K_i \); hence from (8) it follows that \( \|g(x_k)\| \to 0 \) and \( f'_D(x_k; d_k) \to 0, k \to \infty, k \in K_i \).

b) There is a subset \( K_2 \subset K_i \) such that \( \lim_{k \to \infty} i(k) = \infty \).
This part of proof is analogous to the proof in [1].

In order to have a finite value \( i(k) \), it is sufficient that \( s_k \) and \( d_k \) have descent properties, i.e.
\[
g(x_k)^T s_k < 0 \quad \text{and} \quad g(x_k)^T d_k < 0
\]
whenever \( g(x_k) \neq 0 \). The first relation follows from Lemma 4 and the second follows from (6). At a saddle point the relation (3) becomes
\[ f(x_k) - f(x_{k+1}) \geq \sigma \left[ \frac{1}{2} q^{d(k)} f_D'(x_k, d_k) \right] \tag{9} \]

In that case by Lemma 3 \( d_k \neq 0 \) and hence, by (5), \( f'(x_k, d_k) > 0 \); so (9) clearly can be satisfied.

**Convergence rate theorem:** Under the assumptions of the previous theorem we have that the following estimate holds for the sequence \( \{x_i\} \) generated by the algorithm.

\[ f(x_n) - f(\bar{x}) \leq \mu_0 \left[ 1 + \frac{\mu_0}{\eta} \sum_{i=0}^{n-1} \frac{f(x_k) - f(x_{k+1})}{\|\nabla f(x_k)\|} \right]^{-1}, \]

\( n = 1, 2, \ldots \) where \( \mu_0 = f(x_0) - f(\bar{x}) \), and \( \text{diam} L(x_0) = \eta < \infty \) since by Proposition it follows that \( L(x_0) \) is bounded.

**Proof:** The proof directly follows from the Theorem 9.2, page 167 in [2], since the assumptions of that theorem are fulfilled.

4. **CONCLUSION**

The algorithm presented in this paper is based on the algorithms from [1] and [4]. The convergence is proved under mild conditions. This method uses previous multi-step iterative information and curve search rule to generate a new iterative point at each iteration. Relating to the algorithms in [1] and [4], in [4] it is supposed that the function \( f \) has a lower bound on the level set \( L(x_0) \) and that the gradient \( g(x) \) of \( f(x) \) is uniformly continuous on an open convex set \( B \) that contains \( L(x_0) \), while in this paper and in the previous paper [1] we supposed that \( f : D \subset \mathbb{R}^n \rightarrow \mathbb{R} \) is a \( LC^1 \) function on the open convex set \( D \), and that the second order Dini upper directional derivative satisfies the condition (5).

**Acknowledgment:** This research was supported by Science Fund of Serbia, grant number 144015G, through institute of Mathematics, SANU.

**REFERENCES**