FACTA UNIVERSITATIS (NIŠ)
Ser. Math. Inform. Vol. 30, No 5 (2015), 583–596

MINIMIZATION OF A NONDIFFERENTIABLE FUNCTION: AN ALGORITHM USING TRUST REGION TECHNIQUE AND CONJUGATE SUBGRADIENT METHOD

Milanka Gardašević-Filipović and Nada Djuranović-Miličić

Abstract. In this paper we present a method for minimization of a nondifferentiable function. The method uses trust region strategy combined with a conjugate subgradient method. 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.

Keywords: non-smooth convex optimization, trust region method, conjugate subgradient method, bundle methods.

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.

Nonsmooth 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 [9], [15], [27], [28]); (ii) bundle techniques (see [14], [16], [18], [20], [25]), (iii) Moreau-Yosida regularization (see [5], [6], [7], [8], [13], [19], [22], [23]). A good overview can be found in [1] and [21].

The algorithm we are going to present here combines the trust region method with the conjugate subgradient method. We build our iterative method on the

Received September 30, 2015; Accepted November 03, 2015
2010 *Mathematics Subject Classification*. Primary 65k05; Secondary 90c30,90c25

following idea. If we, on the k -th iteration, approximate the objective function f of the problem (1.1) by a function Φ then it is reasonable to assume that it could be possible to define some neighborhood about x_k where the approximation function Φ of the objective function f agrees with the objective function in some sense. Then it would be appropriate to choose for the next iteration the minimizer of the approximation function Φ, that is to apply a trust region technique. If the approximation function Φ of the objective function f does not agree with the objective function in an appropriate degree then we proceed to solve problem (1.1) by conjugate subgradient method (instead of bundle philosophy which is used in [12]) for solving the latter problem.

In the second section some basic theoretical preliminaries are given. In the third section a model algorithm is suggested. In the fourth section 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 $(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, ..., \frac{\partial}{\partial x_n})^T$. The Euclidean product is denoted by $\langle .,. \rangle$ and $\|.\|$ is the associated norm.

The *domain* of a given function $f : R^n \to R \cup \{+\infty\}$ is the set $dom(f) = \{x \in R^n | f(x) < +\infty\}$. 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 $f : R^n \to R \cup \{+\infty\}$. 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 inequality $f(z) \le f(x) + g^T(z - x)$ 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 $x \in dom(f)$ and the set $\partial f(x)$ is bounded for $\forall x \in B \subset dom(f)$, where B is compact. If a proper convex function $f : R^n \to R \cup \{+\infty\}$ is a differentiable function at a point $x \in dom(f)$, then $\partial f(x) = \{\nabla f(x)\}$.

Proposition 2.1. *Let $f : R^n \to R \cup \{+\infty\}$ be a proper convex function. The condition $0 \in \partial f(x)$ is a first order necessary and sufficient condition for a global minimizer at $x \in R^n$. This can be stated alternatively as:*

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

Proof. See in [10] , [17] or [24] □

In [10] it is proved that for the sequence $x_k \to x'$, defined by $x_k = x' + \epsilon_k s_k$, such that $\epsilon_k > 0, \epsilon_k \to 0$ and $s_k \to s$ if $g_k \in \partial f(x_k)$ then all accumulation points of the sequence $\{g_k\}$ lie in the set $\partial f(x')$, where $f : S \to R \cup \{+\infty\}$ is a convex function defined on a convex set $S \subseteq R^n$.

The *directional derivative* of a real function defined on R^n at the point $x' \in R^n$ in the direction $s \in R^n$, denoted by $f'(x', s)$, is $f'(x', s) = \lim_{t \downarrow 0} \frac{f'(x'+ts)-f(x')}{t}$ 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 [21]). In [10] it is proved that if the sequence $x_k \to x'$, defined by $x_k = x' + \epsilon_k s_k$, such that $\epsilon_k > 0$, $\epsilon_k \to 0$ and $s_k \to s$, then it holds that $f'(x', s) = \lim_{k \to +\infty} \frac{f'(x_k)-f(x')}{\epsilon_k} = \max_{g \in \partial f(x')} s^T g$, where $f : S \to R \cup \{+\infty\}$ is a convex function defined on a convex set $S \subseteq R^n$. Hence, it follows that if the function f is convex then $f(x' + ts) = f(x') + t f'(x', s) + o(t)$ holds, which can be considered as one linearization of the function f (see in [11]). It also follows that the first order necessary and sufficient condition for a global minimizer of the convex function f at the point $x \in R^n$ can be stated alternatively as $\max_{g \in \partial f(x)} s^T g \geq 0, \forall s \in R^n, \|s\| = 1$ (see in [2]).

Lemma 2.1. *Let $f_i : R^n \to R \cup \{+\infty\}$ for $i \in I = \{1, 2, ..., n\}$, $n \in N$ be convex functions, and $f(x) = \max_{i \in \{1,2,...,n\}} f_i(x)$. Then the function f is a convex function, and its subdifferential at the point $x \in R^n$, is given by $\partial f(x) = \left\{ \sum_{i \in \hat{I}} \lambda_i g_i \;\middle|\; \sum_{i \in \hat{I}} \lambda_i = 1, \lambda_i \geq 0, g_i \in \partial f_i(x), \text{ for } i \in \hat{I} \right\}$, where $\hat{I}$ is the set $\hat{I} = \{i \in I \mid f(x) = f_i(x)\}$.*

Proof. See in [24] □

3. A MODEL ALGORITHM

We suppose that at the k-th iteration, there is an index set $I_k = \{1, 2, ..., k\}$, and we store information as a bundle $B_k = \{(x_i, f(x_i), g_i) \mid i \in I_k\}$, 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 follows:

$$f_i(x) = f(x_i) + g_i^T (x - x_i), \text{ where } i \in I_k. \tag{3.1}$$

If f is a convex function then $f(x) = \max_{z \in R^n} \{f(z) + g^T (x - z)\}$ holds (proved in [4]), where $g \in \partial f(z)$. Hence, it is reasonable to consider the next function

$$\hat{f}_k(x) = \max_{0 \leq i \leq k} f_i(x) = \max_{0 \leq i \leq k} \{f(x_i) + g_i^T (x - x_i)\} \tag{3.2}$$

which is known as a cutting plane function [1].

It is easy to see that $f(x) \geq \hat{f}_{k+1}(x) \geq \hat{f}_k(x)$ for all $x \in R^n$.

Now, let us consider the next function:

$$\Phi_k(x) = \hat{f}_k(x) + \frac{1}{2}\|x_k - x\|^2 \tag{3.3}$$

Clearly, we have that $\Phi_k(x_k) = \hat{f}_k(x_k)$ holds. 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 two type of functions - the maximum function and the linear functions, i.e. $\max_{0\le i\le k} f_i(x)$. Then the function defined by (3.3) is a sum of one differentiable convex quadratic function and maximum of the linear functions. So, the function defined by (3.3) is a closed convex nonsmooth function.

If $z \in \partial\Phi_k(x)$ then $z \in \partial\hat{f}_k(x) + \partial(\frac{1}{2}\|x_k - x\|^2)$. Hence, $z \in \partial\hat{f}_k(x) + x - x_k$ and $z = \hat{g} + x - x_k$ for some $\hat{g} \in \partial\hat{f}_k(x)$. Since $\hat{f}_k(x)$ is a polyhedral function we have, according to Lemma 2.1, that $\hat{g} = \sum_{i\in\hat{I}_k}\lambda_i g_i$, where $\hat{I} = \{i \in I_k | \hat{f}_k(x) = f_i(x)\}$ and $g_i \in \partial f(x_i), i \in \hat{I}_k, \sum_{i\in\hat{I}_k}\lambda_i = 1, \lambda_i \ge 0$, i.e. $\hat{g}$ is a convex combination of the subgradients from the bundle B_k.

The algorithm we are going to present constructs a sequence $\{x_k\}$ in R^n in the following way. On the k-th iteration, we consider the function defined by (3.3) as an approximation of the objective function f about the point x_k i.e. $\Phi_k(x) = \hat{f}_k(x) + \frac{1}{2}\|x_k - x\|^2$. For the next iteration it would be appropriate to choose $x_{k+1} = x_k + \delta_k$, where the correction δ_k minimizes $\Phi_k(x)$ for all $x = x_k + \delta_k \in \Omega_k$, where Ω_k denotes the trust region. Namely, Ω_k denotes the neighborhood about the point x_k where the function defined by (3.3) approximates the objective function of the problem (1.1). Hence, we have that the function defined by (3.3) can be written as:

$$\Phi_k(x_k+\delta) := \hat{f}_k(x_k+\delta) + \frac{1}{2}\|\delta\|^2 = \tag{3.4}$$

$$\max_{1\le i\le k}\{f(x_i) + g(x_i)^T(x_k + \delta - x_i)\} + \frac{1}{2}\|\delta\|^2.$$

It is convenient to consider the case $\Omega_k = B(x_k, h_k)$ for some positive h_k, i.e. δ_k is a solution of the subproblem

$$\min_{\delta}\Phi_k(x_k+\delta) \quad \text{subject to } \|\delta\| \le h_k. \tag{3.5}$$

If $\Phi_k(x_k + \delta)$ attains its minimum at δ_k, then we have that $0 \in \partial\Phi_k(x_k + \delta_k)$, i.e. $0 \in \partial\hat{f}(x_k + \delta_k) + \delta_k$.
Since δ_k has to satisfy the condition $\|\delta_k\| \le h_k$, we consider the following problem.

For a given $g_i \in \partial f(x_i), i \in \hat{I}_k$ compute λ_i as a solution of the next problem:

$$\max \|\sum_{i\in\hat{I}_k} \lambda_i g_i\| \text{ such that } \textstyle\sum_{i\in\hat{I}_k} \lambda_i = 1, \lambda_i \geq 0. \tag{3.6}$$

We can choose the radius h_k as a maximum value of the objective function in the problem (3.6), i.e. such that the next equality:

$$h_k = \max_{\sum_{i\in\hat{I}_k} \lambda_i = 1, \lambda_i \geq 0} \|\sum_{i\in\hat{I}_k} \lambda_i g_i\| \tag{3.7}$$

holds. Under the assumption (3.7) the subproblem (3.5) has a solution. Namely, the next theorem holds.

Theorem 3.1. *If the point x_k is given and h_k is chosen such that the equality (3.7) holds, then the subproblem (3.5) has the unique solution. More than that, this solution is the solution of the problem* $\min_{\delta\in R^n} \Phi_k(x_k + \delta)$.

Proof. At the given point x_k the objective function for the subproblem (3.5) has a subgradient $\hat{g}(x_k + \delta) + \delta \in \partial\Phi(x_k + \delta)$, where $\hat{g}(x_k + \delta) := \hat{g} = \sum_{i\in\hat{I}_k} \lambda_i g_i$ for $g_i \in \partial f(x_i)$, $i \in \hat{I}_k$ and $\sum_{i\in\hat{I}_k} \lambda_i = 1, \lambda_i \geq 0$, and

$$\hat{I}_k = \left\{i \in I_k \;\middle|\; \hat{f}_k(x_k + \delta) = f_i(x_k + \delta) = f(x_i) + g_i^T(x_k + \delta - x_i)\right\}.$$

If the point $x_k + \delta_k$ is a minimum point of $\Phi_k(x_k + \delta)$ on R^n then by Proposition 2.1 it follows that $0 \in \partial\Phi_k(x_k + \delta_k)$ and hence $\delta_k = -\hat{g}(x_k + \delta_k) = -\sum_{i\in\hat{I}_k} \lambda_i g_i$, for $g_i \in \partial f(x_i), i \in \hat{I}_k$ and $\sum_{i\in\hat{I}_k} \lambda_i = 1, \lambda_i \geq 0$, where $\hat{I}_k = \{i \in I_k | \hat{f}_k(x_k + \delta_k) = f_i(x_k + \delta_k)\}$.

Because of the equality (3.7), we have that

$$h_k = \max_{\sum_{i\in\hat{I}_k} \lambda_i = 1, \lambda_i \geq 0} \|\sum_{i\in\hat{I}_k} \lambda_i g_i\| \geq \|\sum_{i\in\hat{I}_k} \lambda_i g_i\| = \|\hat{g}\| = \| - \delta_k\|$$

hold. So, it follows that $\delta_k = -\hat{g}$ is a solution of the subproblem (3.5). The solution is unique since Φ_k is a strongly convex function (as a sum of maximum of linear functions and strongly convex quadratic function) □

Corollary 3.1. *If the point x_k is given and h_k is chosen such that*

$$h_k = \max_{\sum_{i\in I_k} \lambda_i = 1, \lambda_i \geq 0} \|\sum_{i\in I_k} \lambda_i g_i\|$$

holds (we use the whole index set I_k instead of the $\hat{I}_k$ used in (3.7)), then the subproblem (3.5) has the unique solution. More than that, this solution is the solution of the problem $\min_{\delta\in R^n} \Phi_k(x_k + \delta)$.

Proof. If we choose h_k such that the condition (3.7) holds, then, by Theorem 3.1, the solution of the problem (3.5) there exists. More precisely, there exists the solution of the problem (3.5) over the set

$$S_0 = \left\{\delta \;\middle|\; \|\delta\| \le h_k, h_k = \max_{\sum_{i\in\hat{I}_k}\lambda_i=1,\lambda_i\ge 0} \|\sum_{i\in\hat{I}_k}\lambda_i g_i\| \;\right\}.$$

If we choose $h_k = \max_{\sum_{i\in I_k}\lambda_i=1,\lambda_i\ge 0}\|\sum_{i\in I_k}\lambda_i g_i\|$, i.e. if we use the whole index set I_k instead of the $\hat{I}_k$ used in (3.7), then we have to solve the problem (3.5) on the larger set

$$S = \left\{\delta \;\middle|\; \|\delta\| \le h_k, h_k = \max_{\sum_{i\in I_k}\lambda_i=1,\lambda_i\ge 0} \|\sum_{i\in I_k}\lambda_i g_i\| \;\right\}$$

such that $S_0 \subseteq S$ (since $\hat{I}_k \subseteq I_k$). So, if a solution there exists on the subset S_0, then a solution there exists on the set S, where $S_0 \subseteq S$.

The solution is unique since Φ_k is a strongly convex function (as a sum of maximum of linear functions and strongly convex quadratic function) □

On the k-th iteration we denote by $\triangle f_k = f(x_k) - f(x_k + \delta_k)$ the actual reduction of the function $f(.)$, and by $\triangle \Phi_k = f(x_k) - \Phi_k(x_k + \delta_k)$ the predicted reduction. Hence, the ratio $r_k = \frac{\triangle f_k}{\triangle \Phi_k}$ measures the accuracy to which $\Phi_k(x_k + \delta)$ approximates $f(x_k + \delta)$. The ratio r_k plays an important role in selecting a new iteration point x_{k+1} and updating the trust-region radius h_k. If the ratio is close to 1, it means that there is good agreement. If the ratio is close to zero or negative then we quit the trust region and find the next iteration by a conjugate subgradient method.

We will present the algorithm now.

Algorithm

- **Step 0:** Let β and η be constants such that $0.5 < \beta < 1$ and $0 < \eta < 1$. Let ϵ, γ and μ be real positive numbers small enough. Suppose that we have a given initial point $x_1 \in R^n$. For a given $x_1 \in R^n$ calculate $g_1 = g(x_1) \in \partial f(x_1)$. Set $k = 1$ and $I_0 = \emptyset, B_0 = \emptyset$.

- **Step 1:** For a given x_k calculate $f_k = f(x_k)$.
  Set $I_k = \{k\} \cup I_{k-1} \backslash S_k$, where $S_k = \{i \in I_{k-1} \;\big|\; \|x_i - x_k\| \ge \mu\}$.
  Set $B_k = \{(x_i, f(x_i), g_i) \;\big|\; i \in I_k\}$.
  Solve the problem (3.6) with I_k instead of $\hat{I}_k$ and denote by h_k its solution.

- **Step 2:** Solve the problem (3.5) $\min_{\delta\in R^n} \Phi_k(x_k+\delta)$ and denote by δ_k its solution. Denote by

$$\hat{I}_k = \{\, i \in I_k \;\big|\; \hat{f}_k(x_k+\delta_k) = f_i(x_k+\delta_k) = f(x_i) + g_i^T(x_k+\delta_k-x_i)\,\}$$

- **Step 3:** If $\|\delta_k\| < \epsilon$ then stop. Otherwise, solve the problem $\min \|\sum_{i\in\hat{I}_k} \lambda_i g_i\|$ such that $\sum_{i\in\hat{I}_k} \lambda_i = 1, \lambda_i \geq 0$, and denote by $\lambda_i^{(k)}$ its solution. If $\|\lambda_i^{(k)} g_i\| \leq \epsilon$, then stop. Otherwise go to Step 4.

- **Step 4:** Calculate $f(x_k+\delta_k), \phi(x_k+\delta_k), \triangle f_k, \triangle \phi_k, r_k$.

- **Step 5:** If $r_k < \beta$ then go to Step 6.
  Otherwise set $x_{k+1} = x_k + \delta_k$ and calculate

$$g_{k+1} = g(x_{k+1}) \in \partial f(x_{k+1}).$$

  Set $k = k+1$ and go to Step 1.

- **Step 6:** Set $\bar{\delta}_{(k)} = -\sum_{i\in\hat{I}_k} \lambda_i^{(k)} g_i$. Compute $\alpha_k > \gamma > 0$ such that $\alpha_k\|\bar{\delta}_k\| \leq \epsilon\eta\mu$ holds.
  Compute $g_{k+1} \in \partial f(x_k + \alpha_k\bar{\delta}_k)$.

- **Step 7:** If the inequality $g_{k+1}^T\bar{\delta}_k \leq -\frac{\eta}{2}\|\bar{\delta}_k\|^2$ holds then set $x_{k+1} = x_k + \alpha_k\bar{\delta}_k$, else set $x_{k+1} = x_k$.
  Set $k = k+1$ and go to Step 1.

Remark 1. Now we see in detail how to implement Step 2 of Algorithm. The problem $\min_{\delta\in R^n} \Phi_k(x_k+\delta)$ to be solved at Step 2, by Corollary 3.1, can be written in the following way:

$$\min_{\delta\in R^n, v\in R} \{v + \frac{1}{2}\|\delta\|^2\} \text{subject to } f(x_i) + g(x_i)^T(x_k+\delta-x_i) \leq v, i \in I_k.$$

The above problem has a nonlinear structure and a nonempty feasible region, and by duality solving the above problem is equivalent to finding multipliers λ_i^k for $i \in I_k$ that solve the quadratic problem

$$\min_{\delta\in R^n, v\in R} \{\frac{1}{2}\|\sum_{i\in I_k} \lambda_i g(x_i)\|^2 + \sum_{i\in I_k} \lambda_i[f(x_i) + g(x_i)^T(x_i - x_k)]\}$$

subject to $\sum_{i\in I_k}\lambda_i = 1, \lambda_i \geq 0$ for $i \in I_k$. If the multipliers λ_i^k for $i \in I_k$ solve the above problem then the solution of the problem $\min_{\delta\in R^n}\Phi_k(x_k+\delta)$ is $\delta_k = -\sum_{i\in I_k}\lambda_i^k g(x_i)$. The set $\hat{I}_k$ is the set of indices $i \in I_k$ such that $\lambda_i^k \geq 0$.

Remark 2. It is possible to consider $e_i^k := f(x_k) - f_i(x_k) = f(x_k) - f(x_i) - g_i^T(x_k - x_i)$ named linearization error, and rewrite the polyhedral function in the form $\hat{f}_k(x) = \max_{i\in I_k} f_i(x) = f(x_k) + \min_{i\in I_k} e_i^k$. In that case the bundle consists of $\left(g_j, e_j^k\right)$ instead of $(x_j, f(x_j), g_j)$, and that can be a way to cope with a finite storage.

Remark 3. Note that the passing to the next iteration is faster at Step 5 (the case $r_k \geq \beta$, i.e. when the quadratic approximation (defined by (3.4)) of the objective function is in the trust region) than at Step 7. Namely, in the first case ($r_k \geq \beta$) we made a step in the direction $\delta_k = -\sum_{i\in \hat{I}_k}\lambda_i g_i$ such that $\sum_{i\in\hat{I}_k}\lambda_i = 1, \lambda_i \geq 0$. In the second case $r_k < \beta$ at Step 7, when $g_{k+1}^T\bar{\delta}_k \leq -\frac{\eta}{2}\|\bar{\delta}_k\|^2$ holds, we made a step in the direction $\bar{\delta}_k = -\sum_{i\in\hat{I}_k}\lambda_i^k g_i$, where λ_i^k is a solution of the problem $\min\|\sum_{i\in\hat{I}_k}\lambda_i g_i\|$ such that $\sum_{i\in\hat{I}_k}\lambda_i = 1, \lambda_i \geq 0$, with a step lenght $\alpha_k \in (0, \eta\mu)$ or we do not move at all, when $g_{k+1}^T\bar{\delta}_k \leq -\frac{\eta}{2}\|\bar{\delta}_k\|^2$ does not hold . The moving at the Step 7 is in the spirit of the conjugate subgradients method [29], so this Algorithm has a faster part (case $r_k \geq \beta$) than the conjugate subgradients algorithm (ref. [26], page 617).

4. CONVERGENCE PROOF

Lemma 4.1. *Let $\{x_k\}$ be a sequence generated by the Algorithm such that $x_k \to x'$, $x_k = x' + \alpha_k s_k, s_k \to s, \alpha_k \downarrow 0$, where $\|s_k\| = 1$. Then $\hat{f}_k(x_k) - \hat{f}_k(x') = o(\alpha_k)$ holds. More than that then $\lim_{k\to+\infty}\hat{f}_k(x_k) = f(x')$ holds.*

Proof. We denote by $f_k = f(x_k), f' = f(x'), g_k = g(x_k) \in \partial f(x_k)$ and $g' = g(x') \in \partial f'$.
Since Φ_k is a convex function we have that $\Phi_k(x'+\alpha_k s_k) = \Phi_k(x') + \alpha_k\Phi_{k'}(x', s_k) + o(\alpha_k)$ holds. Hence, it follows that:

$$\Phi_k(x'+\alpha_k s_k) = \Phi_k(x') + \alpha_k \max_{g\in\partial\Phi_k(x')} g^T s_k + o(\alpha_k) \tag{4.1}$$

If $z \in \partial\Phi_k(x')$ such that $z^T s_k = \max_{g\in\partial\Phi_k(x')} g^T s_k$ then $z = \hat{g} + x' - x_k$ and $\hat{g} = \sum_{i\in\hat{I}_k}\lambda_i g_i$, where $\hat{I}_k = \left\{i \in I_k \;\middle|\; \hat{f}_k(x) = f_i(x)\right\}$ and $g_i \in \partial f(x_i)$, $i \in \hat{I}_k$, $\sum_{i\in\hat{I}_k}\lambda_i = 1$, $\lambda_i \geq 0$. Hence, from (4.1) it follows that:

$$\hat{f}_k(x_k) = \Phi_k(x_k) = \Phi(x'+\alpha_k s_k) = \Phi_k(x') + \alpha_k z^T s_k + o(\alpha_k) = \Phi_k(x') + z^T(x_k - x') + o(\alpha_k)$$

$$= \max_{0\leq i\leq k}\left\{f(x_i) + g(x_i)^T(x'-x_i)\right\} + \frac{1}{2}\|x_k - x'\|^2 + (\hat{g} + x' - x_k)^T(x_k - x') + o(\alpha_k)$$

$$= f(x_i) + g(x_i)^T(x'-x_i) + \frac{1}{2}\|\alpha_k s_k\|^2 + \Big(\sum_{i\in\hat{I}_k}\lambda_i g_i - \alpha_k s_k\Big)^T(\alpha_k s_k) + o(\alpha_k), \quad \text{for some } i \in \hat{I}_k$$

$$= f(x_i) + g(x_i)^T(x' - x_i) - \frac{1}{2}\|\alpha_k s_k\|^2 + \Big(\sum_{i\in\hat{I}_k}\lambda_i g_i\Big)^T(\alpha_k s_k) + o(\alpha_k), \quad i \in \hat{I}_k$$

$$= \hat{f}_k(x') - \frac{1}{2}\alpha_k^2 + \Big(\sum_{i\in\hat{I}_k}\lambda_i g_i\Big)^T(\alpha_k s_k) + o(\alpha_k) = \hat{f}_k(x') + o(\alpha_k).$$

So, we get that $\hat{f}_k(x_k) - \hat{f}_k(x') = o(\alpha_k)$ holds.

As $k \to +\infty$ we have that $\lim_{k\to+\infty} \hat{f}_k(x_k) = \lim_{k\to+\infty}\left\{\max_{1\le i\le k}\left(f_i + g_i^T(x_k - x_i)\right)\right\}$.

Since $f(x) \ge \hat{f}_{k+1}(x) \ge \hat{f}_k(x)$ for any $x \in R^n$ it follows that the sequence $\{\hat{f}_k(x)\}$ is an increasing and bounded above. The sequence $x_k \to x'$ and starting from k large enough infinitely many points of this sequence lies in the neighbourhood of the point x'. Hence, according to the fact that $f(x') = \max_{z\in R^n}\left\{f(z) + g^T(x' - z)\right\}$ where $g \in \partial f(z)$, we can state that $f(x') = \lim_{k\to+\infty}\left\{\max_{1\le i\le k}\left(f_i + g_i^T(x_k - x_i)\right)\right\}$.

If we suppose contrary, that is $f(x') > \lim_{k\to+\infty}\left\{\max_{1\le i\le k}\left(f_i + g_i^T(x_k - x_i)\right)\right\}$, then we have that

$$f(x') > \max_{1\le i\le k}\left(f_i + g_i^T(x_k - x_i)\right) \ge f_k + g_k^T(x_k - x_k) = f_k$$

holds for every k. Hence, as $k \to +\infty$ we get $f(x') > f(x')$ (because of convexity of the function f) $\square$

Lemma 4.2. *Let $\{x_k\}$ be a sequence generated by the Algorithm such that $x_k \to x'$, $x_k = x' + \alpha_k s_k$, $s_k \to s, \alpha_k \downarrow 0$, where $\|s_k\| = 1$. Then $\lim_{k\to+\infty} \phi_k(x_k) = f(x')$.*

Proof. As $k \to +\infty$ by Lemma 4.1 and $\Phi_k(x) = \hat{f}_k(x) + \frac{1}{2}\|x_k - x\|^2$ we have that $lim_{k\to+\infty}\Phi_k(x_k) = lim_{k\to+\infty}\hat{f}_k(x_k) = f(x')$ holds $\square$

If we denote by $\Phi_\infty(x) = \hat{f}_\infty(x) + \frac{1}{2}\|x' - x\|^2$ then we have $\Phi_\infty(x') = \hat{f}_\infty(x') = f(x')$.

Now we can prove the main result of this paper.

Theorem 4.1. *Let $\{x_k\}$ be a sequence generated by the Algorithm such that $x_k \in B \subset R^n$, $\forall k$ where B is compact. Then there exists an accumulation point x_∞ of the sequence $\{x_k\}$ satisfying the first order necessary and sufficient conditions, that is the following*

$$0 \in \partial f_\infty \tag{4.2}$$

holds (where $\partial f_\infty = \partial f(x_\infty)$).

Proof. Since the set B is compact and $x_k \in B$, it follows that the sequence $\{x_k\}$ has an accumulation point x_∞. Hence it follows that there exists a convergent subsequence $x_k \to x_\infty$, $k \in K$. The Algorithm generates a subsequence for which either:

(i) $r_k < \beta$ and hence $\|\bar{\delta}_k\| \to 0, \ \ k \in K_1 \subseteq K$ or
(ii) $r_k \geq \beta$ and $\inf(h_k) > 0, \ \ k \in K_2 \subseteq K$
is fulfilled. We will prove that (4.2) holds in either case (i) or (ii). In the case (i) we have that $r_k < \beta, \ \ k \in K_1$, and at Step 6 we calculate $g_{k+1} = g(x_{k+1}) \in \partial f(x_k + \alpha_k \bar{\delta}_k)$, where $\alpha_k > \gamma > 0$ is such that $\alpha_k \|\bar{\delta}_k\| \leq \eta\mu$ holds. Since g_{k+1} is a subgradient of the function f we have that:

$$f(z) \geq f(x_k + \alpha_k \bar{\delta}_k) + g_{k+1}^T (z - x_k - \alpha_k \bar{\delta}_k) \tag{4.3}$$

holds for every $z \in R^n$. If the inequality

$$g_{k+1}^T \bar{\delta}_k \leq -\frac{\eta}{2} \|\bar{\delta}_k\|^2 \tag{4.4}$$

holds at Step 7, then from (4.3) we have (for $z = x_k$):

$$\begin{aligned} f(x_k) \geq f(x_k + \alpha_k \bar{\delta}_k) + \alpha_k(-g_{k+1}^T \bar{\delta}_k) &\geq f(x_k + \alpha_k \bar{\delta}_k) + \frac{\alpha_k \eta}{2} \|\bar{\delta}_k\|^2 \\ &> f(x_k + \alpha_k \bar{\delta}_k), \end{aligned} \tag{4.5}$$

i.e. $f(x_k) > f(x_{k+1})$. Since the sequence $\{f(x_k)\}, \ \ k \in K_1$, is decreasing and bounded below on the compact B it follows that $f(x_{k+1}) - f(x_k) \to 0$ as $k \to +\infty, \ \ k \in K_1$. From (4.5) it follows that

$$f(x_k) - f(x_{k+1}) \geq \frac{\alpha_k \eta}{2} \|\bar{\delta}_k\|^2 > \frac{\gamma \eta}{2} \|\bar{\delta}_k\|^2 \tag{4.6}$$

holds, and since $f(x_{k+1}) - f(x_k) \to 0$ as $k \to +\infty, \ \ k \in K_1$, then from (4.6) it follows that $\|\bar{\delta}_k\| \to 0$ as $k \to +\infty, \ \ k \in K_1$. Since ∂f_k is bounded in a neighborhood of x_∞ there exists a subsequence for which $g_k \to g_\infty, \ \ k \in K_3 \subseteq K_1$, and $g_\infty \in \partial f_\infty$. At Step 6 we have that $\|\bar{\delta}_k\| > \epsilon$ and $\alpha_k \|\bar{\delta}_k\| \leq \epsilon\eta\mu$ hold, consequently we have that $0 < \epsilon\alpha_k < \alpha_k \|\bar{\delta}_k\| \leq \epsilon\eta\mu$, and hence it follows that $0 < \epsilon\alpha_k \leq \epsilon\eta\mu$ hold. Dividing the last inequalities by ϵ we get $0 < \alpha_k \leq \eta\mu$. Since $\|x_{k+1} - x_k\| = \alpha_k \|\bar{\delta}_k\| \to 0$ as $k \to +\infty, \ \ k \in K_1$, over the subsequence (because of the boundness of $\{\alpha_k\}$) then according to the Step 1 of the Algorithm and Lemma 2.1 it follows that $\|x_i - x_k\| \to 0, \ \ i \in \hat{I}_k, \ \ k \in K_1$. It also follows that since ∂f_i for $i \in \hat{I}_k$ are bounded in a neighborhood of x_∞ that there are corresponding subsequences for which $g_i \to g_\infty^i$ for $i \in \hat{I}_k \ \ k \in K_3$, and $g_\infty^i \in \partial f_\infty$. Since ∂f_∞ is a convex set it follows that every convex combination of points from the set ∂f_∞ belongs to ∂f_∞, i.e. $\bar{\delta}_\infty = \sum_{i \in \hat{I}_\infty} \lambda_i^{(\infty)} g_i^\infty \in \partial f_\infty$, where $\bar{\delta}_k = \sum_{i \in \hat{I}_k} \lambda_i^{(k)} g_i \to \bar{\delta}_\infty = \sum_{i \in \hat{I}_\infty} \lambda_i^{(\infty)} g_\infty$. Since $\bar{\delta}_k \to \bar{\delta}_i^\infty, \ \ \ k \in K_3$, and $\|\bar{\delta}_k\| \to 0, \ \ k \in K_1$, it follows that $0 \in \partial f_\infty$ (because of $K_3 \subseteq K_1$).

If the inequality (4.4) does not hold at Step 7, i.e. $g_{k+1}^T \bar{\delta}_k > -\frac{\eta}{2} \|\bar{\delta}_k\|^2$ we choose $x_{k+1} = x_k$ and consequently we have that $f(x_k) - f(x_{k+1}) = 0$ holds. Since $\|x_{k+1} - x_k\| = 0 < \mu$, at the next iteration we change the set I_k (adding the new k) and the bundle B_k (adding a new triplet $(x_k, f(x_k), g(x_{k+1}))$. In that case we find another solution of the problem (3.6), and consequently we find another solution of the problem (3.5), because we change the trust region radius. So, for the next iteration we get a new r_k for which either $r_k < \beta$ or $r_k \geq \beta$. In that case if $r_k < \beta$ holds, then at the new iteration condition (4.4) maybe holds. If not, then we compute another $g_{k+1} = g(x_{k+1}) \in \partial f(x_k + \alpha_k \bar{\delta}_k)$, and try to find another solution. This situation can be repeated finitely many times.

Suppose contrary, that is, the algorithm is passing infinitely many iterations when the condition (4.4) does not hold. Then for some k large enough the function $\Phi_k(x_k + \delta)$ attains its minimum at the point δ_k. So, $0 \in \partial\Phi_k(x_k + \delta_k)$, i.e. $0 \in \partial \hat{f}_k(x_k + \delta_k) + \delta_k$ and $0 = \hat{g} + \delta_k$ for some $\hat{g} \in \partial \hat{f}_k(x_k + \delta_k)$, i.e. $-\delta_k \in \partial \hat{f}_k(x_k + \delta_k)$. From $\Phi_k(x_k + 0) \geq \Phi_k(x_k + \delta_k)$ it follows that $\hat{f}_k(x_k) \geq \hat{f}_k(x_k + \delta_k) + \frac{1}{2}\|\delta_k\|^2$ and by Lemma 4.1 (as $k \to +\infty$) $f(x_k) \geq f(x_k + \delta_k) + \frac{1}{2}\|\delta_k\|^2 > f(x_k + \delta_k)$ (since at Step 3 we have that $\|\delta_k\| > \epsilon > 0$). So, we have the following inequality:

$$f(x_k) > f(x_k + \delta_k). \tag{4.7}$$

Since $r_k < \beta$ holds, i.e. $\frac{f(x_k) - f(x_k+\delta_k)}{f(x_k) - \Phi_k(x_k+\delta_k)} < \beta$, it follows that $\frac{f(x_k) - \Phi_k(x_k+\delta_k)}{f(x_k) - f(x_k+\delta_k)} > \frac{1}{\beta}$, i.e. $\frac{\Phi_k(x_k+\delta_k) - f(x_k)}{f(x_k+\delta_k) - f(x_k)} > \frac{1}{\beta}$. Hence we have that:

$$\begin{aligned} \frac{1}{\beta} < \frac{\Phi_k(x_k + \delta_k) - f(x_k)}{f(x_k + \delta_k) - f(x_k)} &= \frac{\hat{f}_k(x_k + \delta_k) + \frac{1}{2}\|\delta_k\|^2 - f(x_k)}{f(x_k + \delta_k) - f(x_k)} \\ &\leq \frac{f(x_k + \delta_k) + \frac{1}{2}\|\delta_k\|^2 - f(x_k)}{f(x_k + \delta_k) - f(x_k)} \quad (\text{since } f(x_k + \delta_k) \geq \hat{f}_k(x_k + \delta_k)) \\ &= 1 + \frac{\frac{1}{2}\|\delta_k\|^2}{f(x_k + \delta_k) - f(x_k)} \end{aligned}$$

hold, i.e. $\frac{\frac{1}{2}\|\delta_k\|^2}{f(x_k+\delta_k) - f(x_k)} > \frac{1}{\beta} - 1 > 0$, where the last inequality holds since $0.5 < \beta < 1$. From $\frac{\frac{1}{2}\|\delta_k\|^2}{f(x_k+\delta_k) - f(x_k)} > 0$ it follows that $f(x_k + \delta_k) > f(x_k)$ (since $\|\delta_k\| > \epsilon > 0$) which contradicts (4.7).

In the case (ii) we have that $f_1 - f_\infty \geq \sum_{k \in K_2} \triangle f_k$ (where sum is taken over the subsequence) and by the assumption $r_k \geq \beta, \quad k \in K_2$, implies $\triangle\Phi_k \to 0, \quad k \in K_2$, since $f_1 - f_\infty$ is constant. Let $\bar{h}$ satisfy the inequality $0 < \bar{h} < \inf(h_k)$ and let $\bar{\delta}$ be the minimum point of $\Phi_\infty(x_\infty + \delta)$ on $\|\delta\| \leq \bar{h}$. Since the point $\bar{x} = x_\infty + \bar{\delta}$ belongs to the set $\Omega_k = B(x_k, h_k)$ for k large enough, $k \in K_2$, by the definition of $\triangle\Phi_k$ and the fact that the minimum over the smaller set is not less than the minimum over the larger set it follows that

$$\Phi_k(x_k + \bar{\delta}) = \Phi_k(x_k + \bar{x} - x_\infty) \geq \Phi_k(x_k + \delta_k) = f_k - \triangle\Phi_k. \tag{4.8}$$

As $k \to +\infty, \quad k \in K_2$, since f and Φ_k are continuous as convex functions, and since $\triangle\Phi_k \to 0$ and $\bar{x} - x_k \to \bar{\delta}$, from (4.8) by Lemma 4.2 we have that:

$$\Phi_\infty(x_\infty + \bar{\delta}) \geq f_\infty - 0 = \Phi_\infty(x_\infty + 0)$$

holds. Notice that since $\delta = 0$ is the minimum point of $\Phi_\infty(x_\infty + \delta)$ for $\|\delta\| \leq \bar{h}$ then it follows that $\bar{\delta} = 0$. Since $\bar{\delta} = \underline{0}$ minimizes $\Phi_\infty(x_\infty + \delta)$ and since the later constraint is not active (because of $0 < \bar{h} < \inf(h_k)$) it follows that the first necessary condition holds, that is $0 \in \partial f_\infty$(because of $\Phi_\infty(x_\infty + 0) = \hat{f}_\infty(x_\infty) = f(x_\infty) = f_\infty$). Since the function f is convex it is also sufficient condition for a global minimum of the problem (1.1) at the point x_∞ □

5. CONCLUSION

Our algorithm defined above combines the bundle trust region and the conjugate subgradient method. To our knowledge it is a new approach to solve the nonsmooth problem defined by (1.1). If we use some other norms in (3.3) then it would be reasonable to expect some new results.

6. ACKNOWLEDGMENT

This research was supported by Science Fund of Serbia, grant number III44007.

REFERENCES

1. Bonnans J.F., J.C. Gilbert, C. Lemarchal, and C.A. Sagastizbal, *Numerical Optimization. Theoretical and Practical Aspects*, Springer, 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. Demyanov, V.F., Vasiliev, P.V., *Nondifferentiable Optimization. Moscow*, Nauka, 384 (in Russian), 1981.
5. Djuranović Miličić N., Gardašević Filipović M., *A Multi-step Curve Search Algorithm in Nonlinear Optimization - Nondifferentiable Convex Case*, Facta Universitatis Series Mathematics and Informatics, Vol.25, 11-24, 2010.
6. Djuranović Miličić N., Gardašević Filipović M., *An Algorithm For Minimization Of A Nondifferentiable Convex Function*, Lecture Notes in Engineering and Computer Science, WCE 2009, **VOL. II**, 1241-1246, 2009.
7. Djuranović Miličić N., Gardašević Filipović M., *On An Algorithm In Nondifferential Convex Optimization*, YUJOR **Vol.22 , No. 2**, 2012.
8. Djuranović Miličić N., Gardašević Filipović M., *Nondifferentiable Convex Optimization: An Algorithm Using Moreau-Yosida Regularization*, Lecture Notes in Engineering and Computer Science, WCE 2012, **VOL. I**, 131-136, 2012.
9. Ermoliev, Y. M.; Kryazhimskii, A. V.; Ruszczynskii, A., *Constraint Aggregation Principle in Convex Optimization*, Mathematical Programming **76** , No. 3, 53-372, 1997.
10. Fletcher R.,*Practical Methods of Optimization*, Volume 2, Department of Mathematics, University of Dundee, Scotland, U.K, Wiley-Interscience, Publication, 1981.

11. Fuduli A., *Metodi Numerici per la Minimizzazione di Funzioni Convesse Nondifferenziabili*, PhD Thesis, Departimento di Electronica Informatica e Sistemistica, Univerzita della Calabria, 1998.

12. Gardašević Filipović M., Djuranović Miličić N.,*An Algorithm Using Trust Region Strategy For Minimization Of A Nondifferentiable Function*, Numerical Functional Analysis and Optimization, **Vol. 32**, No. 12, 1239-1251, 2011.

13. Gardašević Filipović M., Djuranović Miličić N.,*An Algorithm Using Moreau-Yosida Regularization for Minimization of a Nondifferentiable Convex Function*, Filomat **27:1**, 13-20, 2013.

14. Karas E., Ribeiro A., Sagastizbal C., Solodov M., *A bundle-filter method for nonsmooth convex constrained optimization*, Mathematical Programming **116, No. 1-2**, 297-320, 2009.

15. Kiwiel K., *Convergence Of Approximate And Incremental Subgradient Methods For Convex Optimization*, SIAM Journal on Optimization, **Vol 14**, Issue 5, 807-840, 2004.

16. Kiwiel, K., *Proximity Control in Bundle Methods for Convex Nondifferentiable Minimization*, Mathematical Programming **46, No. 1**, 105-122, 1990.

17. Kusarev A.G., Kutateladze S.S., *Subdifferential Calculus*, Moscow, Nauka, 224 (in Russian), 1987.

18. Lemarchal, C., Nemirovskii, A.; Nesterov, Y, *New Variants of Bundle Methods*, Mathematical Programming,**Vol. 69, No. 1**, 111-147, 1995.

19. Lemarchal C., Sagastizabal C., *Practical aspects of the Moreau-Yosida regularization I: theoretical preliminaries*, SIAM Journal on Optimization, **Vol.7**, 367-385, 1997.

20. Lemarchal, 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. 245-282, Academic Press: New York, 1981.

21. Makela M., *Survey of Bundle Methods For Nonsmooth Optimization*, Optimization Methods and Software, **Vol 17**, 1-29, 2002.

22. Meng F., Zhao G., *On Second-order Properties of the Moreau-Yosida Regularization for Constrained Nonsmooth Convex Programs*, Numer. Funct. Anal. Optim , **Vol. 25; No. 5/6**, 515-530, 2004.

23. 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.

24. Rockafellar R., *Convex Analysis*, Princeton University Press, New Jersey, 1970.

25. 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.**, 146-169, 2005.

26. Sun W., Yuan Y., *Optimization theory and methods*, Nonlinear programming, ISBN-10:0-387-24975-3 Springer Science+Business Media, LLC, 2006.

27. Shor N.Z., Zhurbenko N.G., *A minimization method using the operation of extension of the space in the direction of the difference of the two successive gradients*, Cybernetics, **7(3)**, 450 - 459, 1971.

28. Shor N.Z., *Minimization Methods for Non-Differentiable Functions*, Springer-Verlag, Berlin, Heidelberg, New York, Tokyo, 1985.

29. Wolfe P., *A Method of Conjugate Subgradients for Minimizing Nondifferentiable Functions*, Nondifferentiable Optimization, Mathematical Programming Study **3**, (Eds. Balinski M. L. and Wolfe P.), 145-173, 1975.

Milanka Gardašević Filipović
Vocational College of Technology
Arandjelovac, Serbia
milanka.filipovic@vtsar.edu.rs

Nada DJuranović-Miličić
Faculty of Technology and Metallurgy
University of Belgrade
Belgrade, Serbia
nmilicic@tmf.bg.ac.rs