FACTA UNIVERSITATIS (NIŠ)
Ser. Math. Inform. 25 (2010), 11–24

A MULTI-STEP CURVE SEARCH ALGORITHM IN NONLINEAR OPTIMIZATION: NONDIFFERENTIABLE CONVEX CASE

Nada I. Đuranović-Miličić and Milanka Gardašević-Filipović

Abstract. In this paper a multi-step algorithm for minimization of a nondifferentiable function is presented. It is based on the results from [5] and [6]. The algorithm uses the Moreau-Yosida regularization of the objective function and its second order Dini upper directional derivative. This method uses previous multi-step iterative information and curve search to generate new iterative points. It is proved that the algorithm is well defined, as well as the convergence of the sequence of points generated by the algorithm to an optimal point. An estimate of the rate of convergence is given, too.

Keywords. multi-step, Moreau-Yosida regularization, unconstrained non-smooth convex optimization, second order Dini upper directional derivative.

1. Introduction

The following minimization problem is considered:

$$\min_{x \in R^n} f(x) \tag{1.1}$$

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

For nonsmooth programs, many approaches have been presented so far and they are often restricted to the convex unconstrained case. It is reasonable because a constrained problem can be easily transformed to an unconstrained problem using a distance function. In general, the various approaches are based on combinations of the following methods: subgradient methods; bundle techniques and the Moreau-Yosida regularization.

For a function f it is very important that its Moreau-Yosida regularization is a new function which has the same set of minima as f and is differentiable with Lipschitz

Received November 13, 2009.
2010 *Mathematics Subject Classification.* Primary 65K05; 90C30; Secondary 90C25

continuous gradient, even when f is not differentiable. In [13], [14] and [23] the second order properties of the Moreau-Yosida regularization of a given function f are considered.

Having in mind that the Moreau-Yosida regularization of a proper closed convex function is an LC^1 function, we present an optimization algorithm (using the second order Dini upper directional derivative (described in [1] and [2])) based on the results from [5]. That is the main idea of this paper.

We shall present an iterative algorithm for finding an optimal solution of problem (1.1) by generating the sequence of points $\{x_k\}$ of the following form:

$$x_{k+1} = x_k + \alpha_k s_k + \alpha_k^2 d_k \quad k = 0, 1, \ldots, s_k \neq 0, \quad d_k \neq 0 \tag{1.2}$$

where the step-size α_k and the directional vectors s_k and d_k are defined by the particular algorithms.

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 order Dini upper directional derivative and the basic properties are given; in the fifth section the semi-smooth functions and conditions for their minimization are described. Finally in the sixth section a model algorithm is suggested and its convergence is proved, and an estimate rate of its convergence is given, too.

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; $B(x, \rho)$ is the ball centred at x with radius ρ. 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 $f : R^n \to R \cup \{+\infty\}$ is the set $\mathrm{dom}(f) = \{x \in R^n | \, f(x) < +\infty\}$. We say that 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 $f : R^n \to R \cup \{+\infty\}$.

The *epigraph* of a given function $f : R^n \to R \cup \{+\infty\}$ is the set $\mathrm{epi}\, f = \{(\alpha, x) \in R \times R^n | \, \alpha \geqslant f(x)\}$. The concept of the epigraph gives us a possibility to define convexity and closure of a function in a new way. We say that f is convex if its epigraph is a convex set, and f is closed if its epigraph is a closed set.

In this section we will give the definitions and statements necessary in this work.

Definition 2.1. A vector $g \in R^n$ is said to be a *subgradient* of a given proper convex function $f : R^n \to R \cup \{+\infty\}$ at a point $x \in R^n$ if the next inequality

$$f(z) \geqslant f(x) + g^T \cdot (z - x) \tag{2.1}$$

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 $x \in \mathrm{dom}(f)$.

For a convex function f it follows that $f(x) = \max_{z \in R^n}\{f(z) + g^T(x - z)\}$ holds, where $g \in \partial f(z)$ (see [7]).
The concept of the subgradient is a simple generalization of the gradient for non-differentiable convex functions.

Lemma 2.1. *Let $f : S \to R \cup \{+\infty\}$ be a convex function defined on a convex set $S \subseteq R^n$, and $x' \in \mathrm{int}\, S$. Let $\{x_k\}$ be a sequence such that $x_k \to x'$, where $x_{k+1} = x_k + \varepsilon_k s_k + \varepsilon_k^2 d_k$, $k = 0, 1, \ldots, s_k \neq 0$, $d_k \neq 0$, $\varepsilon_k > 0$, $\varepsilon_k \to 0$ and $s_k \to s$, $d_k \to d$ and $g_k \in \partial f(x_k)$. Then all accumulation points of the sequence $\{g_k\}$ lie in the set $\partial f(x')$.*

Proof. Since $g_k \in \partial f(x_k)$ then the inequality $f(y) \geqslant f(x_k) + g_k^T \cdot (y - x_k)$ holds for any $y \in S$. Hence, taking any subsequence for which $g_k \to g'$ it follows that $f(y) \geqslant f(x') + g'^T \cdot (y - x')$, which means that $g' \in \partial f(x')$. □

Definition 2.2. 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

$$f'(x', s) = \lim_{t \downarrow 0} \frac{f(x' + t \cdot s) - f(x')}{t} \tag{2.2}$$

when this limit exists.

Hence, it follows that if the function f is convex and $x' \in \mathrm{dom}\, f$, then

$$f(x' + t \cdot s) = f(x') + t \cdot f'(x', s) + o(t) \tag{2.3}$$

holds, which can be considered as one linearization of the function f (see in [8]).

Lemma 2.2. *Let $f : S \to R \cup \{+\infty\}$ be a convex function defined on a convex set $S \subseteq R^n$, and $x' \in \mathrm{int}\, S$. If the sequence $x_k \to x'$, where $x_k = x' + \varepsilon_k s_k$, $\varepsilon_k > 0$, $\varepsilon_k \to 0$ and $s_k \to s$ then the next formula:*

$$f'(x', s) = \lim_{k \to \infty} \frac{f(x_k) - f(x')}{\varepsilon_k} = \max_{g \in \partial f(x')} s^T g \tag{2.4}$$

holds.

Proof. See in [9] or [17]. □

Lemma 2.3. *Let $f : S \to R \cup \{+\infty\}$ be a convex function defined on a convex set $S \subseteq R^n$. Then $\partial f(x)$ is bounded for $\forall x \in B \subset \operatorname{int} S$, where B is a compact.*

Proof. See in [10] or [12]. □

Proposition 2.1. *Let $f : R^n \to R \cup \{+\infty\}$ be a proper convex function. The condition:*

$$0 \in \partial f(x) \tag{2.5}$$

is a first order necessary and sufficient condition for a global minimizer at $x \in R^n$. This can be stated alternatively as:

$$\forall s \in R^n, \quad \|s\| = 1 \max_{g \in \partial f(x)} s^T g \geqslant 0. \tag{2.6}$$

Proof. See [16]. □

Lemma 2.4. *If a proper convex function $f : R^n \to R \cup \{+\infty\}$ is a differentiable function at a point $x \in \operatorname{dom}(f)$, then:*

$$\partial f(x) = \{\nabla f(x)\}. \tag{2.7}$$

Proof. The statement follows directly from Definition 2.2. □

Lemma 2.5. *Let $f_i : R^n \to R \cup \{+\infty\}$ for $i \in \{1, 2, \dots, n\}$, $n \in N$ be convex functions, and $f(x) = \max_{i \in \{1,2,\dots,n\}} f_i(x)$. Then the function f is a convex function, and its subgradient at the point $x \in R^n$, i.e. $g \in \partial f(x)$ is given as follows:*

$$g = \left\{ \sum_{i \in \hat{I}} \lambda_i g_i \middle| \sum_{i \in \hat{I}} \lambda_i = 1, \quad \lambda_i \geqslant 0, \quad g_i \in \partial f_i(x) \quad \textit{for} \quad i \in \hat{I} \right\} \tag{2.8}$$

where $\hat{I}$ is the set $\hat{I} = \{i \in I | f(x) = f_i(x)\}$.

Proof. See in [7]. □

Definition 2.3. The real function f defined on R^n is LC^1 function on the open set $D \subseteq R^n$ if it is continuously differentiable and its gradient ∇f is locally Lipschitz, i.e.

$$\|\nabla f(x) - \nabla f(y)\| \leqslant L\|x - y\| \quad \text{for} \quad x, y \in D \tag{2.9}$$

for some $L > 0$.

3. The Moreau-Yosida regularization

Definition 3.1. Let $f : R^n \to R \cup \{+\infty\}$ 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:

$$F(x) := \min_{y\in R^n}\left\{f(y) + \frac{1}{2}\|y - x\|_M^2\right\} \tag{3.1}$$

The above function is an *infimal convolution*. In [18] it is proved that the infimal convolution of a convex function is also a convex function. Hence the function defined by (3.1) is a convex function and has the same set of minima as the function f (see in [8]), so the motivation of the study of Moreau-Yosida regularization is due to the fact that $\min_{x\in R^n} f(x)$ is equal to $\min_{x\in R^n} F(x)$.

Definition 3.2. The minimum point $p(x)$ of the function (3.1):

$$p(x) := \arg\min_{y\in R^n}\left\{f(y) + \frac{1}{2}\|y - x\|_M^2\right\} \tag{3.2}$$

is called the *proximal point* of x.

Proposition 3.1. *The function Fdefined by (3.1) is always differentiable.*

Proof. See in [8]. □

The first order regularity of F is well known (see in [8] and [13]): without any further assumptions, F has a Lipschitzian gradient on the whole space R^n. More precisely, for all $x_1, x_2 \in R^n$ the next formula:

$$\|\nabla F(x_1) - \nabla F(x_2)\|^2 \leqslant \Lambda\langle\nabla F(x_1) - \nabla F(x_2), x_1 - x_2\rangle \tag{3.3}$$

holds (see in [13]), where $\nabla F(x)$ has the following form:

$$G := \nabla F(x) = M(x - p(x)) \in \partial f(p(x)) \tag{3.4}$$

and $p(x)$ is the unique minimum in (3.1). So, according to above consideration and Definition 2.3, we conclude that F is an LC^1 function (see in [14]).
Note 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 (3.1) then (see in [8] and [13]):

$$p(x) = x - M^{-1}g \quad \text{where} \quad g \in \partial f(p(x)). \tag{3.5}$$

In [13] it is also proved that for all $x_1, x_2 \in R^n$ the next formula:

$$\|p(x_1) - p(x_2)\|_M^2 \leqslant \langle M(x_1 - x_2), p(x_1) - p(x_2)\rangle \tag{3.6}$$

is valid, namely the mapping $x \to p(x)$, where $p(x)$ is defined by (3.2), is Lipschitzian with constant $\dfrac{\Lambda}{\lambda}$ (see Proposition 2.3. in [13]).

Lemma 3.1. *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 [8] or [23]. □

4. Dini second upper directional derivative

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

Definition 4.1. [22] The *second order Dini upper directional derivative* of the function $f \in LC^1$ at the point $x \in R^n$ in the direction $d \in R^n$ is defined to be

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

If ∇f is directionally differentiable at x_k, we have that

$$f''_D(x_k,d) = f''(x_k,d) = \lim_{\alpha \downarrow 0} \frac{[\nabla f(x + \alpha d) - \nabla f(x)]^T \cdot d}{\alpha}$$

for all $d \in R^n$.

Since the Moreau-Yosida regularization of a proper closed convex function f is an LC^1 function, we can consider its second order Dini upper directional derivative at the point $x \in R^n$ in the direction $d \in R^n$. Using (3.4) we can state that:

$$F''_D(x,d) = \limsup_{\alpha \downarrow 0} \frac{g_1 - g_2}{\alpha} d,$$

where $F(x)$ is defined by (3.1) and $g_1 \in \partial f(p(x + \alpha d))$, $g_2 \in \partial f(p(x))$.

Lemma 4.1. *Let* $f : R^n \to R$ *be a closed convex proper function and* F *is its Moreau – Yosida regularization. Then the next statements are valid.*

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

(ii) $F''_D(x_k, d_1 + d_2) \leqslant 2(F''_D(x_k, d_1) + F''_D(x_k, d_2))$

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

Proof. See in [22] and [2]. □

Lemma 4.2. *Let* $f : R^n \to R$ *be a closed convex proper function and let* F *be its Moreau – Yosida regularization. Then the next statements are valid.*

(i) $F''_D(x,d)$ *is upper semicontinuous with respect to* (x,d), *i.e.* $\limsup_{i\to\infty} F''_D(x_i,d_i) \leqslant F''_D(x,d)$ *when* $(x_i,d_i) \to (x,d)$

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

Proof. See in [22] and [2]. □

5. Semi-smooth functions and optimality conditions

Definition 5.1. 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 $x \in R^n$ and the limit $\lim_{\substack{h \to d \\ \lambda \downarrow 0}} \{Vh\}, V \in \partial^2 F(x + \lambda h)$ 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.

Lemma 5.1. [22]: *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 that $h^T Vh - F''(x, h) = o(\|h\|^2)$.*

Lemma 5.2. *Let $f : R^n \to R$ be a proper closed convex function and let F be its Moreau-Yosida regularization. So, if $x \in R^n$ is solution of the problem (1.1) then $F'(x, d) = 0$ and $F''_D(x, d) \geqslant 0$ for all $d \in R^n$.*

Proof. See in [6]. □

Lemma 5.3. *Let $f : R^n \to R$ be a proper closed convex function, F its Moreau-Yosida regularization, and x a point from R^n. If $F'(x, d) = 0$ and $F''_D(x, d) > 0$ for all $d \in R^n$, then $x \in R^n$ is a strict local minimizer of the problem (1.1).*

Proof. See in [6]. □

6. A model algorithm

In this section an algorithm for solving the problem (1.1) is introduced. We suppose that at each $x \in R^n$ it is possible to compute $f(x)$, $F(x)$, $\nabla F(x)$ and $F''_D(x, d)$ for a given $d \in R^n$.

At the k-th iteration we consider the following problem

$$\min_{d \in R^n} \Phi_k(d), \quad \Phi_k(d) = \nabla F(x_k)^T d + \frac{1}{2} F''_D(x_k, d) \tag{6.1}$$

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 a Lipschitzian constant for F it is also a Lipschitzian constant for ∇F. The function $\Phi_k(d)$ is called an iteration function. It is easy to see that $\Phi_k(0) = 0$ and $\Phi_k(d)$ is Lipschitzian on R^n.

We generate the sequence $\{x_k\}$ of the form $x_{k+1} = x_k + \alpha_k s_k + \alpha_k^2 d_k$, where the directional vectors s_k and d_k are defined by the particular algorithms called Direction vector rule 1 and 2, and the step-size α_k is defined by the particular algorithm called Curve search rule.

We suppose that at the k-th iteration, there is an index set $I_k = \{1, 2, \ldots, k\}$, and we store information as a bundle $B_k = \{(x_i, f(x_i), g_i) | \ i \in I_k\}$ i.e a set of triplets indexed by I_k consisting of the generic point x_i, the value $f(x_i)$ of the objective function f at the point x_i, and an arbitrary subgradient $g_i \in \partial f(x_i)$. Each triplet in the bundle B_k defines one linearization $f_i(x)$ of the objective function given as

$$f_i(x) = f(x_i) + g_i^T(x - x_i),$$

where $i \in I_k$. If f is a convex function then $f(x) = \max_{z \in R^n}\{f(z) + g^T(x - z)\}$ holds, where $g \in \partial f(z)$. Hence, we have that the function

$$\hat{f}_k(x) = \max_{0 \leqslant i \leqslant k} f_i(x) = \max_{0 \leqslant i \leqslant k}\{f(x_i) + g_i^T(x - x_i)\}$$

(which is, in literature, known as a cutting plane function) is a good approximation of the function f. It is easy to see that $f(x) \geqslant \hat{f}_{k+1}(x) \geqslant \hat{f}_k(x)$ hold for all $x \in R^n$.

Function $\hat{f}_k(x)$ is a polyhedral function (piecewise linear function) and hence it is a closed convex function. More than that, $\hat{f}_k(x)$ could be considered as a composition of the linear functions, i.e. $\max_{0 \leqslant i \leqslant k} f_i(x)$. If $\hat{g} \in \partial \hat{f}_k(x)$ then since $\hat{f}_k(x)$ is a polyhedral function according to Lemma 2.5 we have that $\hat{g} = \sum_{i \in \hat{I}_k} \lambda_i g_i$, where $\hat{I}_k = \{i \in I_k | \ \hat{f}_k(x) = f_i(x)\}$ and $g_i \in \partial f(x_i)$, $i \in \hat{I}_k$, $\sum_{i \in \hat{I}_k} \lambda_i = 1$, $\lambda_i \geqslant 0$, i.e. $\hat{g}$ is a convex combination of the subgradients from the bundle B_k.

Lemma 6.1. *Let $f : R^n \to R$ be a proper closed convex function, $\{x_k\}$ is a sequence of points from R^n and $g_i \in \partial f(x_i)$. Let*

$$\hat{f}_k(x) = \max_{0 \leqslant i \leqslant k} f_i(x) = \max_{0 \leqslant i \leqslant k}\{f(x_i) + g_i^T(x - x_i)\}$$

be a polyhedral function. If $\hat{g} \in \partial \hat{f}_k(x)$ then $\hat{g} \in \partial f(x)$ for any $x \in R^n$.

Proof. Let $\hat{g} \in \partial \hat{f}_k(y)$ for some $y \in R^n$. Then by Definition 2.1 the inequality

$$\hat{f}_k(x) \geqslant \hat{f}_k(y) + \hat{g}^T(x - y)$$

holds for any $x \in R^n$. If we suppose that $\hat{g} \notin \partial f(y)$ then by Definition 2.1 the next inequality $f(x) < f(y) + \hat{g}^T(x - y)$ holds for any $x \in R^n$. So, we get that

$$f(x) - f(y) < \hat{g}^T(x - y) \leqslant \hat{f}_k(x) - \hat{f}_k(y)$$

hold, and hence $f(x) - \hat{f}_k(x) < f(y) - \hat{f}_k(y)$. Since $f(x) \geqslant \hat{f}_{k+1}(x) \geqslant \hat{f}_k(x)$ hold for all $x \in R^n$ then from $f(x) - \hat{f}_k(x) < f(y) - \hat{f}_k(y)$ it follows that $f(y) < \hat{f}_k(y)$ holds for some $y \in R^n$ which contradicts $f(x) \geqslant \hat{f}_{k+1}(x) \geqslant \hat{f}_k(x)$ for all $x \in R^n$. □

Algorithm. Let $0 < \rho < 1$, $0 < \sigma < 1$, $x_1 \in R^n$, and ε and μ be real positive numbers small enough, $k := 1$ and $I_0 = \emptyset$, $B_0 = \emptyset$.

Step 1. For a given x_k calculate $f_k = f(x_k)$ and $g_k = g(x_k)$. Set $I_k = \{k\} \cup I_{k-1} \backslash S_k$, where $S_k = \{i \in I_{k-1} | \ \|x_i - x_k\| \geqslant \mu\}$. Set $B_k = \{(x_i, f(x_i), g_i) | \ i \in I_k\}$.

Step 2. If $\|g_k\| \leqslant \varepsilon$ then STOP; else solve the problem $\min \left\| \sum_{i \in \hat{I}_k} \lambda_i g_i \right\|$ such that $\sum_{i \in \hat{I}_k} \lambda_i = 1$, $\lambda_i \geqslant 0$ where $\hat{I}_k = \{i \in I_k | \ \hat{f}_k(x) = f_i(x)\}$, $\hat{f}_k(x) = \max_{0 \leqslant i \leqslant k}\{f(x_i) + g_i^T(x - x_i)\}$ and $g_i \in \partial f(x_i)$, $i \in \hat{I}_k$ and denote by $\lambda_i^{(k)}$ its solution. If $\left\| \sum_{i \in \hat{I}_k} \lambda_i^{(k)} g_i \right\| \leqslant \varepsilon$, then stop. Otherwise go to step 3.

Step 3. Set $x_{k+1} = x_k + \alpha_k s_k(\alpha_k) + \alpha_k^2 d_k(\alpha_k)$, where α_k is selected by the Curve search rule, and $s_k(\alpha_k)$ and $d_k(\alpha_k)$ are computed by the Direction vector rules 1 and 2. For simplicity we denote $s_k(\alpha_k)$ by s_k, $d_k(\alpha_k)$ by d_k and $g(x_k)$ by g_k.

Curve search rule:

Choose $\alpha_k = q^{i(k)}$, $0 < q < 1$, where $i(k)$ is the smallest integer from $\{0, 1, 2, \ldots\}$ such that

$$F(x_k) - F(x_k + q^{i(k)} s_k + q^{2i(k)} d_k) \geqslant \sigma \left(-q^{i(k)} g_k^T s_k + \frac{1}{2} q^{4i(k)} F_D''(x_k; d_k) \right). \tag{6.2}$$

Direction vector rule 1:

$$s_k(\alpha) = \begin{cases} s_k^\star & k \leqslant m-1 \\ -\left[\left(1 - \sum_{i=2}^m \alpha^{i-1} p_k^i \right) g_k + \sum_{i=2}^m \alpha^{i-1} p_k^i s_{k-i+1} \right] & k \geqslant mk \end{cases}$$

where $m = \operatorname{card} I_k$, $m > 1$,

$$p_k^i = \frac{\rho \|g_k\|^2}{(m-1)\left[\|g_k\|^2 + |g_k^T s_{k-i+1}| \right]}, \quad i = 2, 3, \ldots, m,$$

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

Direction vector rule 2.

The direction vector d_k^*, $k \leqslant m-1$, presents a solution of the problem (6.1) and

$$d_k(\alpha) = \begin{cases} d_k^*, & k \leqslant m-1 \\ \sum_{i=2}^m \alpha^{i-1} d_{k-i+1}^*, & k \geqslant m. \end{cases}$$

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

We make the following assumptions.

A1. We suppose that there exist constants $c_2 \geqslant c_1 > 0$ such that $c_1 \|d\|^2 \leqslant F''_D(x_k, d) \leqslant c_2 \|d\|^2$ hold for every $d \in R^n$.

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

Lemma 6.2. *Under the assumption A1 the function $\Phi_k(\cdot)$ is coercive.*

Proof. See in [6]. □

Remark 6.1. Coercivity of the function $\Phi_k(\cdot)$ assures that the optimal solution of the problem (6.1) exists (see in [23]).

Proposition 6.1. *If the Moreau-Yosida regularization $F(\cdot)$ of the proper closed convex function $f(\cdot)$ satisfies the condition A1, then:*

(i) *the function $F(\cdot)$ is uniformly and, hence, strictly convex;*

(ii) *the level set $L(x_0) = \{x \in R^n : F(x) \leqslant F(x_0)\}$ is a compact convex set, and*

(iii) *there exists a unique point x^* such that $F(x^*) = \min_{x \in L(x_0)} F(x)$.*

Proof. See in [6]. □

Lemma 6.3. *The following statements are equivalent:*

(i) *$d = 0$ is globally optimal solution of the problem (6.1)*

(ii) *0 is the optimum of the objective function in (6.1)*

(iii) *the corresponding x_k is such that $0 \in \partial f(x_k)$*

Proof. See in [6]. □

Lemma 6.4. *For $\alpha \in [0,1]$ and all $k \geqslant m$, we have $g_k^T s_k(\alpha) \leqslant -(1-\rho)\|g_k\|^2$.*

Proof. >From Direction rule 1 we have for $k \geqslant m$ that:

$$g_k^T s_k = g_k^T \left\{ -\left[\left(1 - \sum_{i=2}^m \alpha^{i-1} p_k^i \right) g_k + \sum_{i=2}^m \alpha^{i-1} p_k^i s_{k-i+1} \right] \right\} =$$

$$= -\|g_k\|^2 + \frac{\rho \|g_k\|^2}{m-1} \sum_{i=2}^m \alpha^{i-1} \frac{\|g_k\|^2 - g_k^T s_{k-i+1}}{\|g_k\|^2 + |g_k^T s_{k-i+1}|}$$

$$= -\|g_k\|^2 + \frac{\rho \|g_k\|^2}{m-1} \sum_{i=1}^{m-1} \alpha^i (\text{ since } g_k^T s_{k-i+1} \leqslant 0) = -\|g_k\|^2 + \frac{\rho \|g_k\|^2}{m-1} \alpha \frac{1-\alpha^{m-1}}{1-\alpha} \leqslant -(1-\rho)\|g_k\|^2$$

where the last inequality holds by assumptions. □

(From assumptions $\alpha \in [0,1]$ and $m > 1$ it follows that $\alpha^{m-1} \leqslant \alpha$, $\alpha\dfrac{1-\alpha^{m-1}}{1-\alpha} \leqslant 1$ and $\dfrac{1}{m-1} \leqslant 1$.)

Convergence theorem. Suppose that f is a proper closed convex function and F is its Moreau-Yosida regularization, and the assumptions A1 and A2 hold. Then for any initial point $x_0 \in R^n$, $x_k \to \bar{x}$, as $k \to \infty$, where $\bar{x}$ is a unique minimal point.

Proof. If $d_k \neq 0$ is a solution of (6.1), it follows that $\Phi_k(d_k) \leqslant 0 = \Phi_k(0)$. Consequently, we have by assumption A1 that

$$g(x_k)^T d_k \leqslant -\frac{1}{2}F_D''(x_k; d_k) \leqslant -\frac{1}{2}c_1\|d_k\|^2 < 0, \tag{6.3}$$

i.e. d_k is a descent direction at $x_k$for the function F. From (6.2), A1 and Lemma 6.4 it follows that

$$\begin{aligned} F(x_k) - F(x_{k+1}) &\geqslant \sigma\left[-q^{i(k)} g_k^T s_k + \frac{1}{2}q^{4i(k)}F_D''(x_k; d_k)\right] \geqslant \\ &\geqslant q^{i(k)}\sigma(1-\rho)\|g_k\|^2 + \frac{\sigma}{2}q^{4i(k)}c_1\|d_k\|^2 > 0. \end{aligned} \tag{6.4}$$

Hence $\{F(x_k)\}$ is a decreasing sequence and consequently $\{x_k\} \subset L(x_0)$. Since$L(x_0)$ by Proposition 6.1 is a compact convex set, it follows that the sequence $\{x_k\}$ is bounded. Therefore there exist accumulation points of $\{x_k\}$. Since the gradient $G = \nabla F$ is by assumption continuous (because $F \in LC^1$), then, if $G_k = \nabla F(x_k) \to 0$ as $k \to \infty$, it follows that every accumulation point $\bar{x}$ of the sequence $\{x_k\}$ satisfies $\bar{G} = \nabla F(\bar{x}) = 0$ and hence by Lemma 3.1 follows that $\bar{x}$ is the minimum point of the function f. Since F by Proposition 6.1 is strictly convex, it follows that there exists a unique point $\bar{x} \in L(x_0)$ such that $\bar{G} = \nabla F(\bar{x}) = 0$. □

Hence, $\{x_k\}$ has a unique limit point $\bar{x}$ and it is a global minimizer. Therefore we have to prove that $G_k = \nabla F(x_k) \to 0$ as $k \to \infty$. There are two cases to consider.

a) The set of indices $\{i(k)\}$ for $k \in K_1$, is uniformly bounded above by a number I, i.e. $i(k) \leqslant I < \infty$ for $k \in K_1$. Consequently, from (6.2) and (6.4) and since $g(x_k)^T s_k \leqslant 0$ and $F_D''(x_k; d_k) > 0$ it follows that

$$\begin{aligned} F(x_k) - F(x_{k+1}) &\geqslant \sigma\left[-q^{i(k)} g_k^T s_k + \frac{1}{2}q^{4i(k)}F_D''(x_k; d_k)\right] \geqslant \\ \geqslant \sigma\left[-q^{I} g_k^T s_k + \frac{1}{2}q^{4I}F_D''(x_k; d_k)\right] &\geqslant q^{I}\sigma(1-\rho)\|g_k\|^2 + \frac{\sigma}{2}q^{4I}F_D''(x_k; d_k). \end{aligned} \tag{6.5}$$

Since $\{F(x_k)\}$ is bounded below (on the compact set $L(x_0)$) and monotone (by (6.4)), it follows that $F(x_{k+1}) - F(x_k) \to 0$ as $k \to \infty$, $k \in K_1$. Hence from (6.5) it follows that $\|g(x_k)\| \to 0$ and $F_D''(x_k, d_k) \to 0$, $k \to \infty$, $k \in K_1$.

b) There is a subset $K_2 \subset K_1$ such that $\lim_{k\to\infty} i(k) = \infty$.

This part of proof is analogous to the proof in [6].

In order to have a finite value $i(k)$, it is sufficient that s_k and d_k have descent properties, i.e.$g(x_k)^T s_k < 0$ and $g(x_k)^T d_k < 0$ whenever $g(x_k) \neq 0$. The first relation follows from Lemma 6.4 and the second follows from (6.3). At a saddle point the relation (6.2) becomes

$$F(x_k) - F(x_{k+1}) \geqslant \sigma \left[\frac{1}{2} q^{4i(k)} F''_D(x_k; d_k) \right] \tag{6.6}$$

In that case by Lemma 6.3 $d_k \neq 0$ and hence, by A1 it follows that $F''_D(x_k; d_k) > 0$; so (6.6) clearly can be satisfied.

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

$$F(x_n) - F(\bar{x}) \leqslant \mu_0 \left[1 + \frac{\mu_0}{\eta^2} \sum_{k=0}^{n-1} \frac{F(x_k) - F(x_{k+1})}{\|\nabla F(x_k)\|^2} \right]^{-1},$$

$n = 1, 2, \ldots$ where $\mu_0 = F(x_0) - F(\bar{x})$, and $diam L(x_0) = \eta < \infty$ since by Proposition 6.1 it follows that $L(x_0)$ is bounded.

Proof. The proof directly follows from the Theorem 9.2, page 167 in [11], since the assumptions of that theorem are fulfilled. □

CONCLUSION

The algorithm presented in this paper is based on the algorithms from [21], [5] and [6]. 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 algorithm in [21], the presented algorithm is defined and converges under weaker assumptions than the algorithm given in [21]. Relating to the algorithm in [5], the presented algorithm is defined and converges for nondifferentiable convex function.

ACKNOWLEDGMENT

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

REFERENCES

1. J. M. Borwein and A. S. Lewis: *Convex Analysis and Nonlinear Optimization: Theory and Examples*. CMS Books in Mathematics, Springer, New York, 2000.
2. F. H. Clarke: *Optimization and Nonsmooth Analysis*. Wiley, New York, 1983.
3. N. I. Đuranović-Miličić: *On an optimization algorithm for LC^1 unconstrained optimization*. Facta Universitatis, Ser. Math. Inform. 20 (2005), 129–135.

4. N. I. Đuranović-Miličić: *An algorithm for LC^1 optimization*. YUJOR 15 (2) (2005), 301–307.
5. N. I. Đuranović-Miličić: *A multi-step curve search algorithm in nonlinear optimization*. YUJOR 18 (1) (2008), 47–52.
6. N. I. Đuranović-Miličić and M. Gardašević-Filipović: *An algorithm for minimization of a nondifferentiable convex function*. Proceedings of the 2009 International Conference of Applied and Engineering Mathematics, World Congress on Engineering 2009, London, 1-3 July 2009, pp 1241–1246.
7. V. F. Demianov and L. V. Vasilev: *Nondifferentiable Optimization*. Nauka, Moscow, 1981 (in Russian).
8. A. Fuduli: *Metodi Numerici per la Minimizzazione di Funzioni Convesse Nondifferenziabili*. PhD Thesis, Departimento di Electronica Informatica e Sistemistica, Univerzita della Calabria, 1998.
9. R. Fletcher: *Practical Methods of Optimization, Volume 2 – Constrained Optimization*. Wiley-Interscience Publication, 1981.
10. A. D. Ioffe and V. M. Tikhomirov: *Theory of Extremal Problems*. Nauka, Moscow, 1974 (in Russian).
11. V. G. Karmanov: *Mathematical Programming*, Nauka, Moscow, 1975 (in Russian).
12. A. G. Kusarev and S. S. Kutateladze: *Subdifferential Calculus*, Nauka, Moscow, 1987 (in Russian).
13. C. Lemarechal and C. Sagastizabal: *Practical aspects of the Moreau-Yoshida regularization I: theoretical preliminaries*, SIAM Journal of Optimization 7 (1997), 367–385.
14. F. Meng and G. Zhao: *On Second-order Properties of the Moreau-Yoshida Regularization for Constrained Nonsmooth Convex Program*, Numer. Funct. Anal. Optim. 25 (5/6) (2004), 515–530.
15. B. N. Pshenichny: *Convex Analysis and Extremal Problems*. Nauka, Moscow, 1980 (in Russian).
16. B. N. Pshenichny: *Necessary Conditions for an Extremum*, Marcel Dekker, New York, 1971.
17. B. N. Pshenichny: *The Linearization Method*. Nauka, Moscow, 1983 (in Russian).
18. R. Rockafellar: *Convex Analysis*. Princeton University Press, New Jersey, 1970 (Russ. trans.: Mir, Moscow 1973).
19. E. Polak: *Computational Methods in Optimization. A Unified Approach*. Mathematics in Science and Engineering, Vol. 77, Academic Press, New York-London, 1971 (Russ. trans.: Mir, Moscow, 1974).
20. R. Pytlak: *Conjugate Gradient Algorithms in Nonconvex Optimization*, Springer-Verlag, Berlin-Heidelberg, 2009.
21. Z. J. Shi: *Convergence of multi-step curve search method for unconstrained optimization*, J. Numer. Math. 12 (4) (2004), 297–309.
22. W. Sun, R. J. B. Sampaio, and J. Yuan: *Two algorithms for LC^1 unconstrained optimization*, Journal of Computational Mathematics 18 (6) (2000), 621–632.
23. L. Qi: *Second Order Analysis of the Moreau-Yoshida Regularization*, School of Mathematics, The University of New Wales Sydney, New South Wales 2052, Australia, 1998.

Nada I. Đuranović-Miličić
Department of Mathematics,
Faculty of Technology and Metallurgy
University of Belgrade, Belgrade, Serbia
nmilicic@tmf.bg.ac.rs

Milanka Gardašević-Filipović
Vocational College of Technology, Arandjelovac, Serbia
milankafilipovic@yahoo.com