Filomat 27:1 (2013), 15–22
DOI 10.2298/FIL1301015G

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

An Algorithm Using Moreau-Yosida Regularization for Minimization of a Nondifferentiable Convex Function

Milanka Gardašević-Filipović[a], Nada Ðuranović-Miličić[b]

[a] *Vocational College of Technology, Arandjelovac, Serbia*
[b] *Faculty of Technology and Metallurgy,University of Belgrade, Serbia*

Abstract. In this paper we present an algorithm for minimization of a nondifferentiable proper closed convex function. Using the second order Dini upper directional derivative of the Moreau-Yosida regularization of the objective function we make a quadratic approximation. It is proved that the sequence of points generated by the algorithm has an accumulation point which satisfies the first order necessary and sufficient conditions.

1. Introduction

The following minimization problem is considered:

$$\min_{x\in R^n} f(x) \tag{1}$$

where $f : R^n \to R\bigcup\{+\infty\}$ is a convex and not necessary differentiable function with a nonempty set X^* of minima.

Non-smooth optimization problems, in general, are difficult to solve, even when they are unconstrained. For nonsmooth programs, many approaches have been presented so far and they are often restricted to the convex unconstrained case. In general, the various approaches are based on combinations of the following three methods: (i) subgradient methods (see [4], [11]); (ii) bundle techniques (see [12], [14], [15], [22]), (iii) Moreau-Yosida regularization (see [13], [20], [17]).

For a function f it is very important that its Moreau-Yosida regularization is a new function which has the same set of minima and is differentiable with Lipchitz continuous gradient, even when f is not differentiable.

In [23] the optimality conditions and an algorithm for minimizing an LC^1 function are given. Having in mind that the Moreau-Yosida regularization of a proper closed convex function is an LC^1 function, we present an optimization algorithm based on the results from [23] and [6] (using the second order Dini upper directional derivative (described in [2] and [3])). We shall present an iterative algorithm for finding an optimal solution of the problem (1) by generating the sequence of points $\{x_k\}$ of the following form:

$$x_{k+1} = x_k + d_k \quad k = 0, 1, \ldots, d_k \neq 0 \tag{2}$$

2010 *Mathematics Subject Classification*. Primary 65K05; Secondary 90C30, 90C25.
Keywords. trust region method, non-smooth convex optimization, Moreau-Yosida regularization, Dini directional derivative, semismooth function
Received: 28 August 2010; Accepted: 26 August 2011
Communicated by Dragan S. Djordjevic
Email addresses: milankafilipovic@yahoo.com (Milanka Gardašević-Filipović), nmilicic@tmf.bg.ac.rs (Nada Ðuranović-Miličić)

where the directional vector d_k is defined by the particular algorithm. That is the main idea of this paper.

Paper is organized as follows: in the second section some basic theoretical preliminaries are given; in the third section the Moreau-Yosida regularization and its properties are described; in the fourth section the definition of the second upper Dini directional derivative and the basic properties are given; in the fifth section the semismooth functions and conditions for their minimization are described. Finally, in the sixth section a model algorithm is suggested and the convergence of the algorithm is proved.

2. Theoretical preliminaries

Throughout the paper we will use the following notation. A vector s refers to a column vector, and ∇ denotes the gradient operator $\left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \ldots, \frac{\partial}{\partial x_n}\right)^T$. The Euclidean product is denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ is the associated norm. For a given symmetric positive definite linear operator M we set $\langle \cdot, \cdot \rangle_M := \langle M \cdot, \cdot \rangle$; hence it is shortly denoted by $\|x\|_M^2 := \langle x, x \rangle_M$. The smallest and the largest eigenvalue of M we denote by λ and Λ respectively.

The *domain* of a given function $f : R^n \to R \bigcup \{+\infty\}$ is the set $\mathrm{dom}(f) = \{x \in R^n | f(x) < +\infty\}$. We say f is proper if its domain is nonempty. The point $x^* = \arg\min_{x \in R^n} f(x)$ refers to the minimum point of a given function $f : R^n \to R \bigcup \{+\infty\}$.

A vector $g \in R^n$ is said to be a *subgradient* of a given proper convex function $f : R^n \to R \bigcup \{+\infty\}$ at a point $x \in R^n$ if the next inequality

$$f(z) \geq f(x) + g^T \cdot (z - x) \tag{3}$$

holds for all $z \in R^n$. The set of all subgradients of $f(x)$ at the point x, called the *subdifferential* at the point x, is denoted by $\partial f(x)$. The subdifferential $\partial f(x)$ is a nonempty set if and only if $x \in \mathrm{dom}(f)$. The condition $0 \in \partial f(x)$ is a first order necessary and sufficient condition for a global minimizer for the convex function f at the point $x \in R^n$ (see in [1] or [18]).

For convex function f it follows that $f(x) = \max_{z \in R^n} \{f(z) + g^T(x - z)\}$ holds, where $g \in \partial f(z)$ ([5]).

The concept of the subgradient is a simple generalization of the gradient for nondifferentiable convex functions.

The *directional derivative* of a real function f defined on R^n at the point $x' \in R^n$ in the direction $s \in R^n$, denoted by $f'(x', s)$, is

$$f'(x', s) = \lim_{t \downarrow 0} \frac{f(x' + t \cdot s) - f(x')}{t} \tag{4}$$

when this limit exists. For a real convex function a directional derivative at the point $x' \in R^n$ in the direction s exists in any direction $s \in R^n$ (see Theorem 2.1.3, page 10 in [16]).

At the end of this section we recall the definition of LC^1 function.

Definition 1. *A real function f defined on R^n is an LC^1 function on the open set $D \subseteq R^n$ if it is continuously differentiable and its gradient ∇f is locally Lipschitzian, i.e., $\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$ for $x, y \in D$ holds for some $L > 0$.*

3. The Moreau-Yosida regularization

Definition 2. *Let $f : R^n \to R \bigcup \{+\infty\}$ be a proper closed convex function. The Moreau-Yosida regularization of a given function f, associated to the metric defined by M, denoted by F, is defined as follows:*

$$F(x) := \min_{y \in R^n} \left\{ f(y) + \frac{1}{2}\|y - x\|_M^2 \right\}. \tag{5}$$

The above function is an *infimal convolution*. In [21] (Theorem 5.4, page 50) it is proved that infimal convolution of a convex function is also a convex function. Hence the function defined by (5) is a convex function and has the same set of minima as the function f (see in [7]), so the motivation of the study of Moreau-Yosida regularization is due to the fact that $\min_{x\in R^n} f(x)$ is equivalent to $\min_{x\in R^n} F(x)$.

The minimum point $p(x)$ of the function (5), i.e.: $p(x) := \arg\min_{y\in R^n} \left\{ f(y) + \frac{1}{2}\|y-x\|_M^2 \right\}$ is called the *proximal point* of x. In [7] it is proved that the function F defined by (5) is always differentiable.

The first order regularity of F is well known (see in [7] and [13]): without any further assumptions, F has a Lipschitzian gradient on the whole space R^n. More precisely,

$$\|\nabla F(x_1) - \nabla F(x_2)\|^2 \leq \Lambda \langle \nabla F(x_1) - \nabla F(x_2), x_1 - x_2 \rangle$$

holds for all $x_1, x_2 \in R^n$ (see in [13]), where $\nabla F(x)$ has the following form:

$$G := \nabla F(x) = M(x - p(x)) \in \partial f(p(x)) \tag{6}$$

and $p(x)$ is the unique minimizer in (5). So, according to the above consideration and Definition 1, we conclude that F is an LC^1 function (if Λ is Lipschitzian constant for F then it is also Lipschitzian constant for ∇F).

Note in particular that the function f has nonempty subdifferential at any point p of the form $p(x)$. Since $p(x)$ is the minimum point of the function (5) then it follows (see in [7] and [13]) that $p(x) = x - M^{-1}g$ where $g \in \partial f(p(x))$.

Lemma 1. *The following statements are equivalent:*
(i) x minimizes f;
(ii) $p(x) = x$;
(iii) $\nabla F(x) = 0$;
(iv) x minimizes F;
(v) $f(p(x)) = f(x)$;
(vi) $F(x) = f(x)$.

Proof. See in [7] or [20].

4. Dini second upper directional derivative

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

Definition 3. *The second order Dini upper directional derivative of the function $f \in LC^1$ at $x \in R^n$ in the direction $d \in R^n$ is defined to be*

$$f_D''(x,d) = \limsup_{\alpha\downarrow 0} \frac{\left[\nabla f(x+\alpha d) - \nabla f(x)\right]^T \cdot d}{\alpha}.$$

If ∇f is directionally differentiable at x, we have

$$f_D''(x,d) = f''(x_k,d) = \lim_{\alpha\downarrow 0} \frac{\left[\nabla f(x+\alpha d) - \nabla f(x)\right]^T \cdot d}{\alpha} \quad \textit{for all} \quad d \in R^n.$$

Remark 1. *For a locally Lipschitzian function the directional derivative may not exist but the Dini directional derivative always there exists. (see [24], page 598).*

According to the Rademacher's theorem, which state that any Lipschitzian continuous function from R^n to R is differentiable almost everywhere in R^n, it follows that ∇F is differentiable almost everywhere on R^n. Let $D_{\nabla F}$ be the *set of points where ∇F is differentiable. The generalized Hessian in the sense of Clarke is defined to be $\partial^2 F(x_k) = co\{\lim_{x_i \to x_k} \nabla^2 F(x_i) | \, x_i \in D_{\nabla F}\}$, where co stands for the convex hull of all $n \times n$ matrices obtained as a limit of sequence of Hessian matrices $\nabla^2 F(x_i)$. So, $\partial^2 F(x)$ is nonempty convex compact subset of $R^{n\times n}$. According to Caratheodory theorem (see [9], page 195 or [25],page 155) it follows that if $V \in \partial^2 F(x)$ then there exist $V_j \in \{\lim_{x_i \to x} \nabla^2 F(x_i) | \, x_i \in D_{\nabla F}\}$ and $\lambda_j \in [0,1]$, $r \leq n^2 + 1$ such that $\sum_{j=1}^r \lambda_j = 1$ and $V = \sum_{j=1}^r \lambda_j V_j$. In other words, $V \in \partial^2 F(x)$ is obtained as a convex combination of limit of semipositive definite matrices.*

Lemma 2. *Let $f : R^n \to R$ be a closed convex proper function and F be its Moreau – Yosida regularization for $M = I$. Then $V \in \partial^2 F(x)$ is a positive semidefinite matrix.*

Proof. Let V and d be elements from the sets $\partial^2 F(x)$ and R^n respectively. Then by Charatheodory theorem there exist $V_j \in \{\lim_{x_i \to x} \nabla^2 F(x_i) | \, x_i \in D_{\nabla F}\}$ and $\lambda_j \in [0,1]$, $r \leq n^2 + 1$ such that $\sum_{j=1}^r \lambda_j = 1$ and $V = \sum_{j=1}^r \lambda_j V_j$. So, we have that

$$\begin{aligned} d^T V d = d^T \left(\sum_{j=1}^r \lambda_j V_j \right) d &= d^T \left(\sum_{j=1}^r \lambda_j \lim_{x_i \to x_j} \nabla^2 F(x_j) \right) d, x_i \in D_{\nabla F} \\ &= \sum_{j=1}^r \lambda_j \lim_{x_i \to x_j} (d^T \nabla^2 F(x_j) d), x_i \in D_{\nabla F} \\ &\geq \sum_{j=1}^r \lambda_j \cdot 0 = 0 \end{aligned}$$

where the last inequality holds because $\nabla^2 F(x_i)$ is positive semidefinite at every $x_i \in D_{\nabla F}$ as the Hessian of the convex function F.

Lemma 3. *Let $f : R^n \to R$ be a closed convex proper function and F be its Moreau –Yosida regularization for $M = I$. Then the next statements are valid.*

(i) $F''_D(x,d)$ is upper semicontinous with respect to (x,d) i.e. if $x_i \to x$ and $d_i \to d$, then

$$\limsup_{i\to\infty} F''_D(x_i, d_i) \leq F''_D(x,d)$$

(ii) $F''_D(x,d) = \max\left\{ d^T V d | V \in \partial^2 F(x) \right\}$

(iii) $|F''_D(x_k,d)| \leq K \cdot \|d\|^2$, where K is some constant.

(iv) $F''_D(x_k, kd) = k^2 F''_D(x_k, d)$

Proof. See in [3].

5. Semi-smooth functions and optimality conditions

Definition 4. *A function $\nabla F : R^n \to R^n$ is said to be semi-smooth at the point $x \in R^n$ if ∇F is locally Lipschitzian at the point $x \in R^n$ and the limit $\lim_{\substack{h\to d \\ \lambda \downarrow 0}} \{Vh\}$, $V \in \partial^2 F(x + \lambda h)$ there exists for any $d \in R^n$.*

Note that for a closed convex proper function, the gradient of its Moreau-Yosida regularization is a semi-smooth function (see in [26]).

Lemma 4. *[23]: If $\nabla F : R^n \to R^n$ is semi-smooth at the point $x \in R^n$ then ∇F is directionally differentiable at $x \in R^n$ 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 $h^T V h - F''(x,h) = o(\|h\|^2)$.*

Lemma 5. *Let $f : R^n \to R$ be a proper closed convex function and F be its Moreau-Yosida regularization for M=I. Then if $x \in R^n$ is a solution of the problem (1), then $F'(x,d) = 0$ and $F''_D(x,d) \geq 0$ for all $d \in R^n$.*

Proof. From the definition of the directional derivative and by Lemma 1 we have that $F'(x,d) = \nabla F(x)^T d = 0$. Since $x \in R^n$ is a solution of the problem (1) then according to Lemma 1, Theorem 23.1 in [21] and the fact that the next inequalities $F'(x+td,d) \geq \frac{1}{t}(F(x+td) - F(x)) \geq 0$ hold, we have that

$$F''_D(x,d) = \lim_{t\downarrow 0} \frac{F'(x+td,d) - F'(x,d)}{t} \geq 0.$$

Lemma 6. *Let $f : R^n \to R$ be a proper closed convex function and F be its Moreau-Yosida regularization for $M = I$. Let x be a point from R^n such that $F'(x,d) = 0$ and $F''_D(x,d) > 0$ hold for all $d \in R^n$. Then $x \in R^n$ is a strict local minimizer of the problem (1).*

Proof. Suppose that $x \in R^n$ is not a strict minimizer of the function f. According to Lemma 1 then $x \in R^n$ is neither strict minimizer of the function F, nor a proximal point of the function F. Then there exists a sequence $\{x_k\}, x_k \to x$ such that $F(x_k) \leq F(x)$ holds for every k. If we define the sequence $\{x_k\}, x_k \to x$ by $x_k = x + t_k d$, where $t_k = \frac{\|x_k - x\|}{\|d\|}$, then by Lemma 3 it follows that $F(x_k) - F(x) - t_k \nabla F(x)^T d = \frac{1}{2} t_k^2 F''_D(x,d) + o(\|d\|^2)$ holds. Since $\nabla F(x) = 0$ it follows that $\frac{1}{2} t_k^2 F'''_D(x,d) \leq 0$, which contradicts the assumption.

6. A model algorithm

In this section an algorithm for solving the problem (1) is introduced. We suppose that at each $x \in R^n$ it is possible to compute $f(x)$ and $\nabla F(x)$, and $F''_D(x,d)$ for a given $d \in R^n$, where F is Moreau-Yosida regularization of the function f for $M = I$.

At the k-th iteration we consider the following problem

$$\min_{d \in R^n} \tilde{\Phi}_k(d), \tilde{\Phi}_k(d) = F(x_k) + \nabla F(x_k)^T d + \frac{1}{2} F''_D(x_k, d) \tag{7}$$

where $F''_D(x_k,d)$ stands for the second order Dini upper directional derivative at x_k in the direction d. Note that if Λ is Lipschitzian constant for F it is also Lipschitzian constant for ∇F. Since F is an LC^1 function, it follows that $F''_D(x_k,d)$ always there exists for all $d \in R^n$. $\tilde{\Phi}_k(d)$ is called an *iteration function*. It is easy to see that $\tilde{\Phi}_k(0) = F(x_k)$ and $\tilde{\Phi}_k(d)$ is Lipschitzian on R^n (with the same Lipshitzian constant as a function F).

Lemma 7. *The objective function of the problem (7) is a convex function.*

Proof. Let d_1 and d_2 be vectors from R^n and $\hat{V}_k \in \partial^2 F(x_k)$. Let λ be a scalar such that $0 < \lambda < 1$. Then

$$\begin{aligned} \tilde{\Phi}_k(\lambda d_1 + (1-\lambda)d_2) &= F(x_k) + \nabla F(x_k)^T(\lambda d_1 + (1-\lambda)d_2) + \frac{1}{2} F''_D(x_k, (\lambda d_1 + (1-\lambda)d_2)) \\ &= \lambda(F(x_k) + \nabla F(x_k)^T d_1) + (1-\lambda)(F(x_k) + \nabla F(x_k)^T d_2) + \\ &+ \frac{1}{2} \max_{\hat{V}_k \in \partial^2 F(x_k)} (\lambda d_1 + (1-\lambda)d_2)^T \hat{V}_k (\lambda d_1 + (1-\lambda)d_2) \end{aligned} \tag{8}$$

holds. Since $\hat{V}_k \in \partial^2 F(x_k)$ is positive semidefinite as stated in Lemma 2 then it follows:

$$(d_1 - d_2)^T \hat{V}_k (d_1 - d_2) \geq 0 \Rightarrow d_1^T \hat{V}_k d_2 + d_2^T \hat{V}_k d_1 \leq d_1^T \hat{V}_k d_1 + d_2^T \hat{V}_k d_2.$$

Hence

$$\begin{aligned} (\lambda d_1 + (1-\lambda)d_2)^T \hat{V}_k (\lambda d_1 + (1-\lambda)d_2) &= \lambda^2 d_1^T \hat{V}_k d_1 + (1-\lambda)^2 d_2^T \hat{V}_k d_2 + \\ &+ \lambda(1-\lambda) d_2^T \hat{V}_k d_1 + \lambda(1-\lambda) d_1^T \hat{V}_k d_2 \\ &\leq \lambda^2 d_1^T \hat{V}_k d_1 + (1-\lambda)^2 d_2^T \hat{V}_k d_2 + \\ &+ \lambda(1-\lambda)(d_1^T \hat{V}_k d_1 + d_2^T \hat{V}_k d_2) \\ &= \lambda d_1^T \hat{V}_k d_1 + (1-\lambda) d_2^T \hat{V}_k d_2. \end{aligned} \tag{9}$$

So, from (8) and (9) then it follows:

$$\begin{aligned}\tilde{\Phi}_k(\lambda d_1 + (1-\lambda)d_2) &\leq \lambda(F(x_k) + \nabla F(x_k)^T d_1) + (1-\lambda)(F(x_k) + \nabla F(x_k)^T d_2) + \\ &+ \frac{1}{2} \max_{\hat{V}_k \in \partial^2 F(x_k)} (\lambda d_1 + (1-\lambda)d_2)^T \hat{V}_k (\lambda d_1 + (1-\lambda)d_2) \\ &\leq \lambda(F(x_k) + \nabla F(x_k)^T d_1) + (1-\lambda)(F(x_k) + \nabla F(x_k)^T d_2) + \\ &+ \frac{1}{2} \max_{\hat{V}_k \in \partial^2 F(x_k)} (\lambda d_1^T \hat{V}_k d_1 + (1-\lambda) d_2^T \hat{V}_k d_2) \\ &\leq \lambda(F(x_k) + \nabla F(x_k)^T d_1) + (1-\lambda)(F(x_k) + \nabla F(x_k)^T d_2) + \\ &+ \lambda \frac{1}{2} \max_{\hat{V}_k \in \partial^2 F(x_k)} d_1^T \hat{V}_k d_1 + (1-\lambda) \frac{1}{2} \max_{\hat{V}_k \in \partial^2 F(x_k)} d_2^T \hat{V}_k d_2 \\ &= \lambda \tilde{\Phi}_k(d_1) + (1-\lambda)\tilde{\Phi}_k(d_2)\end{aligned}$$

(where the last inequality holds since the maximum over the sum of two nonnegative functions is less or equal to sum of maximum of this functions).

Lemma 8. *The following two statements are equivalent:*
(i) x_k is a solution of the problem (1)
(ii) $d_k = 0$ is a solution of the problem (7).

Proof. i) ⇒ ii): Suppose that $d_k \neq 0$ is a solution of the problem (7) and x_k is a solution of the problem (1). Then $\nabla F(x_k) = 0$ holds by Lemma 1 and by Lemma 2 it follows that

$$\tilde{\Phi}_k(d_k) = F(x_k) + \nabla F(x_k)^T d_k + \frac{1}{2} F''_D(x_k, d_k) = F(x_k) + \frac{1}{2} F''_D(x_k, d_k) \geq F(x_k) = \tilde{\Phi}_k(0) \tag{10}$$

which contradicts $\tilde{\Phi}_k(d_k) = \min_{d \in R^n} \tilde{\Phi}_k(d) \leq \tilde{\Phi}_k(0) = F(x_k)$. So $d_k = 0$ is a solution of the problem (7).

ii) ⇒ i): Suppose that $d_k = 0$ is a solution of the problem (7). Then we have that $0 \in \partial\tilde{\Phi}_k(d_k)$ holds, and then there exists some $\hat{V}_k \in \partial^2 F(x_k)$ such that $0 = \hat{V}_k d_k + \nabla F(x_k)$. Since by assumption $d_k = 0$ it follows that $0 = \nabla F(x_k)$ which means (by Lemma 1) that x_k is a solution of the problem (1).

Lemma 9. *If x_k is not a solution of the problem (1), then a minimum point d_k of the function $\tilde{\Phi}_k(d) = F(x_k) + \nabla F(x_k)^T d + \frac{1}{2} F''_D(x_k, d)$ over the set R^n is an non ascent direction of the function F at the point x_k.*

Proof. If d_k is a solution of the problem (7) then it follows that $0 \in \partial\tilde{\Phi}_k(d_k)$, i.e. $0 \in \nabla F(x_k) + \frac{1}{2}\partial(F''_D(x_k, d_k)) \Rightarrow 0 \in \nabla F(x_k) + \frac{1}{2}\partial(\max_{V \in \partial^2 F(x_k)} d_k^T V d_k)$. Then for some $\hat{G}_k \in \frac{1}{2}\partial(\max_{V \in \partial^2 F(x_k)} d_k^T V d_k)$, where $\hat{G}_k = \hat{V}_k d_k$, $\hat{V}_k = \sum_{j=1}^{r} \lambda_j V_j$, $r \leq n^2 + 1$, $V_j \in \partial^2 F(x_k)$, it follows that

$$\hat{V}_k d_k = -\nabla F(x_k) \tag{11}$$

holds.

If $\|\hat{V}_k\| = 0$ then from (11) by consistency of norms (if we take norm $\|\cdot\|_2$, see [24], page 5) it follows that $\|\nabla F(x_k)\| = 0$ (because of $\|\nabla F(x_k)\| = \|\hat{V}_k d_k\| \leq \|\hat{V}_k\| \, \|d_k\| = 0$). Hence by (6) and Lemma 1 it follows that x_k is a solution of the problem (1). So, since by assumption x_k is not a solution of the problem (1), then $\|\hat{V}_k\| \neq 0$. From (11) by Lemma 2 then it follows that $0 \leq d_k^T \hat{V}_k d_k = -\nabla F(x_k)^T d_k$, i.e.$\nabla F(x_k)^T d_k \leq 0$. Hence, d_k is an non ascent direction of the function F at the point x_k.

We will present the algorithm now.

Algorithm 1.

Step 1 *Given* $x_0 \in R^n$, $\varepsilon > 0$. *Set* $k = 0$.

Step 2 *Compute* $F(x_k)$ *and* $\nabla F(x_k)$.

If $\|\nabla F(x_k)\| < \varepsilon$ *then stop. The point* x_k *is a solution of the problem (1).*

Otherwise solve the problem (7), i.e.:

$$\min_{d \in R^n} \tilde{\Phi}_k(d), \tilde{\Phi}_k(d) = F(x_k) + \nabla F(x_k)^T d + \frac{1}{2} F''_D(x_k, d)$$

and denote by d_k *its solution.*

Step 3 *If* $\|d_k\| < \varepsilon$ *then stop. The point* x_k *is a solution of the problem (1). Else go to Step 4.*

Step 4 *Set* $x_{k+1} = x_k + d_k$ *and* $k = k + 1$. *Go to Step 2.*

Remark 2. *If Algorithm terminates at the Step 2, then by Lemma 1 the point* x_k *is a solution of the problem (1). If Algorithm terminates at the Step 3, then by Lemma 8 the point* x_k *is a solution of the problem (1).*

Theorem 1. *Let* $f : R^n \to R$ *be a closed convex proper function and* F *be its Moreau –Yosida regularization for* $M = I$. *Let the sequence* $\{x_k\}$ *be generated by Algorithm. Suppose that* $\{x_k\} \subseteq B$, *where* B *is a compact. Then for any initial point* $x_0 \in R^n$, $x_k \to x_\infty$ *as* $k \to +\infty$, *where* x_∞ *is a minimal point of the function* f.

Proof. Since we suppose that $\{x_k\} \subseteq B$, where B is a compact, then there exist accumulation points of the sequence $\{x_k\}$. Since ∇F is continuous, then, if $\nabla F(x_k) \to 0, k \to +\infty$ then it follows that every accumulation point x_∞ of the sequence $\{x_k\}$ satisfies $\nabla F(x_\infty) = 0$. Hence, by Lemma 1 the point x_∞ is a minimizer of F and, also, a minimizer of the function f.

Therefore we have to prove that $\nabla F(x_k) \to 0, k \to +\infty$.

Since from Algorithm we have that $x_{k+1} = x_k + d_k$, then it follows that $d_k \to 0$ as $k \to +\infty$, i.e. $d_\infty = 0$. From Algorithm we have that $\nabla F(x_k)^T d_k \leq 0$ and $\nabla F(x_\infty)^T d_\infty \leq 0$ as $k \to +\infty$. If we denote by $\Phi_\infty(d_\infty) = F(x_\infty) + \nabla F(x_\infty)^T d_\infty + \frac{1}{2} F''_D(x_\infty, d_\infty) = F(x_\infty)$ then according to the fact that $F''_D(x, d)$ is upper semicontinous with respect to (x, d) i.e. if $x_i \to x$ and $d_i \to d$ then $\limsup_{i\to\infty} F''_D(x_i, d_i) \leq F''_D(x, d)$, and by Lemma 3 and Lemma 2, we get

$$\begin{aligned}
F(x_\infty) = \tilde{\Phi}_\infty(d_\infty) &\geq \limsup_{k\to\infty} \tilde{\Phi}_k(d_k) = \limsup_{k\to\infty} (F(x_k) + \nabla F(x_k)^T d_k + \frac{1}{2} F''_D(x_k, d_k)) \\
&\geq \liminf_{k\to\infty} (F(x_k) + \nabla F(x_k)^T d_k + \frac{1}{2} F''_D(x_k, d_k)) \\
&= \liminf_{k\to\infty} \tilde{\Phi}_k(d_k) \geq \liminf_{k\to\infty} (F(x_k) + \nabla F(x_k)^T d_k) = \lim_{k\to\infty} (F(x_k) + \nabla F(x_k)^T d_k) \\
&= F(x_\infty) + \nabla F(x_\infty)^T d_\infty = F(x_\infty) = \tilde{\Phi}_\infty(d_\infty)
\end{aligned}$$

i.e. $\tilde{\Phi}_\infty(d_\infty) = \lim_{k\to\infty} \tilde{\Phi}_k(d_k)$.

Since $g_k = 0 \in \partial \tilde{\Phi}_k(d_k)$ then by Theorem 24.4 1 in [21] page 249 it follows that $g_\infty = 0 \in \partial \tilde{\Phi}_\infty(d_\infty)$ as $k \to +\infty$. Then for some $\hat{G}_\infty \in \frac{1}{2} \partial(\max_{V \in \partial^2 F(x_k)} d_\infty^T V d_\infty)$, where $\hat{G}_\infty = \hat{V}_\infty d_\infty$, $\hat{V}_\infty = \sum_{j=1}^r \lambda_j V_j$, $r \leq n^2 + 1$, $V_j \in \partial^2 F(x_\infty)$, it follows that $\hat{V}_\infty d_\infty = -\nabla F(x_\infty)$ holds. Hence by consistency of norms we have that $\|\nabla F(x_\infty)\| = \|\hat{V}_\infty d_\infty\| \leq \|\hat{V}_\infty\| \|d_\infty\| = \|\hat{V}_\infty\| \cdot 0 = 0$, i.e. $\|\nabla F(x_\infty)\| = 0$.

7. Conclusion

The Moreau-Yosida regularization is a powerful tool for smoothing nondifferentiable functions. It allows us to transform the solving an NDO problem into the solving an LC^1 optimization problem using the properties of this regularization.

To our knowledge this is a new approach to solving NDO problems, and in some sense it is close to the proximal quasi Newton algorithm (see [6]).

Acknowledgment. This research was supported by the Science Fund of Serbia, grant number III-44007.

References

[1] Boyd S., Vandenberghe L.: *Convex Optimization*, Cambridge University Press, 2006.
[2] Browien J., Lewis A.: *Convex analysis and nonlinear optimization*, Canada, 1999.
[3] Clarke F. H.: *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.
[4] Ermoliev, Y. M.; Kryazhimskii, A. V.; Ruszczynskii, *A. Constraint Aggregation Principle in Convex Optimization.* Math. Programming 76 (3),pp 353-372, 1997.
[5] Демьянов В. Ф. Васильев: Недифференцируемая оптимизация, Москва, "Наука", 1981.
[6] Djuranovic Milicic Nada: *On An Optimization Algorithm For LC^1 Unconstrained Optimization*, Facta Universitatis Ser. Math. Inform. 20, pp 129-135, 2005.
[7] Fuduli A.: *Metodi Numerici per la Minimizzazione di Funzioni Convesse Nondifferenziabili*, PhD Thesis, Departimento di Electronica Informatica e Sistemistica, Univerzita della Calabria, 1998.
[8] Fuduli A., Gaudioso M.: *Tuning Strategy for the Proximity Parameter in Convex Minimization*, JOTA Journal on Optimization Theory and Applications, Vol. 130, No. 1. pp 95-112, July 2006.
[9] Иоффе А. Д., Тихомиров В. М.: Теория экстремальныих задач, Москва, "Наука", 1974.
[10] Кусарев А. Г., Кутателадзе С. С.: Субдифференциальное исчисление, Новосибирск, Издатеьство, "Наука", 1987.
[11] Kiwiel K.: *Convergence Of Approximate And Incremental Subgradient Methods For Convex Optimization*, SIAM J. Optim. Vol 14, No. 5, pp 807-840, 2004.
[12] Kiwiel, K. *Proximity Control in Bundle Methods for Convex Nondifferentiable Minimization*. Math. Programming 46 (1), pp 105-122, 1990.
[13] Lemarechal C., Sagastizabal C.: *Practical aspects of the Moreau-Yosida regularization I: theoretical preliminaries*, SIAM Journal of Optimization, Vol.7, pp 367-385, 1997.
[14] Lemarechal, C.; Nemirovskii, A.; Nesterov, *Y. New Variants of Bundle Methods*. Math.Programming. 69 (1), pp 111-147, 1995.
[15] Lemarechal, C.; Strodiot, J.; Bihain, A. *On a Bundle Method for Nonsmooth Optimization*. In Nonlinear Programming 4; Mangasarian, O. L.; Meyer, R. R.; Robinson,S. M. Eds. pp 245-282.Academic Press: New York, 1981.
[16] Makela M., Neittaanmaki P.: Nonsmooth Optimization, World Scientific Publishing Co.Pte.Ltd., 1992.
[17] Meng F., Zhao G.:*On Second-order Properties of the Moreau-Yosida Regularization for Constrained Nonsmooth Convex Programs,* Numer. Funct. Anal. Optim , VOL 25; NUMB 5/6, pages 515-530, 2004.
[18] Пшеничний Б. Н.: Неопходимые условя экстремума, "Наука", 1982.
[19] Пшеничний Б. Н.: Метод линеаризации, Москва, "Наука", 1983.
[20] Qi L.: *Second Order Analysis of the Moreau-Yosida Regularization*, School of Mathematics, The University of New Wales Sydney, New South Wales 2052, Australia, 1998.
[21] Rockafellar R.: *Convex Analysis*, Princeton University Press, New Jersey, 1970.
[22] Sagastizabal C., Solodov M.: An Infeasible Bundle Method For Nonsmooth Convex Constrained Optimization Without A Penality Function Or A Filter, SIAM Journal of Optimization, Vol. 16, No. 1., pp 146-169, 2005.
[23] Sun W., Sampaio R.J.B., Yuan J.: *Two Algorithms for LC^1 Unconstrained Optimization*, Journal of Computational Mathematics, Vol. 18, No. 6, pp. 621-632, 2000.
[24] Sun W., Yuan Y.: *Optimization theory and methods*, Nonlinear programming, ISBN-10:0-387-24975-3 Springer Science+Business Media, LLC, 2006.
[25] Sun D., Sun J.: *Semismooth matrix-valued functions*, Math. Of Oper. Res., Vol. 27, No. 1, pp 150-169, 2002.
[26] Shengyuan S.: *Smooth Convex Approximation and its Applications*, Departmens of Mathematics National University of Singapore, 2004.