An Algorithm For Minimization Of A Nondifferentiable Convex Function

Nada DJURANOVIC-MILICIC [1], *Milanka GARDASEVIC-FILIPOVIC* [2]

Abstract— **In this paper an algorithm for minimization of a nondifferentiable function is presented. The algorithm uses the Moreau-Yoshida regularization of the objective function and its second order Dini upper directional derivative. 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.**

***Index Terms*— Moreau-Yoshida regularization, non-smooth convex optimization, second order Dini upper directional derivative.**

I. Introduction

The following minimization problem is considered:

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

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. In general, the various approaches are based on combinations of the following methods: subgradient methods; bundle techniques and the Moreau-Yoshida regularization.

For a function f it is very important that its Moreau-Yoshida regularization is a new function which has the same set of minima as f and is differentiable with Lipschitz continuous gradient, even when f is not differentiable. In [10], [11] and [19] the second order properties of the Moreau-Yoshida regularization of a given function f are considered.

Having in mind that the Moreau-Yoshida 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 [3]. That is the main idea of this paper.

We shall present an iterative algorithm for finding an optimal solution of problem (1) by generating the sequence of points $\{x_k\}$ of the following form:

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

where the step-size α_k and the directional vector 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-Yoshida 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.

II. 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 $dom(f) = \{x \in R^n \mid f(x) < +\infty\}$. 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 $epi\, f = \{(\alpha, x) \in R \times R^n \mid \alpha \geq f(x)\}$. 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.

Manuscript received February 27, 2009. This work was supported by Science Fund of Serbia, grant number 144015G, through Institute of Mathematics, SANU

[1] Nada DJURANOVIC-MILICIC is with Faculty of Technology and Metallurgy, University of Belgrade, Serbia (corresponding author to provide phone: +381-11-33-03-865; e-mail: nmilicic@elab.tmf.bg.ac.rs

[2] Milanka GARDASEVIC-FILIPOVIC is with Faculty of Architecture, University of Belgrade, Serbia; e-mail: milanka@arh.bg.ac.rs

Definition 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

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

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 $x \in dom(f)$.

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 [4]).

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

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

<u>*Proof.*</u> See in [7] or [6].

Definition 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

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

when this limit exists.

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

$$f(x' + t \cdot s) = f(x') + t \cdot f'(x', s) + o(t) \tag{5}$$

holds, which can be considered as one linearization of the function f (see in [5]).

Lemma 2. Let $f : S \to R \cup \{+\infty\}$ be a convex function defined on a convex set $S \subseteq R^n$, and $x' \in int\, S$. 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:

$$f'(x', s) = \lim_{k \to \infty} \frac{f(x_k) - f(x')}{\varepsilon_k} = \max_{g \in \partial f(x')} s^T g \tag{6}$$

holds.

<u>*Proof.*</u> See in [6] or [14].

Lemma 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 $\forall x \in B \subset int\, S$, where B is a compact.

<u>*Proof.*</u> See in [7] or [9].

Proposition 1 Let $f : R^n \to R \cup \{+\infty\}$ be a proper convex function. The condition:

$$0 \in \partial f(x) \tag{7}$$

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

$$\forall s \in R^n,\ \|s\| = 1 \qquad \max_{g \in \partial f(x)} s^T g \geq 0 \, . \tag{8}$$

<u>*Proof.*</u> See [13].

Lemma 4. If a proper convex function $f : R^n \to R \cup \{+\infty\}$ is a differentiable function at a point $x \in dom(f)$, then:

$$\partial f(x) = \{\nabla f(x)\}. \tag{9}$$

<u>*Proof.*</u> The statement follows directly from Definition 2.

Definition 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.

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\| \text{ for } x, y \in D \tag{10}$$

for some $L > 0$.

III. THE MOREAU-YOSHIDA REGULARIZATION

Definition 4. Let $f : R^n \to R \cup \{+\infty\}$ be a proper closed convex function. The *Moreau-Yoshida regularization* of a given function f, associated to the metric defined by M, denoted by F, is defined as follows:

$$F(x) := \min_{y \in R^n} \left\{ f(y) + \frac{1}{2} \|y - x\|_M^2 \right\} \tag{11}$$

The above function is an *infimal convolution*. In [15] it is proved that the infimal convolution of a convex function is also a convex function. Hence the function defined by (11) is a convex function and has the same set of minima as the function f (see in [5]), so the motivation of the study of Moreau-Yoshida 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 5. The minimum point $p(x)$ of the function (11):

$$p(x) := \arg\min_{y \in R^n} \left\{ f(y) + \frac{1}{2} \|y - x\|_M^2 \right\} \tag{12}$$

is called the *proximal point* of x.

Proposition 2. The function F defined by (11) is always differentiable.

<u>*Proof.*</u> See in [5].

The first order regularity of F is well known (see in [5] and

[10]): 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:

$$\|\nabla F(x_1) - \nabla F(x_2)\|^2 \le \Lambda \langle \nabla F(x_1) - \nabla F(x_2), x_1 - x_2 \rangle \quad (13)$$

holds (see in [10]), where $\nabla F(x)$ has the following form:

$$G := \nabla F(x) = M(x - p(x)) \in \partial f(p(x)) \quad (14)$$

and $p(x)$ is the unique minimum in (11). So, according to above consideration and Definition 3, we conclude that F is an LC^1 function (see in [11]).

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 (11) then (see in [5] and [10]):

$$p(x) = x - M^{-1} g \text{ where } g \in \partial f(p(x)). \quad (15)$$

In [10] it is also proved that for all $x_1, x_2 \in R^n$ the next formula:

$$\|p(x_1) - p(x_2)\|_M^2 \le \langle M(x_1 - x_2), p(x_1) - p(x_2) \rangle \quad (16)$$

is valid, namely the mapping $x \to p(x)$, where $p(x)$ is defined by (12), is Lipschitzian with constant $\frac{\Lambda}{\lambda}$ (see Proposition 2.3. in [10]).

Lemma 5: The following statements are equivalent:

(i) x minimizes f; (iv) x minimizes F;

(ii) $p(x) = x$ (v) $f(p(x)) = f(x)$;

(iii) $\nabla F(x) = 0$ (vi) $F(x) = f(x)$

Proof. See in [5] or [19].

IV. DINI SECOND UPPER DIRECTIONAL DERIVATIVE

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

Definition 6. [18] 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

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

for all $d \in R^n$.

Since the Moreau-Yoshida 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 $d \in R^n$, i.e.:

$$F_D''(x,d) = \limsup_{\alpha \downarrow 0} \frac{g_1 - g_2}{\alpha} d,$$
$$g_1 \in \partial f(p(x + \alpha d)), g_2 \in \partial f(p(x))$$

where $F(x)$ is defined by (11).

Lemma 6: Let $f : R^n \to R$ be a closed convex proper function and F is its Moreau –Yoshida regularization in the sense of definition 5. Then the next statements are valid.

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

(ii) $F_D''(x_k, d_1 + d_2) \le 2(F_D''(x_k, d_1) + F_D''(x_k, d_2))$

(iii) $|F_D''(x_k, d)| \le K \cdot \|d\|^2$, where K is some constant.

Proof. See in [18] and [2].

Lemma 7. Let $f : R^n \to R$ be a closed convex proper function and let F be its Moreau –Yoshida regularization. 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

$$\limsup_{i \to \infty} F_D''(x_i, d_i) \le F_D''(x,d)$$

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

Proof. See in [18] and [2].

V. SEMI-SMOOTH FUNCTIONS AND OPTIMALITY CONDITIONS

Definition 7: 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-Yoshida regularization is a semi-smooth function.

Lemma 8. *[18]:* 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 $h^T V h - F''(x,h) = o(\|h\|^2)$.

Lemma 9: Let $f : R^n \to R$ be a proper closed convex function and let F be its Moreau-Yoshida regularization. So, if $x \in R^n$ is solution of the problem (1) then $F'(x,d) = 0$ and $F_D''(x,d) \ge 0$ for all $d \in R^n$.

Proof. From the definition of the directional derivative and by Lemma 5 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 5, theorem 23.1 in [15] 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

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

Lemma 10. Let $f:R^n\to R$ be a proper closed convex function, F its Moreau-Yoshida 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).

Proof. Suppose that $x\in R^n$ is not a strict minimum of the function f. According to Lemma 5 that means that $x\in R^n$ is not a strict minimum 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 $x_k=x+t_kd$, where $t_k=\frac{\|x_k-x\|}{\|d\|}$ then by Lemma 8 and Lemma 6 it follows that

$$F(x_k)-F(x)-t_k\nabla F(x)^Td=\frac{1}{2}t_k^2F''_D(x,d)+o\left(\|d\|^2\right)$$

holds. Since $\nabla F(x)=0$ (from assumption of Lemma 10) it follows that $\frac{1}{2}t_k^2F''_D(x,d)\leq 0$, which contradicts the assumption.■

VI. 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 $f(x), F(x), \nabla F(x)$ and $F''_D(x,d)$ for a given $d\in R^n$.

At the k-th iteration we consider the following problem

$$\min_{d\in R^n}\Phi_k(d),\ \Phi_k(d)=\nabla F(x_k)^Td+\frac{1}{2}F''_D(x_k,d)\quad (17)$$

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 $x_{k+1}=x_k+\alpha_kd_k$, where the direction vector d_k is a solution of the problem (17), and the step-size α_k is a number satisfying $\alpha_k=q^{i(k)}, 0<q<1$, where $i(k)$ is the smallest integer from $\{0,1,2,...\}$ such that

$$F(x_k+q^{i(k)}d_k)-F(x_k)\leq -\frac{1}{2}q^{i(k)}\sigma(F''_D(x_k,d_k))\quad (18)$$

where $\sigma:[0,+\infty)\to[0,+\infty)$ is a continuous function satisfying $\delta_1t\leq\sigma(t)\leq\delta_2t$ and $0<\delta_1<\delta_2<1$.

We suppose that

$$c_1\|d\|^2\leq F''_D(x_k,d)\leq c_2\|d\|^2\quad (19)$$

hold for some c_1 and c_2 such that $0<c_1<c_2$.

Lemma 11. Under the assumption (19) the function $\Phi_k(\cdot)$ is coercive.

Proof. From the assumption there exists $K>0$ ($0<c_1\leq K\leq c_2$) such that $F''_D(x,d)=\max_{V\in\partial^2F(x)}d^TVd=K\|d\|^2$.

Since ∇F is locally Lipschitzian we have that $\left|\nabla F(x_k)^Td\right|\leq\Lambda\|d\|$ holds (see in [17]), therefore $\left|\nabla F(x_k)^Td+\frac{1}{2}\cdot 0\right|\leq\Lambda\|d\|$ holds and hence we have that: $\left|\nabla F(x_k)^Td+\frac{1}{2}F''_D(x,d)-\frac{1}{2}K\|d\|^2\right|\leq\Lambda\|d\|$. Hence, we have:

$$-\Lambda\|d\|+\frac{1}{2}K\|d\|^2\leq\nabla F(x_k)^Td+\frac{1}{2}F''_D(x,d)\leq\Lambda\|d\|+\frac{1}{2}K\|d\|^2$$

and $-\Lambda+\frac{1}{2}K\|d\|\leq\frac{\Phi_k(d)}{\|d\|}\leq\Lambda+\frac{1}{2}K\|d\|$.

This establish coercivity of the function Φ_k.■

Remark. Coercivity of the function Φ_k assures that the optimal solution of problem (17) exists (see in [18]). It also means that, under the assumption (19) the direction sequence $\{d_k\}$ is bounded sequence on R^n (proof is analogous to the proof in [18]).

Proposition 3. If the Moreau-Yoshida regularization $F(\cdot)$ of the proper closed convex function $f(\cdot)$ satisfies the condition (19), then:

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

(ii) the level set $L(x_0)=\{x\in R^n: F(x)\leq F(x_0)\}$ 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. (i) From the assumption (19) and the mean value theorem it follows that for all $x\in L(x_0)$ ($x\neq x_0$) there exists $\theta\in(0,1)$ such that:

$$F(x)-F(x_0)==\nabla F(x_0)^T(x-x_0)+\frac{1}{2}F''_D(x_0+\theta(x-x_0),x-x_0)$$

$$\geq\nabla F(x_0)^T(x-x_0)+\frac{1}{2}c_1\|x-x_0\|^2>\nabla F(x_0)^T(x-x_0)$$

that is, $F(\cdot)$ is uniformly and consequently strictly convex on $L(x_0)$.

(ii) From [16] it follows that the level set $L(x_0)$ is bounded. The set $L(x_0)$ is closed and convex because the function $F(\cdot)$ is continuous and convex. Therefore the set $L(x_0)$ is a compact convex set.

(iii) The existence of x^* follows from the continuity of the function $f(\cdot)$, and therefore and $F(\cdot)$, on the bounded set $L(x_0)$. From 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 $F(\cdot)$ is strictly convex it follows from [15] that x^* is a unique minimizer.■

Lemma 12. The following statements are equivalent:

(i) $d = 0$ is globally optimal solution of the problem (17)

(ii) 0 is the optimum of the objective function in (17)

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

Proof. (i) $\Rightarrow$ *(ii:)* is obvious

(ii) $\Rightarrow$ *(iii)*: Let 0 be a global optimum value of (17), then for any $\lambda > 0$ and $d \in R^n$:

$$0 = \Phi_k(0) \le \Phi_k(\lambda d) = \lambda \nabla F(x_k)^T d + \frac{1}{2} F_D''(x_k, \lambda d)$$

$$= \lambda \nabla F(x_k)^T d + \frac{1}{2} \lambda^2 F_D''(x_k, d)$$

where the last equality holds by Lemma 6. Hence dividing both sides by λ and letting $\lambda \downarrow 0$ we have that $\nabla F(x_k)^T d \ge 0$ holds for any $d \in R^n$, and consequently it follows that x_k is a stationary point of the function F, i.e. $\nabla F(x_k) = 0$. Hence, by Lemma 5 it follows that x_k is a minimum point of the function f.

(iii) $\Rightarrow$ *(i)*: Let x_k be a point such that $0 \in \partial f(x_k)$. Then by (14), Lemma 5 and Lemma 9 it follows that

$$\nabla F(x_k)^T d \ge 0 \qquad (20)$$

for any $d \in R^n$. Suppose that $d \ne 0$ is the optimal solution of the problem (17). From the property of the iterative function Φ it follows that:

$$\nabla F(x_k)^T d_k \le -\frac{1}{2} F_D''(x_k, d_k) \qquad (21)$$

Hence, (21) implies:

$$\nabla F(x_k)^T d_k < 0 \; . \qquad (22)$$

The above two inequalities (20) and (22) are contradictory.■

Now we shall prove that there exists a finite $i(k)$, i.e. since d_k is defined by (17), that the algorithm is well-defined.

Proposition 4. . If $d_k \ne 0$ is a solution of (17), then for any continuous function $\sigma : [0,+\infty) \to [0,+\infty)$ satisfying $\delta_1 t \le \sigma(t) \le \delta_2 t$ (where $0 < \delta_1 < \delta_2 < 1$) there exists a finite $i^*(k)$ such that for all $q^{i(k)} \in \left(0, q^{i^*(k)}\right)$

$$F\left(x_k + q^{i(k)} d_k\right) - F(x_k) \le -\frac{1}{2} q^{i(k)} \sigma\left(F_D''(x_k, d_k)\right)$$

holds.

Proof. According to Lemma 9.3 from [8] and from the definition of d_k it follows that for $x_{k+1} = x_k + t d_k, t > 0$ we have

$$F(x_{k+1}) - F(x_k) \le t \nabla F(x_k)^T d_k + \frac{\Lambda}{2} t^2 \|d_k\|^2 \le -t \frac{1}{2} F_D''(x_k, d_k) + \frac{\Lambda}{2} t^2 \|d_k\|^2 \qquad (23)$$

$$\le -t \frac{1}{2\delta_2} \sigma\left(F_D''(x_k, d_k)\right) + \frac{\Lambda}{2} t^2 \|d_k\|^2$$

If we choose $t = \dfrac{\sigma\left(F_D''(x_k, d_k)\right)}{\Lambda \|d_k\|^2}$ and put in (23), we get

$$F(x_{k+1}) - F(x_k) \le \frac{1}{2} \frac{\delta_2 - 1}{\delta_2} \frac{\sigma^2\left(F_D''(x_k, d_k)\right)}{\Lambda \|d_k\|^2} = -\frac{K}{2} t \sigma\left(F_D''(x_k, d_k)\right)$$

where $\dfrac{\delta_2 - 1}{\delta_2} = -K < 0$. Taking $q^{i^*(k)} = \dfrac{[Kt]}{q}$, i.e. $i^*(k) = \log_q \dfrac{[Kt]}{q}$ we have that the claim of the theorem holds for all $q^{i(k)} \in \left(0, q^{i^*(k)}\right)$.■

Convergence theorem. Suppose that f is a proper closed convex function and F is its Moreau-Yoshida regularization satisfies (19). Then for any initial point $x_0 \in R^n, x_k \to x_\infty$, as $k \to +\infty$, where x_∞ is a unique minimal point of the function f.

Proof. . If $d_k \ne 0$ is a solution of (17), it follows that $\Phi_k(d_k) \le 0 = \Phi_k(0)$. Consequently, we have by the condition (19) that

$$\nabla F(x_k)^T d_k \le -\frac{1}{2} F_D''(x_k, d_k) \le -\frac{1}{2} c_1 \|d_k\|^2 < 0$$

From the above inequality it follows that the vector d_k is a descent direction at x_k, i.e. from the relations (18) and (19) we get

$$F(x_{k+1}) - F(x_k) = F\left(x_k + q^{i(k)} d_k\right) - F(x_k) \le -\frac{1}{2} q^{i(k)} \sigma\left(F_D''(x_k, d_k)\right)$$

$$\le -\frac{1}{2} q^{i(k)} \delta_1 F_D''(x_k, d_k) \le -\frac{1}{2} q^{i(k)} c_1 \|d_k\|^2$$

for every $d_k \ne 0$. Hence the sequence $\{F(x_k)\}$ has the descent property, and, consequently, the sequence $\{x_k\} \subset L(x_0)$. Since $L(x_0)$ is by the Proposition 3 a compact convex set, it follows that the sequence $\{x_k\}$ is bounded. Therefore there exist accumulation points of the sequence $\{x_k\}$.

Since ∇F is continuous, then, if $\nabla F(x_k) \to 0, k \to +\infty$ it follows that every accumulation point x_∞ of the sequence $\{x_k\}$ satisfies $\nabla F(x_\infty) = 0$. Since F is (by the Proposition 3) strictly convex, there exists a unique point $x_\infty \in L(x_0)$ such that $\nabla F(x_\infty) = 0$. Hence, the sequence $\{x_k\}$ has a unique limit point x_∞ and it is a global minimizer of F and by Lemma 5 it is a global minimizer of the function f.

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

Let K_1 be a set of indices such that $\lim\limits_{k \in K_1} x_k = x_\infty$. Then 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. Because of the descent property, it follows that all points of accumulation have the same function value and

$$F(x_\infty) - F(x_0) = \sum_{k=0}^{+\infty} [F(x_{k+1}) - F(x_k)]$$

$$\leq \sum_{k \in K_1} [F(x_{k+1}) - F(x_k)]$$

$$\leq -\frac{1}{2} \sum_{k \in K_1} q^{i(k)} \sigma(F_D''(x_k, d_k))$$

$$\leq -\frac{1}{2} q^I \delta_1 \sum_{k \in K_1} F_D''(x_k, d_k) \leq -\frac{1}{2} q^I c_1 \sum_{k \in K_1} \|d_k\|^2$$

Since $F(x_\infty)$ is finite, it follows that $\|d_k\| \to 0., k \to \infty, k \in K_1$

By Lemma 12 it follows that $d_\infty = 0$ is a globally optimal point of the problem (17) and, that the corresponding accumulation point x_∞ is a stationary point of the objective function F, i.e. $\nabla F(x_\infty) = 0$. From Lemma 5 and Lemma 12 it follows that x_∞ is a unique optimal point of the function f.

b) There is a subset $K_2 \subset K_1$ such that $\lim\limits_{k \to \infty} i(k) = +\infty$. By definition of $i(k)$, we have for $k \in K_2$ that

$$F(x_k + q^{i(k)-1} d_k) - F(x_k) > -\frac{1}{2} q^{i(k)-1} \sigma(F_D''(x_k, d_k)) \quad (24).$$

Suppose that x_∞ is an arbitrary accumulation point of $\{x_k\}$, but not a stationary point of F *, i.e.* f. Then, from Lemma 12 it follows that the corresponding direction vector $d \neq 0$. Now, dividing both sides in the expression (24) by $q^{i(k)-1}$ and using $\lim\limits_{k \to \infty} q^{i(k)-1} = 0, k \in K_2$, we get

$$\nabla F(x_\infty)^T d_\infty > -\frac{1}{2} \sigma(F_D''(x_\infty, d_\infty)) > -\frac{1}{2} \delta_2 F_D''(x_\infty, d_\infty) > -\frac{1}{2} F_D''(x_\infty, d_\infty)$$

But, from the property of the iterative function in (17), we have $\nabla F(x_\infty)^T d_\infty \leq -\frac{1}{2} F_D''(x_\infty, d_\infty)$. Therefore, we get a contradiction.■

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.

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

$$for \quad n = 1,2,3,...$$

where $\mu_0 = F(x_0) - F(x_\infty)$ and $diam L(x_0) = \eta < +\infty$ *(*since by Proposition 3 it follows that $L(x_0)$ is bounded).

<u>*Proof*</u>. The proof directly follows from the Theorem 9.2, page 167, in [8].

REFERENCES

[1] Browien J., Lewis A.: "*Convex analysis and nonlinear optimization"*, Canada 1999
[2] Clarke F. H.: "Optimization and Nonsmooth Analysis", Wiley, New York, 1983
[3] Djuranovic Milicic Nada: "*On An Optimization Algorithm For LC^1 Unconstrained Optimization",* FACTA UNIVERSITATIS Ser. Math. Inform. 20 , pp 129-135, 2005
[4] Демьянов В. Ф., Васильев: „*Недифференцируемая оптимизация*“, Москва „Наука“, 1981.
[5] Fuduli A.: "Metodi Numerici per la Minimizzazione di Funzioni Convesse Nondifferenziabili", PhD Thesis, Departimento di Electronica Informatica e Sistemistica, Univerzita della Calabria (1998)
[6] Fletcher R.: "Practical Methods of Optimization, Volume 2, Constrained Optimization", Department of Mathematics, University of Dundee, Scotland, U.K,Wiley-Interscience, Publication, (1981)
[7] *Иоффе А. Д., Тихомиров В. М.: „Теория экстремальныџ задач“, Москва „Наука“, 1974*
[8] Karmanov V.G: "*Mathematical Programming"*. Nauka, Moscow, 1975 (in Russian).
[9] Кусарев А. Г., Кутателадзе С. С.:*Субдифференциальное исчисление*, Новосибирск, Издатеьство „Наука“, (1987)
[10] Lemare´chal C. , Sagastiza C.: "*Practical aspects of the Moreau-Yoshida regularization I: theoretical preliminaries"*, SIAM Journal of Optimization, Vol.7, pp 367-385, 1997
[11] Meng F., Zhao G.:"*On Second-order Properties of the Moreau-Yoshida Regularization for Constrained Nonsmooth Convex Programs",* Numer. Funct. Anal. Optim , VOL 25; NUMB 5/6, pages 515-530, 2004
[12] Пшеничнй Б. Н.: „*Выпуклый анализ и экстремальные задачи*“, Москва „Наука“, 1980.
[13] Пшеничнй Б. Н.: Неопходим*ые условя экстремума*, Москва „Наука“, 1982.
[14] Пшеничнй Б. Н.: Метод линеаризации, Москва „Наука“, 1983.
[15] Rockafellar R.: "*Convex Analysis"*, Princeton University Press, New Jersey (1970) (Russ. trans.: Mir, Moscow 1973)
[16] Polak E.: "*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].
[17] Pytlak R.:"*Conjugate Gradient Algorithms in Nonconvex Optimization"*, Library of Congress Control Number: 2008937497, Springer-Verlag Berlin Heidelberg, 2009
[18] 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
[19] Qi L.: "*Second Order Analysis of the Moreau-Yoshida Regularization"*, School of Mathematics, The University of New Wales Sydney, New South Wales 2052, Australia, 1998.