Neural, Parallel, and Scientific Computations 24 (2016) 381-392

HYPERBOLIC SMOOTHING METHOD FOR SUM-MAX PROBLEMS

ALI HAKAN TOR

Department of Mathematics,
Yüzüncü Yıl University, Van, Turkey

ABSTRACT. In this study, an approach for solving non-smooth optimization problem which includes sum of finite maximums of smooth functions is proposed. Minimum l_1-norm approximations is a particular case of this problem. In this approach, the problem is reformulated in order to use the hyperbolic smoothing function and the relationship between the original problem and reformulated problem are proved. This approach allows us to use conventional smooth optimization methods.

Key Words Nonsmooth optimization, sum-max problem, smoothing.

1. **Introduction**

In this study, we focus on the following non-smooth optimization problem

$$\begin{aligned} &\text{minimize} \quad f(x) + \sum_{i=1}^{m} \max[\alpha_i h_i(x), \beta_i h_i(x)] \\ &\text{subject to} \quad x \in \mathbb{R}^n \end{aligned} \tag{1.1}$$

where $f(x)$ and $h_i(x)$, $i = 1, \ldots, m$ are continuously differentiable. In this problem, we can assume $\alpha_i \neq \beta_i$ for all $i = 1, \ldots, m$. If $\alpha_i = \beta_i$ for some $i \in \{1, \ldots, m\}$, then instead of $f(x)$, which is the smooth part of the problem, we can write $f(x) + \sum_{i \in J} \alpha_i h_i(x)$ where $J = \{i | \alpha_i = \beta_i\}$.

Problem (1.1) arises Truss Topology Design in [5, 6] as cited in [18], the many important areas of modern signal/image processing, blind source separation and sparse decomposition [7, 19, 20]. On the other hand, for Problem (1.1), if α_i and β_i are chosen as -1 and 1 respectively and the functions $h_i(x)$ are affine, the problem turns:

$$\begin{aligned} &\text{minimize} \quad f(x) + \|h(x)\|_1 \\ &\text{subject to} \quad x \in \mathbb{R}^n \end{aligned} \tag{1.2}$$

where $h(x) := [h_1(x), \ldots, h_m(x)]^T$. This problem is known minimum l_1-norm approximations. In this study, we interest in the general case.

These problems are considered to be difficult because of the kinks which are introduced into the part $\sum_{i=1}^{m} \max[\alpha_i h_i(x), \beta_i h_i(x)]$ of the objective function in Problem 1.1 by the presence of the "max" operator. This situation is also valid for all

Received April 3, 2016

other problems containing “max” (or “min”) operators. In order to over come this difficulties, [17] introduce different functions to smooth the finite maximum function at the kink points. These kinks cause to fail to reach an optimum for classical algorithms because of the lack of derivatives [12, 14]. This difficulty motivated the development of smoothing algorithms for some type of nonsmooth problems, which could be minimized by using well known smooth algorithms.

Although conventional non-smooth optimization algorithms can be used in order to solve Problem (1.1), for instance variations of the bundle methods [2, 3, 9, 10, 11], smoothing methods are being studied nowadays to over come the drawback of non-smooth algorithms. For examples of the drawback of non-smooth algorithms, the implementation of non-smooth algorithm are generally very complex, and some of nonsmooth algorithms have poor convergency and need a very large amount of memory to retain the information on the computer during implementation. On the other hand, smoothing methods use the smooth algorithms and in smooth algorithms we have very strong tool “derivatives”. Because of the derivatives, smooth algorithms find easily descent direction and stopping criteria when comparing nonsmooth algorithms. Thus, smoothing method can be more useful than non-smoothing method for some type of optimization problems.

In this study, it is aimed to approximate the objective function via hyperbolic smooth functions. This approximation allows us to use well known smooth algorithms. On the other hand, although there are some similarities between the proposed method and the method in [1], this study is completely different from the method given in [1]. One can easily observe this differences from the structure of the problem.

The structure of this article is as follows. Some concepts which are used in this study are given in Section 2. In Section 3, Problem 1.1 is reformulated in order to use the some known facts of hyperbolic smoothing techniques. After that, hyperbolic function is given which approximate Problem 1.1 smoothly in Section 4. In the next Section 5, minimization algorithm to solve Problem 1.1 is proposed, which use the smooth function. Finally, the short conclusion is stated in Section 6.

2. **Preliminary**

In this section, the definitions of some concepts are given for people who are not familiar them. The following notations are used in this study. $\mathbb{R}^n$ is an n-dimensional Euclidean space with the inner product $\langle u, v\rangle = \sum_{i=1}^{n} u_i v_i$ and the associated Euclidean norm $\|\cdot\|$.

Clark gives the definition of subdifferential for Lipschitz continuous functions by using generalized directional derivative in [8]. After that, by using the fact that Lipschitz continuous functions are almost everywhere differentiable, the following set is given as a subdifferential of Lipschitz continuous functions, which more useful than the definition. For more detail please look [4, 8]. Let f be Lipschitz continuous function, its subdifferential is

$$\partial f(x) = \text{co}\Big\{ v \in \mathbb{R}^n \Big| \exists\, (x_k \in D(f), x_k \to x \text{ when } k \to \infty) \\ \text{such that } v = \lim_{k\to\infty} \nabla f(x_k) \Big\}$$

where $D(f)$ denotes the set where f is differentiable and co denotes the convex hull of a set.

Another concept used in this study is regular function. A locally Lipschitz continuous function f is called regular at a point x, if directional derivative $f'(x, d)$ exists and

$$f'(x,d) = \limsup_{y\to x,\, t\downarrow 0} \frac{f(y+td) - f(y)}{t} \tag{2.1}$$

holds. The right hand side of Equation 2.1 is also known as a generalized directional derivative of the function f at the point x in the direction d.

3. **Reformulation of Problem (1.1)**

The hyperbolic function for smoothing the function $\theta(x) = \max\{0, x\}$ is given in [13, 15, 16] as follows:

$$\phi_\tau(x) = \frac{x + \sqrt{x^2 + \tau^2}}{2}, \tag{3.1}$$

where $\tau > 0$ is a precision parameter.

The function $\phi_\tau(x)$ has the following properties as reported in [1]:

1. $\phi_\tau(x)$ is an increasing convex C^∞ function
2. $\theta(x) < \phi_\tau(x) < \theta(x) + \frac{\tau}{2}$

In order to use this fact, the strategy in the next section is applied to problem (1.1).

Since $f(x)$ is smooth, we focus on $\sum_{i=1}^m \max[\alpha_i h_i(x), \beta_i h_i(x)]$, which is the non-smooth part of problem (1.1). The non-smooth part of problem (1.1) can be reformulated as maxima of summation of some smooth function in the following way. Consider the following functions,

$$\sum_{i=1}^m (\mu_i \alpha_i h_i(x) + (\mu_i - 1)\beta_i h_i(x))$$

where either $\mu_i = 1$ or $\mu_i = 0$ for all $i = 1, \ldots, m$. In this way we define 2^m functions, and obviously maximum of these functions equal to the non-smooth part of Problem (1.1). In mathematically,

$$\max_{\mu_i} \sum_{i=1}^{m} (\mu_i \alpha_i h_i(x) + (1-\mu_i)\beta_i h_i(x)) = \sum_{i=1}^{m} \max[\alpha_i h_i(x)\ ,\ \beta_i h_i(x)] \tag{3.2}$$

where either $\mu_i = 1$ or $\mu_i = 0$ for all $i = 1, \dots, m$. In order to simplify the left hand side of Equation (3.2), we use the following notation

$$F_j(x) = \sum_{i=1}^{m} (\mu_i(j)\alpha_i h_i(x) + (1-\mu_i(j))\beta_i h_i(x)) \tag{3.3}$$

where $j = 1, \dots, 2^m$ and the relation between index j and $\mu_i(j)$'s for all $i = 1, \dots, m$ can be given in the following algorithm.

Assume the number m is known.

Algorithm 1. Algorithm to compute coefficients of F_j for all $j = 1, \dots, 2^m$.
Step 1 Set $j = 1$.
Step 2 If $j > 2^m$, then stop. Otherwise, set $k = 1$ and $j_1 = j$.
Step 3 If $k \leq m$, compute the remainder R_k and the quotient Q_k after division of j_k by 2 (i.e. $j_k = 2Q_k + R_k$). Otherwise, set $j = j + 1$ and go to Step 2
Step 4 Set $\mu_k(j) = R_k$, $j_{k+1} = Q_k$ and $k = k + 1$. Go to Step 3.

For instance, consider $m = 3$ and all functions F_j for all $j = 1, \dots, 8$ are as follows,

$$\begin{aligned}
F_1(x) &= \alpha_1 h_1(x) + \beta_2 h_2(x) + \beta_3 h_3(x)\\
F_2(x) &= \beta_1 h_1(x) + \alpha_2 h_2(x) + \beta_3 h_3(x)\\
F_3(x) &= \alpha_1 h_1(x) + \alpha_2 h_2(x) + \beta_3 h_3(x)\\
F_4(x) &= \beta_1 h_1(x) + \beta_2 h_2(x) + \alpha_3 h_3(x)\\
F_5(x) &= \alpha_1 h_1(x) + \beta_2 h_2(x) + \alpha_3 h_3(x)\\
F_6(x) &= \beta_1 h_1(x) + \alpha_2 h_2(x) + \alpha_3 h_3(x)\\
F_7(x) &= \alpha_1 h_1(x) + \alpha_2 h_2(x) + \alpha_3 h_3(x)\\
F_8(x) &= \beta_1 h_1(x) + \beta_2 h_2(x) + \beta_3 h_3(x).
\end{aligned}$$

Now, we can rewrite Problem (1.1) as follows,

$$\begin{array}{ll} \text{minimize} & f(x) + \max\limits_{j=1,\dots,2^m} \{F_j(x)\} \\ \text{subject to} & x \in \mathbb{R}^n, \end{array} \tag{3.4}$$

or equivalently

$$\begin{array}{ll} \text{minimize} & \max\limits_{j=1,\dots,2^m} \{f(x) + F_j(x)\} \\ \text{subject to} & x \in \mathbb{R}^n. \end{array} \tag{3.5}$$

In Problem (3.5), it can be clearly showed that $f(x) + F_j(x)$ are continuously differentiable fol all $j \in \{1, \dots, 2^m\}$.

Stationary point of Problem (3.5) are defined by using subdifferential as follows. At a point $x \in \mathbb{R}^n$ consider sets:

$$R(x) = \left\{ j \in \{1, \dots, 2^m\} \,\middle|\, \max_{j=1,\dots,2^m} \{f(x) + F_j(x)\} = f(x) + F_j(x) \right\},$$

and the subdifferential $\partial \left(\max_{j=1,\dots,2^m} \{f(x) + F_j(x)\} \right)$ at the point x is as follows;

$$\partial \left(\max_{j=1,\dots,2^m} \{f(x) + F_j(x)\} \right) = \mathrm{co} \{ \nabla f(x) + \nabla F_j(x) \,|\, j \in R(x) \}$$

where co denotes convex hull of a set. A point x^* is called a stationary point of problem (3.5) if and only if $0_n \in \partial \left(\max_{j=1,\dots,2^m} \{f(x^*) + F_j(x^*)\} \right)$.

In order to use equation (3.1), the objective function of problem (3.5) is reformulated. Consider the following function:

$$G(x,t) = t + \sum_{j \in J} \max \{0, f(x) + F_j(x) - t\} \tag{3.6}$$

where $J = \{1, \dots, 2^m\}$. For a given $(x,t) \in \mathbb{R}^n \times \mathbb{R}$, the index set J can be divided into tree distinct sets as follows

$$J = J_1 \cup J_2 \cup J_3$$

where

$$\begin{aligned} J_1 = J_1(x,t) &= \{j \in J \mid f(x) + F_j(x) < t\} \\ J_2 = J_2(x,t) &= \{j \in J \mid f(x) + F_j(x) = t\} \\ J_3 = J_3(x,t) &= \{j \in J \mid f(x) + F_j(x) > t\}. \end{aligned}$$

The relations between the function $\max_{j=1,\dots,2^m} \{f(x) + F_j(x)\}$ and $G(x,t)$ are stated in the following propositions.

Proposition 3.1. *Let $G(x,t)$ be defined as in* (3.6).

$$\max_{j \in J} \{f(x) + F_j(x)\} = \min_{t \in \mathbb{R}} G(x,t) \tag{3.7}$$

holds.

Proof. For any fixed point $x \in \mathbb{R}^n$, there are three cases

Case 1. $t > \max_{j \in J} \{f(x) + F_j(x)\}$
Case 2. $t = \max_{j \in J} \{f(x) + F_j(x)\}$
Case 3. $t < \max_{j \in J} \{f(x) + F_j(x)\}$

In Case 1, the index sets $J_2 = J_3 = \emptyset$, so $G(x,t) = t > \max_{j\in J}\{f(x) + F_j(x)\}$. In the next case, the index set $J_3 = \emptyset$, so $G(x,t) = t = \max_{j\in J}\{f(x) + F_j(x)\}$. In the last case $J_3 \neq \emptyset$,

$$G(x,t) = t + \sum_{j\in J_3}(f(x) + F_j(x) - t). \tag{3.8}$$

Let $j_0 = \operatorname{argmax}\{f(x) + F_j(x)\}$, that is for the index j_0, $\max_{j\in J}\{f(x) + F_j(x)\} = f(x) + F_{j_0}(x)$. Here, j_0 may not be unique, in this case we take just one of them as a j_0. Thus,

$$G(x,t) = f(x) + F_{j_0}(x) + \sum_{j\in J_3\setminus\{j_0\}}(f(x) + F_j(x) - t).$$

Obviously, $\sum_{j\in J_3\setminus\{j_0\}}(f(x) + F_j(x) - t) \geq 0$, so $G(x,t) \geq \max_{j\in J}\{f(x) + F_j(x)\}$.

Consequently, $G(x,t)$ greater and equal to $\max_{j\in J}\{f(x) + F_j(x)\}$ for all t and reaches to the value $\max_{j\in J}\{f(x) + F_j(x)\}$ at the point $t = \max_{j\in J}\{f(x) + F_j(x)\}$. In other words, $\min_{t\in\mathbb{R}} G(x,t) = \max_{j\in J}\{f(x) + F_j(x)\}$. □

Proposition 3.2. *(1) If a point $x^* \in \mathbb{R}^n$ is a stationary point of the function $\max_{j\in J}\{f(x) + F_j(x)\}$, then the point $(x^*, t^*) \in \mathbb{R}^{n+1}$ is a stationary point of the function $G(x,t)$ where $t^* = \max_{j\in J}\{f(x^*) + F_j(x^*)\}$.*

(2) If a point $(x^*, t^*) \in \mathbb{R}^{n+1}$ is a stationary point of the function $G(x,t)$, then $x^* \in \mathbb{R}^n$ is a stationary point of the function $\max_{j=1,\dots,2^m}\{f(x) + F_j(x)\}$.

Proof. The proof can be found in [1]. □

Remark 3.3. When observing the proof of Proposition 3.2, which is given in [1], if (x^*, t^*) is a stationary point of the function $G(x,t)$, then either

$$t^* = \max_{j\in J}\{f(x^*) + F_j(x^*)\} \text{ or } t^* < \max_{j\in J}\{f(x^*) + F_j(x^*)\}$$

holds.

Moreover, from the proof we can easily say $J_3(x^*, t^*)$ is a singleton in the latter case. This implies $R(x^*)$ is also a singleton set, so the function $\max_{j\in J}\{f(x) + F_j(x)\}$ is differentiable at the point x^*. On the other hand, it is known that stationary points of minimax problems are generally non-differentiable points. Therefore, for the most of min-max problems $t^* = \max_{j\in J}\{f(x^*) + F_j(x^*)\}$. In other words, the stationary points of $G(x,t)$ is in the form $(x^*, \max_{j\in J}\{f(x^*) + F_j(x^*)\}$.

The following proposition presents the relation of local minimizer between the function $\max_{j\in J}\{f(x) + F_j(x)\}$ and the function $G(x,t)$.

Proposition 3.4. *(1) If a point $x^* \in \mathbb{R}^n$ is a local minimizer of the function $\max_{j\in J}\{f(x)+F_j(x)\}$, then the point $(x^*,t^*) \in \mathbb{R}^{n+1}$ is a local minimizer of the function $G(x,t)$ where $t^* = \max_{j\in J}\{f(x^*)+F_j(x^*)\}$.*

(2) If a point $(x^*,t^*) \in \mathbb{R}^{n+1}$ is a local minimizer of the function $G(x,t)$, then $x^* \in \mathbb{R}^n$ is a local minimize of the function $\max_{j\in J}\{f(x)+F_j(x)\}$.

Proof. The proof can be found in [1]. □

4. **A Hyperbolic Smoothing Function**

In this section, a hyperbolic smoothing function of the objective function in problem (3.5) is given. After that, the relationships of problem (1.1) and (3.5) is stated. By combining the equations (3.1) and (3.6), it is obtained that

$$\Phi_\tau(x,t) = t + \sum_{j\in J} \frac{f(x)+F_j(x)-t+\sqrt{\left(f(x)+F_j(x)-t\right)^2+\tau^2}}{2} \tag{4.1}$$

in the sense of $0 < \Phi_\tau(x,t) - G(x,t) < 2^{m-1}\tau$ as the hyperbolic smoothing of the function $\max_{j\in J}\{f(x)+F_j(x)\}$.

The gradient of the function $\Phi_\tau(x,t)$ is as follows:

$$\nabla\Phi_\tau(x,t) = (G_{1\tau}(x,t), G_{2\tau}(x,t))$$

where

$$G_{1\tau}(x,t) = \frac{1}{2}\sum_{j\in J}\left(1+\Gamma_{j\tau}(x,t)\right)\nabla\left(f(x)+F_j(x)\right),$$

$$G_{2\tau}(x,t) = 1 - 2^{m-1} - \frac{1}{2}\sum_{j\in J}\Gamma_{j\tau}(x,t),$$

$$\Gamma_{j\tau}(x,t) = \frac{f(x)+F_j(x)-t}{\sqrt{(f(x)+F_j(x)-t)^2+\tau^2}}.$$

In this gradient, the first and second components represent derivatives with respect to the variable x and t respectively.

Before stating the following proposition, it is necessary to give the subdifferential $\partial G(x,t)$. Denote $\Psi_j(x,t) = \max\{0, f(x)-F_j(x)-t\}$ for all $j \in J$. Then, the subdifferential of $\Psi_j(x,t)$ for fixed j is following;

$$\partial\left(\Psi_j(x,t)\right) = \begin{cases} \{0_{n+1}\}, & j \in J_1 \\ \mathrm{co}\{0_{n+1}, (\nabla f(x)+\nabla F_j(x), -1)\}, & j \in J_2 \\ \{(\nabla f(x)+\nabla F_j(x), -1)\}, & j \in J_3 \end{cases}$$

Since $f(x)$ and F_j for all $j \in J$ are regular, Ψ_j for all $j \in J$ are regular too. Thus, the subdifferential of $\partial G(x,t)$ can be expressed in the following way;

$$\partial G(x,t) = \{(0_n, 1)\} + \sum_{j \in J_2} \operatorname{co}\{0_{n+1}, (\nabla f(x) + \nabla F_j(x), -1)\} + \sum_{j \in J_3} \{(\nabla f(x) + \nabla F_j(x), -1)\}. \tag{4.2}$$

Proposition 4.1. *Assume that $v = \lim_{\tau \to 0} \nabla \Phi_\tau(x,t)$, then $v \in \partial G(x,t)$.*

Proof. It is clear that for any $x \in \mathbb{R}^{n+1}$ and $t \in \mathbb{R}$,

$$\lim_{\tau \to 0} \Gamma_{i\tau}(x,t) = \begin{cases} -1, & i \in J_1, \\ 0, & i \in J_2, \\ 1, & i \in J_3. \end{cases}$$

Then, we have

$$\lim_{\tau \to 0} G_{1\tau}(x,t) = \sum_{j \in J_2} \frac{1}{2} \nabla \left(f(x) + F_j(x)\right) + \sum_{j \in J_3} \nabla \left(f(x) + F_j(x)\right)$$

and

$$\begin{aligned} \lim_{\tau \to 0} G_{2\tau}(x,t) &= 1 - 2^{m-1} + \frac{1}{2}\left(|J_1| - |J_3|\right) \\ &= 1 - \frac{1}{2}|J_2| - |J_3|. \end{aligned}$$

Therefore,

$$v = \lim_{\tau \to 0} \nabla \Phi_\tau(x,t) = \left(\sum_{j \in J_2} \frac{1}{2} \nabla \left(f(x) + F_j(x)\right) + \sum_{j \in J_3} \nabla \left(f(x) + F_j(x)\right), 1 - \frac{1}{2}|J_2| - |J_3| \right).$$

We can arrange

$$v = (0_n, 1) + \sum_{j \in J_2} \frac{1}{2} \left(\nabla \left(f(x) + F_j(x)\right), -1\right) + \sum_{j \in J_3} \left(\nabla \left(f(x) + F_j(x)\right), -1\right).$$

Obviously, $v \in \partial G(x,t)$ from equation (4.2). □

Proposition 4.2. *Suppose that sequences $\{x_k\}$, $\{t_k\}$ and $\{\tau_k\}$ are given such that $x_k \in \mathbb{R}^n$, $t_k \in \mathbb{R}$, $t_k \geq f(x_k)$ and $\tau_k > 0$, $k = 1, 2, \ldots$. Moreover, $x_k \to x$, $t_k \to t$, $\tau_k \to 0$ when $k \to \infty$ and*

$$v = \lim_{k \to \infty} \nabla \Phi_{\tau_k}(x_k, t_k).$$

Then $v \in \partial G(x,t)$.

Proof. Since $f(x) + F_j(x) - t$ are continuously differentiable (i.e. $F_i(x) - t \in C^1$) for all $i = 1, \ldots, m$, then the function Φ_τ is continuously differentiable. This means $\nabla \Phi_{\tau_k}$ is continuous, by using aforementioned convergences in the statement of Proposition 4.2, we have

$$v = \lim_{k \to \infty} \nabla \Phi_{\tau_k}(x_k, t_k) = \lim_{k \to \infty} \nabla \Phi_{\tau_k}(x,t).$$

On the other hand, since $\tau_k \to 0$ as $k \to \infty$,

$$v = \lim_{\tau \to 0} \nabla \Phi_\tau(x,t).$$

Using Proposition 4.1, we can conclude that $v \in \partial G(x,t)$. □

5. **Minimization Algorithm**

In this section, an algorithm is given to find the solution of problem (1.1) by using the function in (4.1) via a smooth optimization solver. Replacing problem (1.1) by the sequence of the following smooth problems:

$$\begin{aligned} &\text{minimize} \quad \Phi_{\tau_k}\left(x, \max_{j\in J}\{f(x)+F_j(x)\}\right) \\ &\text{subject to} \quad x \in \mathbb{R}^n, \end{aligned} \tag{5.1}$$

where $\tau_k \to 0$ as $k \to 0$. In light of the Section (4), smooth optimization solver can solve problem (5.1) and its minimizer can be used to find the minimizer of problem (1.1).

Remark 5.1. In Problem (5.1), the precision parameter can be chosen sufficiently small $\tau > 0$ instead of the sequence $\{\tau_k\}$. However, this approach may cause more computational efforts. The usage of the sequence $\{\tau_k\}$ may help get rid of this phenomenon.

The following algorithm is proposed for solving problem (1.1). Assume the sequences $\{\tau_k\}$, $\{\epsilon_k\}$ are given such that τ_k, $\epsilon_k > 0$ and τ_k, $\epsilon_k \to 0$ as $k \to \infty$.

Algorithm 2. Algorithm to solve problem (1.1).
Step 1 Apply Algorithm 1 to determine F_j for all $j = 1, \ldots, 2^m$.
Step 2 Select any starting point $x_0 \in \mathbb{R}^n$, set $k := 0$ and compute

$$t_0 = \max_{j\in J}\{f(x_0)+F_j(x_0)\}.$$

Step 3 Using a smooth optimization solver, solve the problem (5.1) by starting from the point x_k such that for the solution $\bar{x}$

$$\left\|\nabla\Phi_{\tau_k}\left(\bar{x}, \max_{j\in J}\{f(\bar{x})+F_j(\bar{x})\}\right)\right\| < \epsilon_k \tag{5.2}$$

holds.
Step 4 Set $x_{k+1} = \bar{x}$, $t_{k+1} = \max_{j\in J}\{f(\bar{x})+F_j(\bar{x})\}$, $k := k+1$ and go to Step 3.

Remark 5.2. In Algorithm 2, the choice of sequences $\{\tau_k\}$, $\{\epsilon_k\}$ has important role on the behavior of the problem. If the sequence $\{\tau_k\}$ converges to 0 quickly, then the problem get ill-conditioned behavior. In addition, a large number of iteration may be needed to satisfy the condition (5.2). In order to overcome this situation, the convergency of the sequence $\{\tau_k\}$ should be slower than the sequence $\{\epsilon_k\}$.

Now, we give a proposition which guarantees that Algorithm 2 terminates finitely many steps. First, the following set which is used in next proposition is defined:

$$\mathcal{L}(x_0) = \left\{ x \in \mathbb{R}^n \,\middle|\, \max_{j\in J} \{f(x) - F_j(x)\} \leq \max_{j\in J} \{f(x_0) - F_j(x_0)\} \right\}.$$

Proposition 5.3. *Assume that the set $\mathcal{L}(x_0)$ is bounded for any starting point $x_0 \in \mathbb{R}^n$. Then any accumulation point of the sequence $\{x_k\}$ generated by Algorithm 2 is a stationary point of problem (1.1).*

Proof. Since all point x_k generated by Algorithm 2 are solution of the sequence of the smooth problems (5.1), $x_k \in \mathcal{L}(x_0)$ for all k. On the other hand, $\mathcal{L}(x_0)$ is bounded by the assumption, so the sequence $\{x_k\}$ has at least one accumulation point. Let say this accumulation point is x^*. In other words, assume $x_k \to x^*$ as $k \to \infty$. According to Step 2 of Algorithm 2, consider $t^* = \max_{j\in J} \{f(x^*) + F_j(x^*)\}$. Since the sequence $\{\epsilon_k\}$ converges to 0, $0_{n+1} = \Phi_{\tau_k}\left(x^*, \max_{j\in J} \{f(x^*) + F_j(x^*)\}\right)$. By Proposition 4.2, $0_{n+1} \in \partial G(x,t)$, that is (x^*, t^*) is a stationary point of $G(x,t)$. Consequently, by applying Proposition 3.2, we have that x^* is a stationary point of problem (1.1). □

As a result, by using Algorithm 2 with any conventional smooth solver, the main problem 1.1 can be solved.

6. Conclusions

In this study, a new approach is proposed to solve the non-smooth optimization problem which is written as a sum of smooth function and the sum of maximum of smooth functions. Because of this approach, one can use well-known optimization solver for smooth problems instead of non-smooth solvers. It means that it is possible to avoid some known drawbacks of non-smooth solvers, which are mentioned in the introduction. On the other hand, this strategy can be used for aforementioned areas of Mathematics and Engineering.

Acknowledgment

I would like to thank to professor Adil Bagirov and referees for their valuable suggestions and ideas to improve the quality of this study.

REFERENCES

[1] A. M. Bagirov, A. Al Nuaimat, and N. Sultanova. Hyperbolic smoothing function method for minimax problems. *Optimization*, 62(6):759–782, 2013.

[2] A. M. Bagirov and A. N. Ganjehlou. A quasisecant method for minimizing nonsmooth functions. *Optim. Methods Softw.*, 25(1):3–18, 2010.

[3] A. M. Bagirov, B. Karasözen, and M. Sezer. Discrete gradient method: derivative-free method for nonsmooth optimization. *J. Optim. Theory Appl.*, 137(2):317–334, 2008.

[4] Adil Bagirov, Napsu Karmitsa, and Marko M. Mäkelä. *Introduction to nonsmooth optimization*. Springer, Cham, 2014. Theory, practice and software.

[5] Aharon Ben-Tal and Martin P. Bendso/e. A new method for optimal truss topology design. *SIAM Journal on Optimization*, 3(2):322–37, 05 1993.

[6] M. P. Bendsøe, A. Ben-Tal, and J. Zowe. Optimization methods for truss geometry and topology design. *Structural optimization*, 7(3):141–159, 1994.

[7] S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1998.

[8] F. H. Clarke. *Optimization and nonsmooth analysis*. Canadian Mathematical Society Series of Monographs and Advanced Texts. John Wiley & Sons Inc., New York, 1983. A Wiley-Interscience Publication.

[9] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms. II*, volume 306 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993. Advanced theory and bundle methods.

[10] K. C. Kiwiel. *Methods of descent for nondifferentiable optimization*, volume 1133 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1985.

[11] M. M. Mäkelä and P. Neittaanmäki. *Nonsmooth optimization*. World Scientific Publishing Co. Inc., River Edge, NJ, 1992. Analysis and algorithms with applications to optimal control.

[12] M. J. D. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Comput. J.*, 7:155–162, 1964.

[13] S. Smale. Algorithms for solving equations. In *Proceedings of the International Congress of Mathematicians*, pages 172–195. Berkeley, CA, USA, 1986.

[14] G. W. Stewart, III. A modification of davidon's minimization method to accept difference approximations of derivatives. *J. ACM*, 14(1):72–83, January 1967.

[15] A. E. Xavier. The hyperbolic smoothing clustering method. *Pattern Recognition*, 43(3):731–737, 2010.

[16] A. E. Xavier and A. A. F. de Oliveira. Optimal covering of plane domains by circles via hyperbolic smoothing. *J. Global Optim.*, 31(3):493–504, 2005.

[17] Israel Zang. A smoothing-out technique for min—max optimization. *Mathematical Programming*, 19(1):61–77, 1980.

[18] M. Zibulevsky. Smoothing method of multipliers for sum-max problems. In *Online]. Available: http://iew3.technion.ac.il/∼mcib*, 2002.

[19] M. Zibulevsky and Yehoshua Y. Zeevi. Blind source separation by sparse decomposition in a signal dictionary. *Neural Computation*, 13(4):863–882, 2001.

[20] M. Zibulevsky and Yehoshua Y. Zeevi. Extraction of a source from multichannel data using sparse decomposition. *Neurocomputing*, 49(1-4):163–173, 2002.