# A TRUST REGION METHOD USING SUBGRADIENT FOR MINIMIZING A NONDIFFERENTIABLE FUNCTION

## Milanka GARDAŠEVIĆ FILIPOVIĆ

*Faculty of Architecture,*
*University of Belgrade, Serbia*
*milanka@arh.bg.ac.yu*

**Abstract:** The minimization of a particular nondifferentiable function is considered. The first and second order necessary conditions are given. A trust region method for minimization of this form of the objective function is presented. The algorithm uses the subgradient instead of the gradient. It is proved that the sequence of points generated by the algorithm has an accumulation point which satisfies the first and second order necessary conditions

**Keywords**: Trust region method, non-smooth convex optimization.

## 1. INTRODUCTION

A motivation for the idea of the trust region method is to circumvent the difficulty caused by non-positive definite Hessian matrix in the well known Newton method. In this case the following quadratic function $q^{(k)}(\delta)$, obtained by truncating the Taylor series for $f(x^{(k)} + \delta)$, given as follows

$$f(x^{(k)} + \delta) \approx q^{(k)}(\delta) = f^{(k)} + g^{(k)^T} \delta + \frac{1}{2} \delta^T G^{(k)} \delta$$

does not have a unique minimum and the method is not well defined.

We used the following notation: denote by $f^{(k)} = f(x^{(k)})$, $g^{(k)} = g(x^{(k)}) = \nabla f(x^{(k)})$, where $\nabla$ denotes the gradient operator $\left(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, ..., \frac{\partial}{\partial x_n}\right)^T$, $G^{(k)} = G(x^{(k)})$ denotes the Hessian at $x^{(k)}$, and a vector refers to a

column vector. Another obstacle is that the region about $x^{(k)}$ in which the Taylor series approximates the function does not include a minimizing point of $q^{(k)}(\delta)$.

A more realistic approach therefore is to assume that it can be defined some neighborhood $\Omega^{(k)}$ of $x^{(k)}$ in which $q^{(k)}(\delta)$ agrees with $f(x^{(k)} + \delta)$ in some sense. Then it would be appropriate to choose $x^{(k+1)} = x^{(k)} + \delta^{(k)}$, where the correction $\delta^{(k)}$ minimizes $q^{(k)}(\delta)$ for all $x^{(k)} + \delta \in \Omega^{(k)}$. This is the reason for the name of the method – the trust region method (referred to the neighborhood $\Omega^{(k)}$).

It is usual to consider the case in which $\Omega^{(k)} = \left\{ x : \left\| x - x^{(k)} \right\| \leq h^{(k)} \right\}$ and $\delta^{(k)}$ is the solution of the subproblem $\min_{\delta} q^{(k)}(\delta)$ subject to $\left\| \delta \right\| \leq h^{(k)}$.

Denote by $\Delta f^{(k)} = f^{(k)} - f\left( x^{(k)} + \delta^{(k)} \right) = f^{(k)} - f^{(k+1)}$ the actual reduction and by $\Delta q^{(k)} = f^{(k)} - q^{(k)}\left( \delta^{(k)} \right)$ the predicted reduction. Then the ratio $r^{(k)} = \dfrac{\Delta f^{(k)}}{\Delta q^{(k)}}$ measures the accuracy with which $q^{(k)}(\delta)$ approximates $f(x^{(k)} + \delta)$. Naturally, accuracy is better when the ratio is closer to unity. Optimality conditions, the trust region algorithm and the convergence proof are given in [4].

A generalization of this case is made for the minimization of the function $\Phi(x) = f(x) + h(c(x))$ where $f : R^n \to R$ and $c : R^n \to R^m$ are twice differentiable functions and $h : R^m \to R$ is a convex function. Optimality conditions, the trust region algorithm and the convergence proof are given in [5].

The issue of this paper is a generalization of the previous case. We consider the following nonlinear programming problem:

$$\min_{x \in R^n} \tilde{\Phi}(x) = f(x) + \sum_{i=1}^{p} h_i(c(x)) \qquad (1)$$

where $f : R^n \to R$, $c : R^n \to R^m$ are smooth (that is continuous and continuously (Fréhet) differentiable) functions and $h_i : R^m \to R, i = 1, 2, ..., p$ are convex but non-smooth functions. It is supposed that interiors of domains have non-empty intersection for these functions ($h_i, i = 1, 2, ..., p$). This condition is used to apply the Moro-Rockafellar theorem.

In Section 2 some basic results necessary for further work are given. In Section 3 and Section 4 necessary and sufficient conditions for the solution of the problem (1) are given, respectively. Finally, in Section 5 a global model algorithm is given and its convergence is proved.

## 2. PRELIMINARIES

The next definition and few lemmas and theorems will be necessary in this work.

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

**Definition.** A vector $g \in R^n$ is said to be a subgradient of a convex function $f : R^n \to R$ at a point $x$ if the next inequality

$$f(z) \geq f(x) + (z - x)^T \cdot g \tag{2}$$

holds for all $z \in R^n$. The set of all subgradients of $f(x)$ at $x$ is called the subdifferential at $x$ is denoted by $\partial f(x)$.

The above definition has a simple geometric interpretation: since $f$ is convex we can find a supporting hyperplane at the boundary point $(x, f(x))$ that supports the epigraph of $f(x)$. The slope of the hyperplane is a subgradient $g$ of $f(x)$ at the point $x$.

**Lemma 1.** Let $f : R^n \to R$ be a differentiable function defined at $x \in R^n$. Then:

$$\partial f(x) = \{\nabla f(x)\}. \tag{3}$$

*Proof.* Follows directly from the Definition.

Obviously, the gradient $\nabla f(x)$ is the only possible subgradient. Furthermore, a point $x$ is a global minimum of a convex function $f(x)$ if and only if zero is contained in the subdifferential $\partial f(x)$. Geometrically, it means that we can draw a horizontal hyperplane which supports the epigraph of $f$ at $(x, f(x))$. This property is a generalization of the fact that the gradient of a function differentiable at a local minimum is zero.

**Lemma 2.** Let $f : S \to R$ be a convex function defined on a convex set $S \subseteq R^n$ and $x' \in \operatorname{int} S$. If $x^{(k)} \to x'$, where $x^{(k)} = x' + \delta^{(k)} s^{(k)}$, $\delta^{(k)} > 0$, $\delta^{(k)} \to 0$ and $s^{(k)} \to s$, then:

$$\lim_{k \to \infty} \frac{f^{(k)} - f'}{\delta^{(k)}} = \max_{g \in \partial f'} s^T g, \tag{4}$$

where $f' = f(x')$.

*Proof.* See [5] or [8].

**Theorem 1.** (Moro-Rockafellar)

Let $f_1 : C_1 \to R$, $f_2 : C_2 \to R$ be convex functions defined on convex sets $C_1, C_2 \subseteq R^n$ respectively and $\operatorname{int} C_1 \bigcap \operatorname{int} C_2 \neq \varnothing$. Then for all $x \in C_1 \bigcap C_2$ the following identity:

$$\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x) \tag{5}$$

holds.

*Proof.* See [1].

**Theorem 2.** Let $C$ be a closed convex set in $R^n$ and $\hat{x} \notin C$. Then there exists a hyperplane $H$ which separates the sets $C$ and $\{\hat{x}\}$.

*Proof.* See [1].

**Theorem 3.** Let $G : R^m \to R^n$ be a Fréchet-differentiable function and let $g : R^n \to R$ be a convex function continuous at the point $y_0 = G(x_0)$. For the function $f : R^m \to R$ defined by $f(x) = g(G(x))$ follows that:

$$\partial f(x_0) = G'(x_0) \partial g(G(x_0)) . \tag{6}$$

*Proof.* See [1].

**Lemma 3.** Let $f : S \to R$ 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 set.

*Proof.* See [1] or [7].

## 3. THE FIRST ORDER NECESSARY CONDITION

**Lemma 4.** Let $x^{(k)}$ be a sequence such that $x^{(k)} \to x'$, $x^{(k)} = x' + \delta^{(k)} s^{(k)}$ where $s^{(k)} \to s$, $\delta^{(k)} > 0$, $\delta^{(k)} \to 0$ and let $\tilde{\Phi}' = \tilde{\Phi}(x')$ and $\tilde{\Phi}^{(k)} = \tilde{\Phi}(x^{(k)})$ be the values of the function defined in (1) at the points $x'$ and $x^{(k)}$, respectively. Then

$$\lim_{k \to \infty} \frac{\tilde{\Phi}^{(k)} - \tilde{\Phi}'}{\delta^{(k)}} = s^T g' + \sum_{i=1}^{p} \max_{\lambda_i \in \partial h_i'} s^T A' \lambda_i \tag{7}$$

(where $f' = f(x')$, $g' = g(x') = \nabla f(x')$ and $c^{(k)} = c(x^{(k)})$, $c' = c(x')$, $A' = A(x') = \nabla c(x')$, $h_i' = h_i(c(x'))$ ).

*Proof.* By Taylor expansion of $f(x^{(k)})$ we have $f^{(k)} = f' + \delta^{(k)} g'^T s^{(k)} + o(\delta^{(k)})$ and $\frac{f^{(k)} - f'}{\delta^{(k)}} \to g'^T s$ (because $f$ is a smooth function). Similarly, we have

$c^{(k)} = c' + \delta^{(k)} A'^T s^{(k)} + o(\delta^{(k)})$ and since $c^{(k)} \to c'$ we have $\frac{c^{(k)} - c'}{\delta^{(k)}} \to A'^T s$. Then:

$$\lim_{k \to \infty} \frac{\tilde{\Phi}^{(k)} - \tilde{\Phi}'}{\delta^{(k)}} = \lim_{k \to \infty} \frac{f^{(k)} - f' + \sum_{i=1}^{p} (h_i(c(x^{(k)})) - h_i(c(x')))}{\delta^{(k)}}$$

$$= \lim_{k \to \infty} \frac{f^{(k)} - f'}{\delta^{(k)}} + \sum_{i=1}^{p} \lim_{k \to \infty} \frac{h_i(c(x^{(k)})) - h_i(c(x'))}{\delta^{(k)}}$$

$$= s^T g' + \sum_{i=1}^{p} \max_{\overline{\lambda}_i \in \partial[h_i(c(\cdot))](x')} s^T \overline{\lambda}_i = s^T g' + \sum_{i=1}^{p} \max_{\lambda_i \in \partial h_i'} s^T A' \lambda_i$$

where the last two equalities hold by Lemma 2 and Theorem 3, respectively.∎

If we denote by $x^*$ a local minimum of $\tilde{\Phi}$, then $\tilde{\Phi}^{(k)} \geq \tilde{\Phi}^*$ holds for all k large enough (where $x^{(k)} \to x^*$). Now, if the notations $h_i^* = h_i(c(x^*))$ for $i = 1, 2, ..., p$; $f^* = f(x^*)$, $g^* = g(x^*) = \nabla f(x^*)$ and $A^* = A(x^*) = \nabla c(x^*)$ are used, then by (7) we have

$$s^T g^* + \sum_{i=1}^{p} \max_{\lambda_i \in \partial h_i^*} s^T A^* \lambda_i \geq 0, \ \forall s : \|s\| = 1 . \tag{8}$$

The condition (8) is the first order necessary condition for a local minimum of the function $\tilde{\Phi}$ (see [2]). The condition (8) means that the directional derivative is non-negative in all directions. This can be stated alternatively as:

$$0 \in \partial \tilde{\Phi}^* \tag{9}$$

where $\partial \tilde{\Phi}^* = \left\{ \mu : \mu = g^* + A^* \sum_{i=1}^{p} \lambda_i^*, \lambda_i^* \in \partial h_i^*, i = 1, 2, ..., p \right\}$. The set $\partial \tilde{\Phi}^*$, defined in this way, is not the subdifferential because $\tilde{\Phi}$ may be not a convex function, but it is convenient to use the same notation.

**Lemma 5.** Let $f : R^n \to R$ be a smooth convex function and $c : R^n \to R^m$ and $h_i : R^m \to R$, $i = 1, 2, ..., p$ are monotone non-decreasing convex functions, such that the interiors of domains for $h_i$, $i = 1, 2, ..., p$ have a non-empty intersection. Then the conditions (8) and (9) are equivalent.

_Proof._ Under the above assumptions the function $\tilde{\Phi}(x) = f(x) + \sum_{i=1}^{p} h_i(c(x))$ is convex. Namely, since we suppose that $f$ is a convex function, we have:

$$\partial \tilde{\Phi}^* = \partial (f + \sum_{i=1}^{p} h_i(c(\cdot)))(x^*)$$

$$= \partial f^* + \partial \left( \sum_{i=1}^{p} h_i(c(\cdot)) \right)(x^*) \text{ (by Moro-Rockafellar theorem)}$$

$$= \partial f^* + \sum_{i=1}^{p} \partial h_i(c(\cdot))(x^*) \quad \text{(by Moro-Rockafellar theorem)}$$

$$= g^* + \sum_{i=1}^{p} A^* \partial h_i \left( c(x^*) \right) \quad \text{(by Theorem 3)}$$

$$= g^* + A^* \sum_{i=1}^{p} \partial h_i^* .$$

If $0 \in \partial \tilde{\Phi}(x^*)$ and $\xi^*$ is a vector from $\partial \tilde{\Phi}(x^*)$ then :

$$0 = s^T \xi^* = s^T \left( g^* + A^* \sum_{i=1}^{p} \lambda_i^* \right) \leq s^T \left( g^* + A^* \sum_{i=1}^{p} \max_{\lambda_i \in \partial h_i^*} \lambda_i \right)$$

where $\lambda_i^* \in \partial h_i^*$, $i = 1, 2, ..., p$ . Hence, (9) implies (8).

Suppose that $0 \notin \partial \tilde{\Phi}^*$. Since $\partial \tilde{\Phi}^*$ is a convex set, it follows that the point $0$ and the set $\partial \tilde{\Phi}^*$ can be separated by a hyperplane. By Theorem 2 there exists a vector $s = -\dfrac{\overline{\xi}}{\|\overline{\xi}\|}$, where $\min_{\xi \in \partial \tilde{\Phi}^*} \|\xi\| = \|\overline{\xi}\|$, such that $s^T \xi < 0, \forall \xi \in \partial \tilde{\Phi}^*$. Then it follows that $\max_{\xi \in \partial \tilde{\Phi}^*} s^T \xi < 0$ , contradictory to (8).

Hence the conditions (8) and (9) are equivalent .∎

Another way to state the condition (9) is to introduce the Lagrangian function:

$$\tilde{L}(x, \lambda_1, \lambda_2, ..., \lambda_p) = f(x) + c^T(x) \sum_{i=1}^{p} \lambda_i . \tag{10}$$

**Theorem 4.** If $x^*$ is a minimum point of the function $\tilde{\Phi}$, then there exist the vectors $\lambda_1^* \in \partial h_1^*, \lambda_2^* \in \partial h_2^*, ..., \lambda_p^* \in \partial h_p^*$ such that:

$$\nabla \tilde{L}(x^*, \lambda_1^*, \lambda_2^*, ..., \lambda_p^*) = g^* + A^* \sum_{i=1}^{p} \lambda_i^* = 0 . \tag{11}$$

_Proof._ The proof is obtained, since $\partial \tilde{\Phi}^*$ is the set of the vectors $\nabla \tilde{L}(x^*, \lambda_1^*, \lambda_2^*, ..., \lambda_p^*)$ for all $\lambda_1^* \in \partial h_1^*, \lambda_2^* \in \partial h_2^*, ..., \lambda_p^* \in \partial h_p^*$ .∎

## 4. THE SECOND ORDER CONDITIONS

Now we can consider the second order conditions for the problem (1). The first step is a restriction of possible directions to those for which the function $\tilde{\Phi}$ has zero directional derivatives, so that the second order effects become important.

Consider the set:

$$\tilde{X} = \left\{ x : \sum_{i=1}^{p} h_i(c(x)) = \sum_{i=1}^{p} h_i(c(x^*)) + (c(x) - c(x^*))^T \sum_{i=1}^{p} \lambda_i^* \right\} \tag{12}$$

where $\lambda_i^*, i = 1, 2, ..., p$ are vectors from (11). Define $\tilde{H}^*$ as the set of normalized feasible directions with respect to the set $\tilde{X}$ at the point $x^*$. It means that if $s \in \tilde{H}^*$ then there exists a sequence $x^{(k)} \to x^*$, ($\{x^{(k)}\}$ is feasible in $\tilde{X}$) such that $s^{(k)} \to s, \|s^{(k)}\|_2 = 1, \delta^{(k)} \to 0$, $\delta^{(k)} > 0$ and $x^{(k)} = x^* + \delta^{(k)} s^{(k)}$. It is possible to prove that these directions are closely related to the set of normalized directions of zero slope which is denoted by $\tilde{G}^*$:

$$\tilde{G}^* = \left\{ s : s^T \left( g^* + A^* \sum_{i=1}^{p} \max_{\lambda_i \in \partial h_i^*} \lambda_i \right) = 0, \ \|s\|_2 = 1 \right\} \tag{13}$$

**Lemma 6.** $\tilde{H}^* \subseteq \tilde{G}^*$

*Proof.* Let $s \in \tilde{H}^*$. Then there exists a directional sequence in $\tilde{X}$, such that $s^{(k)} \to s, \|s\|_2 = 1$. Using (1), (7), (12) and the Taylor expansions for functions $f$ and $c$ it follows that:

$$s^T \left( g^* + A^* \sum_{i=1}^{p} \max_{\lambda_i \in \partial h_i^*} \lambda_i \right) = \lim_{k \to \infty} \frac{\tilde{\Phi}^{(k)} - \tilde{\Phi}^*}{\delta^{(k)}} = \lim_{k \to \infty} \frac{f^{(k)} - f^* + \sum_{i=1}^{p} (h_i(c^{(k)}) - h_i(c^*))}{\delta^{(k)}} =$$

$$= \lim_{k \to \infty} \frac{f^{(k)} - f^* + (c^{(k)} - c^*)^T \sum_{i=1}^{p} \lambda_i^*}{\delta^{(k)}} = s^T \left( g^* + A^* \sum_{i=1}^{p} \lambda_i^* \right) = 0,$$

where the last equality follows from (11). Hence, $s^T \left( g^* + A^* \sum_{i=1}^{p} \max_{\lambda_i \in \partial h_i^*} \lambda_i \right) = 0$ for $\|s\|_2 = 1$,

implies $s \in \tilde{G}^*$. So, $\tilde{H}^* \subseteq \tilde{G}^*$. ∎

We suppose that the regularity condition

$$\tilde{H}^* = \tilde{G}^* \tag{14}$$

is satisfied. Now, it is possible to state the second order conditions. We suppose that the functions $f$ and $c$ are twice continuously differentiable.

**Theorem 5**. (Second order necessary conditions)

If $x^*$ is a minimizing point of $\tilde{\Phi}$ then by Lemma 4 for all vectors $\lambda_1^* \in \partial h_1^*, \lambda_2^* \in \partial h_2^*, ..., \lambda_p^* \in \partial h_p^*$, which thus exist, if (14) holds, we have:

$$s^T \nabla^2 \tilde{L}(x^*, \lambda_1^*, \lambda_2^*, ..., \lambda_p^*)s \geq 0 \ , \ \forall s \in \tilde{G}^* . \tag{15}$$

<u>*Proof.*</u> For any $s \in \tilde{G}^* = \left\{ s : s^T \left( g^* + A^* \sum_{i=1}^{p} \max_{\lambda_i \in \partial h_i^*} \lambda_i \right) = 0, \ \|s\|_2 = 1 \right\}$, by (14) it follows that $s \in \tilde{H}^*$.

Taylor expansion of the function $\tilde{L}(x, \lambda_1^*, \lambda_2^*, ..., \lambda_p^*)$ about $x^*$ yields:

$$\tilde{L}(x^{(k)}, \lambda_1^*, \lambda_2^*, ..., \lambda_p^*) = \tilde{L}(x^*, \lambda_1^*, \lambda_2^*, ..., \lambda_p^*) + (x^{(k)} - x^*)^T \nabla \tilde{L}(x^*, \lambda_1^*, \lambda_2^*, ..., \lambda_p^*)$$

$$+ \frac{1}{2}(x^{(k)} - x^*)^T \nabla \tilde{L}(x^*, \lambda_1^*, \lambda_2^*, ..., \lambda_p^*)(x^{(k)} - x^*) + o(\|x^{(k)} - x^*\|^2)$$

$$= f^* + c^{*T} \sum_{i=1}^{p} \lambda_i^* + \delta^{(k)} s^{(k)T} \nabla \tilde{L}(x^*, \lambda_1^*, \lambda_2^*, ..., \lambda_p^*) + \tag{16}$$

$$\frac{1}{2} \delta^{(k)} s^{(k)T} \nabla^2 \tilde{L}(x^*, \lambda_1^*, \lambda_2^*, ..., \lambda_p^*) \delta^{(k)} s^{(k)} + o(\|\delta^{(k)} s^{(k)}\|^2)$$

$$= f^* + c^{*T} \sum_{i=1}^{p} \lambda_i^* + \frac{1}{2} \delta^{(k)} s^{(k)T} \nabla^2 \tilde{L}(x^*, \lambda_1^*, \lambda_2^*, ..., \lambda_p^*) \delta^{(k)} s^{(k)} + o(\delta^{(k)^2}) \ ,$$

where the last equality follows by (11), writing $\delta^{(k)} s^{(k)} = x^{(k)} - x^*$. Then:

$$\tilde{\Phi}^{(k)} = \tilde{\Phi}(x^{(k)}) = \tilde{\Phi}(x^* + \delta^{(k)} s^{(k)}) = f(x^* + \delta^{(k)} s^{(k)}) + \sum_{i=1}^{p} h_i(c(x^* + \delta^{(k)} s^{(k)}))$$

$$= f(x^* + \delta^{(k)} s^{(k)}) + \sum_{i=1}^{p} h_i(c(x^*)) + (c(x^* + \delta^{(k)} s^{(k)}) - c(x^*))^T \sum_{i=1}^{p} \lambda_i^*$$

$$= f(x^*) + \delta^{(k)} s^{(k)T} \nabla f(x^*) + \frac{1}{2} \delta^{(k)} s^{(k)T} \nabla^2 f(x^*) \delta^{(k)} s^{(k)} + o(\delta^{(k)^2}) + \sum_{i=1}^{p} h_i(c(x^*)) +$$

$$+ \left[ \delta^{(k)} s^{(k)T} \nabla c(x^*) + \frac{1}{2} \delta^{(k)} s^{(k)T} \nabla^2 c(x^*) \delta^{(k)} s^{(k)} + o(\delta^{(k)^2}) \right] \sum_{i=1}^{p} \lambda_i^* =$$

$$= \tilde{\Phi}^* + \delta^{(k)} s^{(k)T} \left[ \nabla f(x^*) + \nabla c(x^*) \sum_{i=1}^{p} \lambda_i^* \right] +$$

$$\frac{1}{2} \delta^{(k)T} \left[ \nabla^2 f(x^*) + \nabla^2 c(x^*) \sum_{i=1}^{p} \lambda_i^* \right] \delta^{(k)} s^{(k)} + o(\delta^{(k)2})$$

$$= \tilde{\Phi}^* + \delta^{(k)} s^{(k)T} \nabla \tilde{L}(x^*, \lambda_1^*, \lambda_2^*, ..., \lambda_p^*) + \frac{1}{2} \delta^{(k)} s^{(k)T} \nabla^2 \tilde{L}(x^*, \lambda_1^*, \lambda_2^*, ..., \lambda_p^*) \delta^{(k)} s^{(k)} +$$

$$o(\delta^{(k)^2}) .$$

Since the first order necessary conditions are satisfied, we have that:

$$\tilde{\Phi}^{(k)} = \tilde{\Phi}^* + \frac{1}{2}\delta^{(k)} s^{(k)T} \nabla^2 \tilde{L}(x^*, \lambda_1^*, \lambda_2^*, ..., \lambda_p^*)\delta^{(k)} s^{(k)} + o(\delta^{{(k)}^2}) \ .$$

Since $x^*$ is a minimum point of the function $\tilde{\Phi}$, it follows that $\tilde{\Phi}^{(k)} \geq \tilde{\Phi}^*$ for every k large enough and consequently,

$$\tilde{\Phi}^{(k)} - \tilde{\Phi}^* = \frac{1}{2}\delta^{(k)} s^{(k)T} \nabla^2 \tilde{L}(x^*, \lambda_1^*, \lambda_2^*, ..., \lambda_p^*)\delta^{(k)} s^{(k)} \geq 0 \ .$$

Hence, dividing by $\frac{1}{2}\delta^{(k)2}$ and taking the limit yields, it follows that:

$$s^{(k)T} \nabla^2 \tilde{L}(x^*, \lambda_1^*, \lambda_2^*, ..., \lambda_p^*)s^{(k)} \geq 0 \ , s^{(k)} \to s \Rightarrow s^T \nabla^2 \tilde{L}(x^*, \lambda_1^*, \lambda_2^*, ..., \lambda_p^*)s \geq 0 \ . \blacksquare$$

**Theorem 6.** (Second order sufficient conditions)

If the vectors $\lambda_1^* \in \partial h_1^*, \lambda_2^* \in \partial h_2^*, ..., \lambda_p^* \in \partial h_p^*$ there exist such that (11) holds and if the inequality

$$s^T \nabla^2 \tilde{L}(x^*, \lambda_1^*, \lambda_2^*, ..., \lambda_p^*)s > 0 \ , \ \forall s \in \tilde{G}^* \tag{17}$$

is satisfied, then it follows that $x^*$ is a local minimum of the function $\tilde{\Phi}(x)$.

_Proof._ Assume the contrary: there exists a sequence $x^{(k)} \to x^*$ such that $\tilde{\Phi}^{(k)} \leq \tilde{\Phi}^*$.

By (11) and (7) it follows that $\mu = \max\limits_{\lambda \in \partial h^*} s^T \left( g^* + A^* \sum\limits_{i=1}^{p} \lambda_i \right) \geq 0$. Then it follows that

$$\max_{\lambda_i \in \partial h_i'} \left( s^T g' + \sum_{i=1}^{p} s^T A' \lambda_i \right) = s^T g' + \sum_{i=1}^{p} \max_{\lambda_i \in \partial h_i'} s^T A' \lambda_i = \lim_{k \to \infty} \frac{\tilde{\Phi}^{(k)} - \tilde{\Phi}'}{\delta^{(k)}} \ .$$

If $\mu > 0$ then $\Phi^{(k)} > \Phi^*$ which contradicts the fact that $\Phi^{(k)} \leq \Phi^*$. So $\mu = 0$, and hence $G^* \neq \varnothing$, because $s \in G^*$.

Because of the definitions of the functions $\tilde{L}(x, \lambda_1, \lambda_2, ..., \lambda_p)$ and $\tilde{\Phi}(x)$ it follows that:

$$\tilde{L}(x^{(k)}, \lambda_1^*, \lambda_2^*, ..., \lambda_p^*) - \tilde{L}(x^*, \lambda_1^*, \lambda_2^*, ..., \lambda_p^*) = f(x^{(k)}) + c^T(x^{(k)})\sum_{i=1}^{p}\lambda_i^* - f(x^*) - c^T(x^*)\sum_{i=1}^{p}\lambda_i^*$$

$$= \tilde{\Phi}(x^{(k)}) - \tilde{\Phi}(x^*) - \left( \sum_{i=1}^{p} h_i(c(x^{(k)}) - c(x^*)) - (c(x^{(k)}) - c(x^*))\sum_{i=1}^{p}\lambda_i^* \right) \leq \tilde{\Phi}(x^{(k)}) - \tilde{\Phi}(x^*)$$

(where the last inequality follows from the subgradient inequality (2)).

From (16) it follows that:

$$0 \geq \tilde{\Phi}(x^{(k)}) - \tilde{\Phi}(x^*) = \frac{1}{2}\delta^{(k)}s^{(k)T}\nabla^2 \ \tilde{L}(x^*,\lambda_1^*,\lambda_2^*,...,\lambda_p^*) \ \delta^{(k)}s^{(k)} + o(\delta^{(k)2}).$$

Hence, dividing by $\frac{1}{2}\delta^{(k)2}$ and taking the limit yields $0 \geq s^T \nabla^2 L(x^*, \lambda^*)s$, which contradicts (17). The theorem is established.∎

## 5. A GLOBALLY CONVERGENT MODEL ALGORITHM

In this part of the paper we apply the trust region method to find a minimum of the function:

$$\tilde{\Phi}(x) = f(x) + \sum_{i=1}^{p} h_i(c(x))$$

defined previously in (1). We suppose that the function $\tilde{\Phi}(x)$ has a minimum point $x^*$, and try to find this point by an iterative method. So, we suppose that we have the point $x^{(k)}$ at the k-th iterative step, and we are going to approximate the function $\tilde{\Phi}(x)$ by Taylor expansion in some neighborhood of this point:

$$\tilde{\psi}^{(k)}(\delta) = f^{(k)} + \sum_{i=1}^{p} h_i^{(k)} + \delta^T\left(g^{(k)} + A^{(k)T}\sum_{i=1}^{p}\lambda_i^{(k)}\right) + \frac{1}{2}\delta^T\left(\nabla^2 f^{(k)} + \nabla^2 c^{(k)T}\sum_{i=1}^{p}\lambda_i^{(k)}\right)\delta \quad (18)$$

where $A^{(k)} = \nabla c(x^{(k)})$, $h_i^{(k)} = h_i(c(x^{(k)}))$, $\lambda_i^{(k)} \in \partial h_i^{(k)}$ for $i = 1, 2, ..., p$ .

We suppose that there exists some neighborhood $\Omega^{(k)}$ of the point $x^{(k)}$ in which the approximation $\tilde{\psi}^{(k)}(\delta)$ agrees with $\tilde{\Phi}(x^{(k)} + \delta)$ in some sense. Then it would be appropriate to choose $x^{(k+1)} = x^{(k)} + \delta^{(k)}$, where the correction $\delta^{(k)}$ is a minimum of $\tilde{\psi}^{(k)}(\delta)$ for all $x^{(k)} + \delta^{(k)} \in \Omega^{(k)}$ .

We suppose that:

$$\Omega^{(k)} = \left\{x : \left\|x - x^{(k)}\right\| \leq h^{(k)}\right\} \quad (19)$$

and $\delta^{(k)}$ is a solution of the problem:

$$min\,\tilde{\psi}^{(k)}(\delta) \quad \text{subject to} \quad \left\|\delta\right\| \leq h^{(k)}. \quad (20)$$

The radius $h^{(k)}$ has to be such that the agreement of the function $\tilde{\Phi}(x^{(k)} + \delta)$ and its approximation $\tilde{\psi}^{(k)}(\delta)$ is satisfied in some sense. This can be quantified as follows.

Denote by

$$\Delta\tilde{\Phi}^{(k)} = \tilde{\Phi}^{(k)} - \tilde{\Phi}(x^{(k)} + \delta^{(k)}) = \tilde{\Phi}^{(k)} - \tilde{\Phi}^{(k+1)} \tag{21}$$

the actual reduction of the function $\tilde{\Phi}$ and by

$$\Delta\tilde{\psi}^{(k)} = \tilde{\Phi}^{(k)} - \tilde{\psi}^{(k)}(\delta^{(k)}) \tag{22}$$

the predicted reduction of the function $\tilde{\Phi}$ . Then the ratio:

$$r^{(k)} = \frac{\Delta\tilde{\Phi}^{(k)}}{\Delta\tilde{\psi}^{(k)}} \tag{23}$$

measures the accuracy by which $\tilde{\psi}^{(k)}(\delta^{(k)})$ approximates $\tilde{\Phi}^{(k)}$ (naturally, agreement is better if the ratio is closer to unity).

The k-th step of the model algorithm is as follows:

1. Given $x^{(k)}$, $\lambda_i^{(k)}$ for $i = 1, 2, ..., p$ and $h^{(k)}$, constants $\gamma$ and $\beta$ such that $\gamma \in (0, 0.5)$ and $\gamma < \beta < 1$;

   calculate $f^{(k)}, h_i^{(k)}, g^{(k)}, \nabla^2 c^{(k)}, A^{(k)}, \nabla^2 f^{(k)}$ and $\tilde{\Phi}^{(k)}$ and $\tilde{\psi}^{(k)}$

2. Find a solution $\delta^{(k)}$ of the problem (20)

3. Evaluate $\tilde{\Phi}(x^{(k)} + \delta^{(k)})$, $\Delta\tilde{\Phi}^{(k)}$, $\Delta\tilde{\psi}^{(k)}$ and $r^{(k)}$

   If $r^{(k)} < \gamma$ then set $h^{(k+1)} = \gamma \|\delta^{(k)}\|$, $\qquad\qquad\qquad$ (24)

   if $r^{(k)} > \beta$ and $h^{(k)} = \|\delta^{(k)}\|$ then set $h^{(k+1)} = 2h^{(k)}$;

   otherwise set $h^{(k+1)} = h^{(k)}$

4. If $r^{(k)} \leq 0$ then set $x^{(k+1)} = x^{(k)}$, $\lambda_i^{(k+1)} = \lambda_i^{(k)}$ for $i = 1, 2, ..., p$;

   else set $x^{(k+1)} = x^{(k)} + \delta^{(k)}$, $\lambda_i^{(k+1)} \in \partial h_i(c^{(k+1)})$ for $i = 1, 2, ..., p$

Now it is possible to prove the main result of this paper.

**Theorem 7.** Let the sequence $\{x^{(k)}\}$ be generated by the algorithm. Let $x^{(k)} \in B \subset R^n, \forall k$, where B is a bounded set and let the functions $f$ and $c$ be twice differentiable with bounded matrices of the second derivatives on the set B. Then there exists an accumulation point $x^\infty$ of the sequence in $\{x^{(k)}\}$ which satisfies the first order necessary condition for the problem (1); i.e. the condition

$$s^T g^\infty + \sum_{i=1}^{p} \max_{\lambda_i \in \partial h_i^\infty} s^T A^\infty \lambda_i \geq 0, \; \forall s \tag{25}$$

is satisfied.

_Proof._ Since the set B is bounded, it follows that every sequence in B has an accumulation point. Hence there exists a convergent subsequence $x^{(k)} \to x^{\infty}$. This sequence satisfies either:

(i)        $r^{(k)} < \gamma$   and   $h^{(k+1)} \to 0$     (and hence $\left\| \delta^{(k)} \right\| \to 0$) or

(ii)       $r^{(k)} \geq \beta$ and   $inf(h^{(k)}) > 0$.

We will prove that (25) holds in any case ((i) or (ii)).

In case (i) we suppose that there exists a descent direction $s$ ($\left\| s \right\| = 1$) at $x^{\infty}$, such that:

$$s^T g^{\infty} + \sum_{i=1}^{p} \max_{\lambda \in \partial h_i^{\infty}} s^T A^{\infty} \lambda_i = -d, d > 0 .  \tag{26}$$

Using Taylor expansion it follows that:

$$f(x^{(k)} + \varepsilon^{(k)} s) = f^{(k)} + \varepsilon^{(k)} s^T g^{(k)} + \frac{1}{2} \varepsilon^{(k)2} s^T \tilde{W}^{(k)} s + o(\varepsilon^{(k)2})  = \tilde{q}^{(k)}(\varepsilon^{(k)} s) + o(\varepsilon^{(k)2})  \tag{27}$$

(where   $\tilde{W}^{(k)} = \nabla^2 f^{(k)} + \nabla^2 c^{(k)T} \sum_{i=1}^{p} \lambda_i^{(k)}$   and   $\tilde{q}^{(k)}(\varepsilon^{(k)} s) = f^{(k)} + \varepsilon^{(k)} s^T g^{(k)} + \frac{1}{2} \varepsilon^{(k)2} s^T \tilde{W}^{(k)} s$),

and similarly :

$$c(x^{(k)} + \varepsilon^{(k)} s) = c^{(k)} + A^{(k)T} \varepsilon^{(k)} s + o(\varepsilon^{(k)}) = l^{(k)}(\varepsilon^{(k)} s) + o(\varepsilon^{(k)})  \tag{28}$$

(where   $l^{(k)}(\varepsilon^{(k)} s) = c^{(k)} + A^{(k)T} \varepsilon^{(k)} s$ ).

Hence by (27) and (28) we get

$$\tilde{\Phi}(x^{(k)} + \varepsilon^{(k)} s) = f(x^{(k)} + \varepsilon^{(k)} s) + \sum_{i=1}^{p} h_i(c(x^{(k)} + \varepsilon^{(k)} s))$$

$$= \tilde{q}^{(k)}(\varepsilon^{(k)} s) + o(\varepsilon^{(k)2}) + \sum_{i=1}^{p} h_i(l^{(k)}(\varepsilon^{(k)} s) + o(\varepsilon^{(k)}))$$

$$= \tilde{q}^{(k)}(\varepsilon^{(k)} s) + \sum_{i=1}^{p} h_i(l^{(k)}(\varepsilon^{(k)} s)) + o(\varepsilon^{(k)})$$

$$= \tilde{\psi}^{(k)}(\varepsilon^{(k)} s) + o(\varepsilon^{(k)})  \tag{29}$$

If we put $\varepsilon^{(k)} = \left\| \delta^{(k)} \right\|$ and take a step $\delta^{(k)}$ from the sub problem (20) along $s$, then because of the optimality of $\delta^{(k)}$ and (7), (18), (26) it follows that:

$$\Delta \tilde{\psi}^{(k)} = \tilde{\Phi}^{(k)} - \tilde{\psi}^{(k)}(\delta^{(k)}) \geq \tilde{\Phi}^{(k)} - \tilde{\psi}^{(k)}(\varepsilon^{(k)} s)$$

$$= \tilde{\Phi}^{(k)} - \tilde{\Phi}(x^{(k)} + \varepsilon^{(k)}s) + o(\varepsilon^{(k)}) + o(\varepsilon^{(k)2})$$

$$= \varepsilon^{(k)}(d + o(1)) = d\varepsilon^{(k)} + o(\varepsilon^{(k)}). \tag{30}$$

So, (29) implies:

$$\Delta\tilde{\Phi}^{(k)} = \tilde{\Phi}^{(k)} - \tilde{\Phi}^{(k+1)} = \tilde{\Phi}^{(k)} - \tilde{\psi}^{(k)}(\delta^{(k)}) + o(\varepsilon^{(k)}) = \Delta\tilde{\psi}^{(k)} + o(\varepsilon^{(k)})$$

and hence $r^{(k)} = \dfrac{\Delta\tilde{\Phi}^{(k)}}{\Delta\tilde{\psi}^{(k)}} = 1 + \dfrac{o(\varepsilon^{(k)})}{\Delta\tilde{\psi}^{(k)}} = 1 + o(1)$ , because $d > 0$, which contradicts $r^{(k)} < \gamma$ .

Thus $d \le 0$ for all $s$ and hence (23) holds at $x^\infty$ .

In the case (ii) we have that $\tilde{\Phi}^{(1)} - \tilde{\Phi}^\infty \ge \sum_k \Delta\tilde{\Phi}^{(k)}$ (where the sum is taken over the subsequence) and by the assumption $r^{(k)} \ge \beta$ it follows that $\Delta\tilde{\psi}^{(k)} \to 0$, because $\tilde{\Phi}^{(1)} - \tilde{\Phi}^\infty$ is constant. Since $r^{(k)} = \dfrac{\Delta\tilde{\Phi}^{(k)}}{\Delta\tilde{\psi}^{(k)}} \Rightarrow \Delta\tilde{\Phi}^{(k)} = r^{(k)}\Delta\tilde{\psi}^{(k)} \ge \beta\Delta\tilde{\psi}^{(k)}$ and $\tilde{\Phi}^{(1)} - \tilde{\Phi}^\infty$ is constant, and $\tilde{\Phi}^{(1)} - \tilde{\Phi}^\infty \ge \sum_k \Delta\tilde{\Phi}^{(k)} \ge \beta\sum_k \Delta\tilde{\psi}^{(k)}$ then it follows that $\sum_k \Delta\tilde{\psi}^{(k)}$ converges, and finally it follows that $\Delta\tilde{\psi}^{(k)} \to 0$ .

Let $\tilde{\psi}^\infty(\delta) = \tilde{q}^\infty(\delta) + \sum_{i=1}^{n} h_i(l^\infty(\delta))$ .

Let $\overline{h}$ satisfy the inequality $0 < \overline{h} < inf(h^{(k)})$ and $\overline{\delta}$ be a minimum of the $\tilde{\psi}^\infty(\delta)$ for all $\|\delta\| \le \overline{h}$ . Since $\|x - x^{(k)}\| = \|x^\infty - x^{(k)} + \overline{\delta}\| \le \|x^\infty - x^{(k)}\| + \|\overline{\delta}\| \le \overline{h} < inf(h^{(k)}) \le h^{(k)}$ holds, the point $\overline{x} = x^\infty + \overline{\delta}$ belongs to the set $\Omega^{(k)} = \{x : \|x - x^{(k)}\| \le h^{(k)}\}$ for k large enough. Hence, by the definition of $\Delta\tilde{\psi}^{(k)}$ it follows that $\tilde{\psi}^{(k)}(\overline{x} - x^\infty) \ge \psi^{(k)}(\delta^{(k)}) = \tilde{\Phi}^{(k)} - \Delta\tilde{\psi}^{(k)}$ (because of the known fact that the minimum over the smaller set is not less then the minimum over the larger set).

Taking the limit as $k \to \infty$ we have that $\tilde{\Phi}^{(k)} \to \tilde{\Phi}^\infty$, $g^{(k)} \to g^\infty$, $A^{(k)} \to A^\infty$, $\Delta\tilde{\psi}^{(k)} \to 0$, and $\overline{x} - x^{(k)} \to \overline{\delta}$. By the continuity of the function $\tilde{\psi}$ it follows that $\tilde{\psi}^\infty(\overline{\delta}) \ge \tilde{\Phi}^\infty - 0 = \tilde{\psi}^\infty(0)$ holds. Notice that since $\delta = 0$ minimizes $\tilde{\psi}^\infty(\delta)$ for $\|\delta\| \le \overline{h}$ it follows that $\overline{\delta} = 0$. Since $\overline{\delta} = 0$ minimizes $\tilde{\psi}^\infty(\delta)$, it follows that the first and second order necessary conditions hold. So, at the point $x^\infty$ the first and second order necessary conditions are satisfied ((15) holds), that is $s^T \nabla^2 \tilde{L}(x^\infty, \lambda_1^\infty, \lambda_2^\infty, ..., \lambda_p^\infty)s \ge 0$ , $\forall s \in \tilde{G}^\infty$,

where $\tilde{G}^* = \left\{ s : s^T \left( g^\infty + A^\infty \sum_{i=1}^{p} \max_{\lambda_i \in \partial h_i^\infty} \lambda_i \right) = 0, \|s\|_2 = 1 \right\}$   ∎

## 6. CONCLUSION

Different numerical procedures for evaluating the subgradient in the fifth step of the model algorithm could give different variants of the proposed algorithm. It will be interesting to apply this algorithm to the primal-dual model given for instance in [3].

## REFERENCES

[1]  Алексеев, В.М., Тихомиров, В.М., Фомин, С.В., *Оптимальное управление*, Наука, Москва, 1979.
[2]  Browien, J., and Lewis, A., *Convex Analysis and Nonlinear Optimization*, Canada, 1999.
[3]  Gertz, E.M., and Gill, P.E., "A primal-dual trust-region algorithm for nonlinear optimization*"*, *Math. Program*, *Ser B,* 100 (2004) 49-94.
[4]  Fletcher, R., *Practical Methods of Optimization, Vol. 1,* J. Wiley, New York, 1980.
[5]  Fletcher, R., *Practical Methods of Optimization*, *Vol. 2*, J. Wiley, New York, 1981.
[6]  Fuduli, A., "Metodi numerici per la minimizzazione di funzioni convesse nondifferenziabili", PhD Thesis, Departimento di Electronica Informatica e Sistemistica, Univerzita della Calabria, 1998.
[7]  Кусарев, А.Г., Кутателадзе, С.С., *Субдифференциальное исчисление*, Издатеьство Наука, Новосибирск, 1987.
[8]  Пшеничнй Б. Н., *Метод линеаризации*, Наука, Москва, 1983.