AN ALGORITHM FOR LC^1 OPTIMIZATION*

Nada I. ĐURANOVIĆ-MILIČIĆ
Department of Mathematics, Faculty of Technology and Metallurgy
University of Belgrade, Belgrade, Yugoslavia
nmilicic@elab.tmf.bg.ac.yu

Abstract. In this paper an algorithm for LC^1 unconstrained optimization problems, which uses the second order Dini upper directional derivative is considered. The purpose of the paper is to establish general algorithm hypotheses under which convergence occurs to optimal points. A convergence proof is given, as well as an estimate of the rate of convergence.

Keywords. Directional derivative, second order Dini upper directional derivative, uniformly convex functions.

1. INTRODUCTION

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

$$\min\left\{ f(x) \mid x \in D \subset R^n \right\}, \tag{1}$$

where $f : D \subset R^n \to R$ 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 ∇f is locally Lipschitzian, i.e.

$$\left\| \nabla f(x) - \nabla f(y) \right\| \leq L \left\| x - y \right\| \quad \text{for} \quad x, y \in D$$

for some $L > 0$.

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

$$x_{k+1} = x_k + \alpha_k s_k + \alpha_k^2 d_k, \quad k = 0,1,..., \quad s_k \neq 0, d_k \neq 0 \qquad (2)$$

where the step-size α_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 [4]). *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*

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

If ∇f is directionally differentiable at x_k, we have

$$f_D''(x_k;d) = f''(x_k;d) = \lim_{\lambda \downarrow 0} \frac{[\nabla f(x_k + \lambda d) - \nabla f(x_k)]^T d}{\lambda} \quad \text{for all } d \in R^n.$$

Lemma 1 (See [4]). *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 (1), then:*

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

and $f_D''(x;d) \geq 0, \forall d \in R^n$.

Lemma 2 (See [4]): *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*

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

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

At the k-th iteration the direction vector $s_k \neq 0$ in (2) is any vector satisfying the descent property, i.e. $\nabla f(x_k)^T s_k \leq 0$ holds, and the direction vector d_k presents a solution of the problem

$$\min\{\Phi_k(d) \mid d \in R^n\} \qquad (3)$$

where $\Phi_k(d) = \nabla f(x_k)^T d + \frac{1}{2} f_D''(x_k,d)$. For given q, where $0 < q < 1$, the step-size $\alpha_k > 0$ is a number satisfying

$$\alpha_k = q^{i(k)},$$

where $i(k)$ is the smallest integer from $i = 0, 1, ...$ such that

$$x_{k+1} = x_k + q^{i(k)} s_k + q^{2i(k)} d_k \in D$$

and

$$f(x_{k+1}) - f(x_k) \le \sigma \left[q^{i(k)} \nabla f(x_k)^T s_k - \frac{1}{2} q^{4i(k)} f_D''(x_k; d_k) \right] \tag{4}$$

where $0 < \sigma < 1$ is a pre assigned constant, and $x_0 \in D$ is a given point.

We make the following assumptions.

A1. We suppose that there exist constants $c_2 \ge c_1 > 0$ such that

$$c_1 \|d\|^2 \le f_D''(x; d) \le c_2 \|d\|^2 \tag{5}$$

for every $d \in R^n$.

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

A3. There exists a value $\beta > 0$ such that

$$\nabla f(x_k)^T s_k \le -\beta \|\nabla f(x_k)\| \|s_k\|, \quad k = 0, 1, 2, ... \tag{6}$$

It follows from Lemma 3.1 in [4] that under the assumption A1 the optimal solution of the problem (3) exists.

Proposition 1. 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 $L(x_0) = \{x \in D : f(x) \le f(x_0)\}$ is a compact convex set; 3) there exists a unique point x^* such that $f(x^*) = \min\limits_{x \in L(x_0)} f(x)$.

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

$$f(x) - f(x_0) = \nabla f(x_0)^T (x - x_0) + \frac{1}{2} f_D''[x_0 + \theta(x - x_0); x - x_0]$$

$$\ge \nabla f(x_0)^T (x - x_0) + \frac{1}{2} c_1 \|x - x_0\|^2 > \nabla f(x_0)^T (x - x_0),$$

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 [5]) 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

$$f(x^*) = \min\limits_{x \in L(x_0)} f(x) = \min\limits_{x \in D} f(x).$$

Since f is strictly convex it follows from [5] that x^* is the unique minimizer.

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

Theorem 1 (Convergence theorem). *Suppose that $f \in LC^1$ and that the assumptions A1, A2 and A3 hold. Then for any initial point $x_0 \in D$, $x_k \to \overline{x}$, as $k \to \infty$, where $\overline{x}$ is the 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

$$\nabla f(x_k)^T d_k \leq -\frac{1}{2} f_D''(x_k; d_k) \leq -\frac{1}{2} c_1 \|d_k\|^2 < 0 \tag{7}$$

d_k is a descent direction at x_k. From (4), (5) and (6) it follows that

$$f(x_{k+1}) - f(x_k) \leq \sigma \left[q^{i(k)} \nabla f(x_k)^T s_k - \frac{1}{2} q^{4i(k)} f_D''(x_k; d_k) \right]$$

$$\leq \sigma \left[-\beta \|\nabla f(x_k)\| \|s_k\| - c_1 \|d_k\|^2 \right] < 0 \tag{8}$$

Hence $\{f(x_k)\}$ is a decreasing sequence and consequently $\{x_k\} \subset L(x_0)$. Since $L(x_0)$ is by Proposition 1 a compact convex set, it follows that the sequence $\{x_k\}$ is bounded. Therefore there exist accumulation points of $\{x_k\}$. Since ∇f is by assumption continuous, then, if $\nabla f(x_k) \to 0$ as $k \to \infty$, it follows that every accumulation point $\overline{x}$ of the sequence $\{x_k\}$ satisfies $\nabla f(\overline{x}) = 0$. Since f is by the Proposition 1 strictly convex, it follows that there exists a unique point $\overline{x} \in L(x_0)$ such that $\nabla f(\overline{x}) = 0$. Hence, $\{x_k\}$ has a unique limit point $\overline{x}$ – and it is a global minimizer. Therefore we have to prove that $\nabla f(x_k) \to 0, k \to \infty$. There are two cases to consider.

a) There exists an infinite set K_1 such that the set of indices $\{i(k)\}$ for $k \in K_1$, is uniformly bounded above by a number I, i.e. $i(k) \leq I < \infty$ for $k \in K_1$. Consequently, from (4) and (6) it follows that

$$f(x_{k+1}) - f(x_k) \leq \sigma \left[q^{i(k)} \nabla f(x_k)^T s_k - \frac{1}{2} q^{4i(k)} f_D''(x_k; d_k) \right]$$

$$\leq \sigma \left[q^I \nabla f(x_k)^T s_k - \frac{1}{2} q^{4I} f_D''(x_k; d_k) \right]$$

$$\leq -\sigma q^I \beta \|\nabla f(x_k)\| \|s_k\| - \frac{\sigma}{2} q^{4I} f_D''(x_k; d_k)$$

$$= -\sigma q^I \beta \|\nabla f(x_k)\| - \frac{\sigma}{2} q^{4I} f_D''(x_k; d_k)$$

(since by A2 $\|s_k\| = 1, k = 0, 1, \dots$).

Multiplying this inequality by (–1) we get

$$f(x_k) - f(x_{k+1}) \geq \sigma q^I \beta \|\nabla f(x_k)\| + \frac{\sigma}{2} q^{4I} f''_D(x_k; d_k) \tag{9}$$

Since $\{f(x_k)\}$ is bounded below and $f(x_{k+1}) - f(x_k) \to 0$ as $k \to \infty, k \in K_1$ from (9) it follows that $\|\nabla f(x_k)\| \to 0$ and $f''_D(x_k, d_k) \to 0, k \to \infty, k \in K_1$.

b) There exists an infinite set K_2 such that $\lim_{k \to \infty} i(k) = \infty, k \in K_2$.

By the definition of $i(k)$ it follows that

$$f(x_{k+1}) - f(x_k) > \sigma \left[q^{i(k)-1} \nabla f(x_k)^T s_k - \frac{1}{2} q^{4i(k)-4} f''_D(x_k; d_k) \right] \tag{10}$$

By the definition of the Dini derivative and by (5) we have

$$f(x_{k+1}) - f(x_k) = q^{i(k)-1} \nabla f(x_k)^T s_k + q^{2i(k)-2} \nabla f(x_k)^T d_k$$
$$+ \frac{1}{2} f''_D \left[x_k; q^{i(k)-1} s_k + q^{2i(k)-2} d_k \right] + o(q^{2i(k)-2}) \geq q^{i(k)-1} \nabla f(x_k)^T s_k$$
$$+ q^{2i(k)-2} \nabla f(x_k)^T d_k + \frac{1}{2} c_1 \| q^{i(k)-1} s_k + q^{2i(k)-2} d_k \|^2 + o(q^{2i(k)-2})$$
$$= q^{i(k)-1} \nabla f(x_k)^T s_k + q^{2i(k)-2} \nabla f(x_k)^T d_k + \frac{1}{2} c_1 q^{2i(k)-2} \| s_k + q^{i(k)-1} d_k \|^2$$
$$+ o(q^{2i(k)-2}) > \sigma \left[q^{i(k)-1} \nabla f(x_k)^T s_k - \frac{1}{2} q^{4i(k)-4} f''_D(x_k; d_k) \right] \quad \text{(according to (10))}.$$

Accumulating all terms of order higher than $o(q^{2i(k)-2})$ into the $o(q^{2i(k)-2})$ term (because $\|s_k\| = \|d_k\| = 1$) and using the fact that $\nabla f(x_k)^T d_k \leq 0$ yields

$$\frac{1}{2} c_1 q^{2i(k)-2} \|s_k\|^2 + o(q^{2i(k)-2}) > (\sigma - 1) q^{i(k)-1} \nabla f(x_k)^T s_k \geq 0$$

since $0 < \sigma < 1$ and $\nabla f(x_k)^T s_k \leq 0$. Dividing by $q^{i(k)-1}$ yields

$$\frac{1}{2} c_1 q^{i(k)-1} \|s_k\|^2 + o(q^{i(k)-1}) > (\sigma - 1) \nabla f(x_k)^T s_k.$$

Dividing by $\frac{1}{2} c_1 \|s_k\|^2 = \frac{1}{2} c_1$ yields

$$q^{i(k)-1} > \frac{2(\sigma - 1)}{c_1} \nabla f(x_k)^T s_k + \frac{o(q^{i(k)-1})}{\frac{1}{2} c_1}.$$

Taking the limit as $k \to \infty, k \in K_2$ and having in view (6), we get

$$q^{i(k)-1} > \frac{2(1-\sigma)\beta}{c_1} \|\nabla f(x_k)\| + o(q^{i(k)-1}).$$

Since $q^{i(k)-1} \to 0$ as $k \to \infty$, $k \in K_2$, it follows that $\|\nabla f(x_k)\| \to 0$ as $k \to \infty$, $k \in K_2$.

In order to have a finite value $i(k)$, it is sufficient that s_k and d_k have descent properties, i.e.

$$\nabla f(x_k)^T s_k < 0 \quad \text{and} \quad \nabla f(x_k)^T d_k < 0 \quad \text{whenever} \quad \nabla f(x_k) \neq 0 .$$

The first relation follows from (6) and the second follows from (7). At a saddle point the relation (4) becomes

$$f(x_{k+1}) - f(x_k) \leq -\frac{\sigma}{2} q^{4i(k)} f_D''(x_k; d_k) \tag{11}$$

In that case $d_k \neq 0$ by Lemma 3 and hence, by (5), $f''(x_k; d_k) > 0$; so (11) clearly can be satisfied.

Theorem 2 (Convergence rate theorem). *Under the assumptions of the previous theorem we have that the following estimate holds for the sequence $\{x_k\}$ generated by the algorithm.*

$$f(x_n) - f(\overline{x}) \leq \mu_0 \left[1 + \mu_0 \frac{1}{\eta^2} \sum_{k=0}^{n-1} \frac{f(x_k) - f(x_{k+1})}{\|\nabla f(x_k)\|^2} \right]^{-1} , \quad n = 1, 2, \ldots,$$

where $\mu_0 = f(x_0) - f(\overline{x})$, and $\operatorname{diam} L(x_0) = \eta < \infty$ since by Proposition 1 it follows that $L(x_0)$ is bounded.

Proof: The proof directly follows from the Theorem 9.2, page 167 in [2].

4. CONCLUSION

As it has already been pointed out, the algorithm presented in this paper is based on the algorithms from [1] and [4]. Note that such an algorithm generates at every iteration a point closer to an optimal point than the algorithms given in [4]. It happens because in [4] minimization is applied along one direction, while here we have minimization along a plane defined by the vectors s_k and d_k. Relating to the algorithm in [1], the presented algorithm is defined and converges under weaker assumptions than the algorithm given in [1].

REFERENCES

[1] Goldfarb, D., "Curvilinear path steplength algorithms for minimization whith directions of negative curvature", *Mathematical Programming,* 18 (1980) 31-40.
[2] Karmanov, V.G., *Matematiceskoe Programirovanie*, Nauka, Moskva, 1975.
[3] Polak, E., *Cislennie metodi optimizacii: edinij podhod* (translation from English), Mir, Moskva, 1974.
[4] Sun, V.J., Sampaio, R.J.B., and Yuan, Y.J., "Two algorithms for LC^1 unconstrained optimization", *Journal of Computational Mathematics,* 6 (2000) 621-632.
[5] Vujčić, V., Ašić, M., and Miličić, N., *Mathematical Programming* , Institute of Mathematics, Belgrade, 1980. (In Serbian)