

# A Smoothing Penalty Function Method for the Constrained Optimization Problem

Bingzhuang Liu

School of Mathematics and Statistics, Shandong University of Technology, Shandong, China

Email: lbzlyj@126.com

**How to cite this paper:** Liu, B.Z. (2019) A Smoothing Penalty Function Method for the Constrained Optimization Problem. *Open Journal of Optimization*, 8, 113-126  
<https://doi.org/10.4236/ojop.2019.84010>

**Received:** October 3, 2019

**Accepted:** December 1, 2019

**Published:** December 4, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

In this paper, an approximate smoothing approach to the non-differentiable exact penalty function is proposed for the constrained optimization problem. A simple smoothed penalty algorithm is given, and its convergence is discussed. A practical algorithm to compute approximate optimal solution is given as well as computational experiments to demonstrate its efficiency.

## Keywords

Constrained Optimization, Penalty Function, Smoothing Method, Optimal Solution

## 1. Introduction

Many problems in industry design, management science and economics can be modeled as the following constrained optimization problem:

$$(P) \quad \begin{array}{ll} \min & f(x) \\ \text{s.t.} & g_j(x) \leq 0, j = 1, 2, \dots, m, \end{array} \quad (1)$$

where  $f, g_j : \mathcal{R}^n \rightarrow \mathcal{R}, j = 1, 2, \dots, m$  are continuously differentiable functions.

Let  $\mathcal{F}_0$  be the feasible solution set, that is,

$\mathcal{F}_0 = \{x \in \mathcal{R}^n \mid g_j(x) \leq 0, j = 1, 2, \dots, m\}$ . Here we assume that  $\mathcal{F}_0$  is nonempty.

The penalty function methods based on various penalty functions have been proposed to solve problem (P) in the literatures. One of the popular penalty functions is the quadratic penalty function with the form

$$F_2(x, \rho) = f(x) + \rho \sum_{j=1}^m \max\{g_j(x), 0\}^2, \quad (2)$$

where  $\rho > 0$  is a penalty parameter. Clearly,  $F_2(x, \rho)$  is continuously differentiable, but is not an exact penalty function. In Zangwill [1], an exact penalty

function was defined as

$$F_1(x, \rho) = f(x) + \rho \sum_{j=1}^m \max\{g_j(x), 0\}. \quad (3)$$

The corresponding penalty problem is

$$\begin{aligned} (P_\rho) \quad & \min F_1(x, \rho) \\ & \text{s.t. } x \in \mathfrak{R}^n. \end{aligned} \quad (4)$$

We say that  $F_1(x, \rho)$  is an exact penalty function for Problem  $(P)$  partly because it satisfies one of the main characteristics of *exactness*, that is, under some constraint qualifications, there exists a sufficiently large  $\rho^*$  such that for each  $\rho > \rho^*$ , the optimal solutions of Problem  $(P_\rho)$  are all the feasible solutions of Problem  $(P)$ , therefore, they are all the optimal solution of  $(P)$  (Di Pillo [2], Han [3]).

The obvious difficulty with the exact penalty functions is that it is nondifferentiable, which prevents the use of efficient minimization methods that are based on Gradient-type or Newton-type algorithms, and may cause some numerical instability problems in its implementation. In practice, an approximately optimal solution to  $(P)$  is often only needed. Differentiable approximations to the exact penalty function have been obtained in different contexts such as in BeaTal and Teboulle [4], Herty *et al.* [5] and Pinar and Zenios [6]. Penalty methods based on functions of this class were studied by Auslender, Cominetti and Haddou [7] for convex and linear programming problems, and by Gonzaga and Castillo [8] for nonlinear inequality constrained optimization problems, respectively. In Xu *et al.* [9] and Lian [10], smoothing penalty functions are proposed for nonlinear inequality constrained optimization problems. This kind of functions is also described by Chen and Mangasarian [11] who constructed them by integrating probability distributions to study complementarity problems, by Herty *et al.* [5] to study the optimization problems with box and equality constraints, and by Wu *et al.* [12] to study global optimization problem. Meng *et al.* [13] propose two smoothing penalty functions to the exact penalty function

$$F_3(x, \rho) = f(x) + \rho \sum_{i=1}^m \sqrt{\max\{g_j(x), 0\}}. \quad (5)$$

In Wu *et al.* [14] and Lian [15], some smoothing techniques for (5) are also given.

Moreover, smoothed penalty methods can be applied to solve optimization problems with large scale such as network-structured problems and minimax problems in [6], and traffic flow network models in [5].

In this paper, we consider another simpler method for smoothing the exact penalty function  $F_1(x, \rho)$ , and construct the corresponding smoothed penalty problem. We show that our smooth penalty function can approximate  $F_1(x, \rho)$  well and has better smoothness. Based on our smooth penalty function, we give for  $(P)$  a simple smoothed penalty algorithm which is different from the existing literature in that the convergence of it can be obtained without the compactness

of the feasible region of  $(P)$ . We also give an approximate algorithm which enjoys some convergence under mild conditions.

The rest of this paper is organized as follows. In Section 2, we propose a method for smoothing the  $l_1$  exact penalty function (3). The approximation function we give is convex and smooth. We give some error estimates among the optimal objective function values of the smoothed penalty problem, of the non-smooth penalty problem and of the original constrained optimization problem. In Section 3, we present an algorithm to compute a solution to  $(P)$  based on our smooth penalty function and show the convergence of the algorithm. In particular, we give an approximate algorithm. Some computational aspects are discussed and some experiment results are given in Section 4.

## 2. A Smooth Penalty Function

We define a function  $P_{\rho\varepsilon}(t)$ :

$$P_{\rho\varepsilon}(t) = \begin{cases} \frac{1}{2}\varepsilon e^{\frac{\rho t}{\varepsilon}}, & \text{if } t \leq 0; \\ \rho t + \frac{1}{2}\varepsilon e^{-\frac{\rho t}{\varepsilon}}, & \text{if } t > 0, \end{cases} \quad (6)$$

given any  $\varepsilon > 0, \rho > 0$ .

Let  $P_{\rho}(t) = \rho \max\{t, 0\}$ , for any  $\rho > 0$ . It is easy to show that

$$\lim_{\varepsilon \rightarrow 0} P_{\rho\varepsilon}(t) = P_{\rho}(t).$$

The function  $P_{\rho\varepsilon}(t)$  is different from the function  $P_{\varepsilon}(t)$  given in [16] since here we use two parameters  $\rho$  and  $\varepsilon$ . The function  $P_{\rho\varepsilon}(t)$  has the following abstractive properties.

(I)  $P_{\rho\varepsilon}(t)$  is at least twice continuously differentiable in  $t$  for  $\forall \varepsilon > 0, \rho > 0$ . In fact, we have that

$$P'_{\rho\varepsilon} = \begin{cases} \frac{1}{2}\rho e^{\frac{\rho t}{\varepsilon}}, & \text{if } t \leq 0; \\ \rho - \frac{1}{2}\rho e^{-\frac{\rho t}{\varepsilon}}, & \text{if } t > 0, \end{cases} \quad (7)$$

and

$$P''_{\rho\varepsilon}(t) = \begin{cases} \frac{1}{2\varepsilon}\rho^2 e^{\frac{\rho t}{\varepsilon}}, & \text{if } t \leq 0; \\ \frac{1}{2\varepsilon}\rho^2 e^{-\frac{\rho t}{\varepsilon}}, & \text{if } t > 0. \end{cases} \quad (8)$$

(II)  $P_{\rho\varepsilon}(t)$  is convex and monotonically increasing in  $t$  for any given  $\varepsilon > 0, \rho > 0$ .

Property (II) can follow from (I) immediately.

Note that  $P_{\rho}(g_j(x)) = \rho \max\{0, g_j(x)\}$ ,  $j = 1, 2, \dots, m$ . Consider the penalty function for  $(P)$  given by

$$F(x, \rho, \varepsilon) = f(x) + \sum_{j=1}^m P_{\rho\varepsilon}(g_j(x)), \quad (9)$$

where  $\rho > 0$  is a penalty parameter. Clearly,  $F(x, \rho, \varepsilon)$  is at least twice continuously differentiable in any  $x \in \mathfrak{R}^n$ , if  $f, g_j, j=1, 2, \dots, m$  are all at least twice continuously differentiable, and  $F(x, \rho, \varepsilon)$  is convex, if  $f, g_j, j=1, 2, \dots, m$  are all convex functions.

The corresponding penalty problem to  $F(x, \rho, \varepsilon)$  is given as follows:

$$(P_{\rho, \varepsilon}) \quad \min_{x \in \mathfrak{R}^n} F(x, \rho, \varepsilon).$$

Since  $\lim_{\varepsilon \rightarrow 0} F(x, \rho, \varepsilon) = F_1(x, \rho)$  for any given  $\rho$ , we will first study the relationship between  $(P_\rho)$  and  $(P_{\rho, \varepsilon})$ .

The following Lemma is easily to prove.

**Lemma 2.1** For any given  $x \in \mathfrak{R}^n$ , and  $\varepsilon > 0, \rho > 0$ ,

$$0 \leq F(x, \rho, \varepsilon) - F_1(x, \rho) \leq \frac{1}{2} m \varepsilon.$$

Two direct results of Lemma 2.1 are given as follows.

**Theorem 2.1** Let  $\{\varepsilon_j\} \rightarrow 0$  be a sequence of positive numbers and assume  $x^j$  is a solution to  $\min_{x \in \mathfrak{R}^n} F(x, \rho, \varepsilon_j)$  for some given  $\rho > 0$ . Let  $\bar{x}$  be an accumulating point of the sequence  $\{x^j\}$ , then  $\bar{x}$  is an optimal solution to  $\min_{x \in \mathfrak{R}^n} F_1(x, \rho)$ .

**Theorem 2.2** Let  $x^*$  be an optimal solution of  $(P_\rho)$  and  $\bar{x} \in \mathfrak{R}^n$  an optimal solution of  $(P_{\rho, \varepsilon})$ . Then

$$0 \leq F(\bar{x}, \rho, \varepsilon) - F_1(x^*, \rho) \leq \frac{1}{2} m \varepsilon.$$

It follows from this conclusion that  $F(x, \rho, \varepsilon)$  can approximate  $F_1(x, \rho)$  well.

Theorem 2.1 and Theorem 2.2 show that an approximate solution to  $(P_{\rho, \varepsilon})$  is also an approximate solution to  $(P_\rho)$  when  $\varepsilon$  is sufficiently small.

**Definition 2.1** A point  $x_\delta \in \mathfrak{R}^n$  is a  $\delta$ -feasible solution or a  $\delta$ -solution if,

$$g_j(x_\delta) \leq \delta, j = 1, \dots, m.$$

Under this definition, we get the following result.

**Theorem 2.3** Let  $x^*$  be an optimal solution of  $(P_\rho)$  and  $\bar{x} \in \mathfrak{R}^n$  an optimal solution of  $(P_{\rho, \varepsilon})$ . Furthermore, let  $x^*$  be feasible to  $(P)$  and  $\bar{x}$  be  $\delta$ -feasible to  $(P)$ . Then,

$$-m\rho\delta - \frac{1}{2} m \varepsilon \leq f(\bar{x}) - f(x^*) \leq \frac{1}{2} m \varepsilon.$$

*Proof* Since  $\bar{x}$  is  $\delta$ -feasible to  $(P)$ , then

$$\begin{aligned} 0 &\leq \sum_{j=1}^m P_{\rho\varepsilon}(g_j(\bar{x})) \\ &= \sum_{(j: g_j(\bar{x}) \leq 0)} \frac{1}{2} \varepsilon e^{\frac{\rho g_j(\bar{x})}{\varepsilon}} + \sum_{(j: 0 < g_j(\bar{x}) \leq \delta)} \left( \rho g_j(\bar{x}) + \frac{1}{2} \varepsilon e^{\frac{\rho g_j(\bar{x})}{\varepsilon}} \right) \\ &\leq \sum_{(j: g_j(\bar{x}) \leq 0)} \frac{1}{2} \varepsilon + \sum_{(j: 0 < g_j(\bar{x}) \leq \delta)} \left( \rho \delta + \frac{1}{2} \varepsilon e^{\frac{\rho \delta}{\varepsilon}} \right) \end{aligned}$$

$$\begin{aligned} &\leq m \max \left\{ \frac{1}{2} \varepsilon, \rho \delta + \frac{1}{2} \varepsilon e^{-\frac{\rho \delta}{\varepsilon}} \right\} \\ &\leq m \rho \delta + \frac{1}{2} m \varepsilon. \end{aligned} \quad (10)$$

Since  $x^*$  is an optimal solution to  $(P)$ , we have

$$\sum_{j=1}^m P_{\rho} (g_j (x^*)) = 0.$$

Then by Theorem 2.2, we get

$$0 \leq f(\bar{x}) + \sum_{j=1}^m P_{\rho\varepsilon} (g_j(\bar{x})) - \left( f(x^*) + \sum_{j=1}^m P_{\rho} (g_j(x^*)) \right) \leq \frac{1}{2} m \varepsilon.$$

Thus,

$$-\sum_{j=1}^m P_{\rho\varepsilon} (g_j(\bar{x})) \leq f(\bar{x}) - f(x^*) \leq -\sum_{j=1}^m P_{\rho\varepsilon} (g_j(\bar{x})) + \frac{1}{2} m \varepsilon.$$

Therefore, by (10), we obtain that

$$-m \rho \delta - \frac{1}{2} m \varepsilon \leq f(\bar{x}) - f(x^*) \leq \frac{1}{2} m \varepsilon.$$

This completes the proof.

By Theorem 2.3, if an approximate optimal solution of  $(P_{\rho,\varepsilon})$  is  $\delta$ -feasible, then it is an approximate optimal solution of  $(P)$ .

For  $l_1$  penalty function  $F_1(x, \rho)$ , there is a well known result of its exactness (see [3]):

(\*) *There exists a  $\rho^* > 0$ , such that whenever  $\rho \geq \rho^*$ , each optimal solution of  $F_1(x, \rho)$  is also an optimal solution of  $(P)$ .*

From the above conclusion, we can get the following result.

**Theorem 2.4** *For the constant  $\rho^* > 0$  in (\*), let  $x^*$  be an optimal solution of  $(P_{\rho^*})$ . Suppose that for any  $\varepsilon > 0$ ,  $\tilde{x}_{\varepsilon}$  is an optimal solution of  $(P_{\rho,\varepsilon})$  where  $\rho \geq \rho^*$ , then  $\tilde{x}_{\varepsilon}$  is a  $\varepsilon$ -feasible solution of  $(P)$ .*

*Proof* Suppose the contrary that the theorem does not hold, then there exists a  $\varepsilon_0 > 0$ , and  $\rho_0 \geq \rho^* + m$ , such that  $\tilde{x}_0$  is an optimal solution for  $(P_{\rho_0,\varepsilon_0})$ , and the set  $\{j = 1, \dots, m \mid g_j(\tilde{x}_0) > \varepsilon_0\}$  is not empty.

Since  $\tilde{x}_0$  is an optimal solution for  $(P_{\rho_0,\varepsilon_0})$  when  $\varepsilon = \varepsilon_0$ , then for any  $x \in \mathfrak{R}^n$ , it holds that

$$F(\tilde{x}_0, \rho_0, \varepsilon_0) \leq F(x, \rho_0, \varepsilon_0). \quad (11)$$

Because  $x^*$  is an optimal solution of  $(P_{\rho^*})$ ,  $x^*$  is a feasible solution of  $(P)$ . Therefore, we have that

$$\sum_{j=1}^m P_{\rho_0\varepsilon_0} (g_j(x^*)) \leq \frac{1}{2} m \varepsilon_0.$$

On the other side,

$$\begin{aligned}
 F(\tilde{x}_0, \rho_0, \varepsilon_0) &= f(\tilde{x}_0) + \sum_{j=1}^m P_{\rho_0 \varepsilon_0} (g_j(\tilde{x}_0)) \\
 &= f(\tilde{x}_0) + \sum_{(j: g_j(\tilde{x}_0) \leq 0)} \frac{1}{2} \varepsilon_0 e^{\frac{\rho_0 g_j(\tilde{x}_0)}{\varepsilon_0}} \\
 &\quad + \sum_{(j: g_j(\tilde{x}_0) > 0)} \left( \rho g_j(\tilde{x}_0) + \frac{1}{2} \varepsilon_0 e^{-\frac{\rho_0 g_j(\tilde{x}_0)}{\varepsilon_0}} \right) \\
 &> f(\tilde{x}_0) + \rho_0 \sum_{j=1}^m g_j^+(\tilde{x}_0) \\
 &\geq f(\tilde{x}_0) + \rho^* \sum_{j=1}^m g_j^+(\tilde{x}_0) + m\varepsilon_0 \\
 &\geq f(x^*) + \rho^* \sum_{j=1}^m g_j^+(x^*) + m\varepsilon_0 \\
 &= f(x^*) + m\varepsilon_0 \\
 &\geq f(x^*) + \sum_{j=1}^m P_{\rho_0 \varepsilon_0} (g_j(x^*)) + \frac{1}{2} m\varepsilon_0 \\
 &= F(x^*, \rho_0, \varepsilon_0) + \frac{1}{2} m\varepsilon_0,
 \end{aligned}$$

which contradicts (11).

Theorem 2.4 implies that any optimal solution of the smoothed penalty problem is an approximately feasible solution of (P).

### 3. The Smoothed Penalty Algorithm

In this section, we give an algorithm based on the smoothed penalty function given in Section 2 to solve the nonlinear programming problem (P).

For  $x \in R^n$ , we denote

$$\begin{aligned}
 J^0(x) &= \{j \mid g_j(x) = 0, j \in I\}, \\
 J^-(x) &= \{j \mid g_j(x) < 0, j \in I\}, \\
 J^+(x) &= \{j \mid g_j(x) > 0, j \in I\}, \\
 J_\varepsilon^>(x) &= \{j \mid g_j(x) > \varepsilon, j \in I\}, \\
 J_\varepsilon^\leq(x) &= \{j \mid g_j(x) \leq \varepsilon, j \in I\},
 \end{aligned}$$

For Problem (P), let  $\mathcal{F}_\delta := \{x \in \mathfrak{R}^n \mid g_j(x) \leq \delta, j = 1, \dots, m\}$ , for some  $\delta > 0$ . We consider the following algorithm.

**Algorithm 3.1**

*Step 1.* Given  $x^0, \varepsilon_1 > 0$ , and  $\rho_1 > 0$ , let  $k = 1$ , go to Step 2.

*Step 2.* Take  $x^{k-1}$  as the initial point, and compute

$$x^k \in \arg \min_{x \in \mathfrak{R}^n} F(x, \rho_k, \varepsilon_k). \tag{12}$$

*Step 3.* Let  $\varepsilon_{k+1} = \frac{1}{2} \varepsilon_k$ , and

$$\rho_{k+1} = \begin{cases} \rho_k, & \text{if } x^k \in F_{\varepsilon_k}, \\ 2\rho_k, & \text{otherwise.} \end{cases} \quad (13)$$

Let  $k = k + 1$ , go to Step 2.

We now give a convergence result for this algorithm under some mild conditions. First, we give the following assumption.

(A1) For any  $\rho \in [\rho_1, +\infty)$ ,  $\varepsilon \in (0, \varepsilon_1]$ , the set  $\arg \min_{x \in \mathfrak{R}^n} F(x, \rho, \varepsilon) \neq \emptyset$ .

By this assumption, we obtain the following lemma firstly.

**Lemma 3.1** Suppose that (A1) holds. Let  $\{x^k\}$  be the sequence generated by Algorithm 3.1. Then there exists a natural number  $k_0$ , such that for any  $k \geq k_0$ ,

$$x^k \in \mathcal{F}_{\varepsilon_k}.$$

*Proof* Suppose the contrary that there exists a subset  $K \subset \mathbb{N}$ , such that for any  $k$ ,  $x^k \notin \mathcal{F}_{\varepsilon_k}$ , and

$$\rho_k \rightarrow +\infty, k \in K, k \rightarrow \infty.$$

Then there exists  $k_1 \in K$ , such that for any  $k \geq k_1$ ,  $\rho_k > \rho^* + m$ , where  $\rho^* + m$  is given in Theorem 2.4. Therefore, by Theorem 2.4, when  $k \geq k_1$ , it holds that

$$x^k \in \mathcal{F}_{\varepsilon_k}.$$

This contradicts  $x^k \notin \mathcal{F}_{\varepsilon_k}$ .

**Remark 3.1** From Lemma 3.1 we know that  $\rho_k$  remains unchanged after finite iterations.

**Lemma 3.2** Suppose that (A1) holds. Let  $\{x^k\}$  be the sequence generated by Algorithm 3.1, and  $\bar{x}$  be any limit point of  $\{x^k\}$ . Then

$$\bar{x} \in \mathcal{F}_0.$$

**Lemma 3.3** Suppose that (A1) holds. Let  $\{x^k\}$  be the sequence generated by Algorithm 3.1. Then for any  $k$ ,

$$x^k \in \left\{ x \in \mathfrak{R}^n \mid f(x) \leq \inf_{x \in \mathcal{F}_0} f(x) + \frac{1}{2} m \varepsilon_k \right\}. \quad (14)$$

From Lemma 3.2 and Lemma 3.3, we have the following theorem.

**Theorem 3.1** Suppose that (A1) holds. Let  $\{x^k\}$  be the sequence generated by Algorithm 3.1. If  $\bar{x}$  is any limit point of  $\{x^k\}$ , then  $\bar{x}$  is the optimal solution of (P).

Before giving another conclusion, we need the following assumption.

(A2) The function  $\varphi(\delta) = \inf_{x \in \mathcal{F}_\delta} f(x)$  with respect to  $\delta$  is lower semi-continuous at  $\delta = 0$ .

**Theorem 3.2** Suppose that (A1) and (A2) holds. Let  $\{x^k\}$  be the sequence generated by Algorithm 3.1. Then

$$f(x^k) \rightarrow \varphi(0), k \rightarrow \infty.$$

*Proof* By Lemma 3.1, there exists  $k_0$ , such that for any  $k \geq k_0$ ,  $x^k \in \mathcal{F}_{\varepsilon_k}$ . Thus,

$$f(x^k) \geq \varphi(\varepsilon_k).$$

Therefore,

$$\liminf_{k \rightarrow \infty} f(x^k) \geq \liminf_{k \rightarrow \infty} \varphi(\varepsilon_k).$$

From Assumption (A2), we know that  $\liminf_{\delta \downarrow 0} \varphi(\delta) \geq \varphi(0)$ . Therefore,

$$\liminf_{k \rightarrow \infty} f(x^k) \geq \varphi(0). \tag{15}$$

On the other side, by Lemma 3.3, when  $k \geq k_1$ , we have

$$f(x^k) - \varphi(0) \leq \frac{1}{2} m \varepsilon_k.$$

Then

$$\limsup_{k \rightarrow \infty} f(x^k) \leq \varphi(0). \tag{16}$$

Therefore, from (15) and (16),

$$\varphi(0) \leq \liminf_{k \rightarrow \infty} f(x^k) \leq \limsup_{k \rightarrow \infty} f(x^k) \leq \varphi(0).$$

Therefore,

$$f(x^k) \rightarrow \varphi(0), k \rightarrow \infty.$$

The above theorem is different from the conventional conclusion in other literatures with respect to the convergence of penalty method.

In the following we give an approximate smoothed penalty algorithm for Problem (P).

**Algorithm 3.2**

*Step 1.* Let  $\delta > 0$ ,  $0 < \eta < 1 < \sigma$ . Given  $x^0, \varepsilon_1 > 0$ , and  $\rho_1 > 0$ , let  $k = 1$ , go to Step 2.

*Step 2.* Take  $x^{k-1}$  as the initial point, and compute

$$x^k \in \arg \min_{x \in \mathbb{R}^n} F(x, \rho_k, \varepsilon_k).$$

*Step 3.* If  $x^k$  is an  $\varepsilon_k$ -feasible solution of (P), and  $\varepsilon_k \leq \delta$ , then stop. Otherwise, update  $\rho_k$  and  $\varepsilon_k$  by applying the following rules:

if  $x^k$  is an  $\varepsilon_k$ -feasible solution of (P), and  $\varepsilon_k > \delta$ , let  $\varepsilon_{k+1} = \eta \varepsilon_k$ , and

$$\rho_{k+1} = \rho_k;$$

if  $x^k$  is not an  $\varepsilon_k$ -feasible solution of (P), let  $\rho_{k+1} = \sigma \rho_k$  and

$$\varepsilon_{k+1} = \max_j g_j(x^k).$$

Let  $k = k + 1$ , go to Step 2.

**Remark 3.1** By the analysis of the error estimates in Section 2, We know that whenever the penalty parameter  $\rho_k$  is larger than some threshold, then for any  $\varepsilon > 0$ , an optimal solution of the smoothed penalty problem is also an  $\varepsilon$ -feasible solution, which conversely gives an error bound for the optimal objective function value of the original problem.

### 4. Computational Aspects and Numerical Results

In this section, we will discuss some computational aspects and give some nu-



merical results.

We apply Algorithm 3.2 to nonconvex nonlinear programming problem (P), for which we do not need to compute a global optimal solution but a local one. And in this case, we can also obtain the convergence by the following theorem.

For  $x \in \mathfrak{R}^n$ , we denote  $J = \{1, \dots, m\}$  and

$$J^0(x) = \{j \mid g_j(x) = 0, j \in J\},$$

$$J^-(x) = \{j \mid g_j(x) < 0, j \in J\}, J^+(x) = \{j \mid g_j(x) > 0, j \in J\},$$

$$J_\varepsilon^>(x) = \{j \mid g_j(x) > \varepsilon, j \in J\}, J_\varepsilon^\leq(x) = \{j \mid g_j(x) \leq \varepsilon, j \in J\}.$$

**Theorem 4.1** *Suppose Algorithm 3.2 does not terminate after finite iterations and the sequence  $\{F(x^k, \rho_k, \varepsilon_k)\}$  is bounded. Then  $\{x^k\}$  is bounded and any limit point  $x^*$  of  $\{x^k\}$  is feasible to (P), and there exist  $\lambda \geq 0$ , and  $\mu_j \geq 0, j = 1, 2, \dots, m$ , such that*

$$\lambda \nabla f(x^*) + \sum_{j \in J^0(x^*)} \mu_j \nabla g_j(x^*) = 0. \tag{17}$$

*Proof* First we show that  $\{x^k\}$  is bounded. By the assumptions, there is some number  $L$  such that

$$L > F(x^k, \rho_k, \varepsilon_k), k = 0, 1, 2, \dots$$

Suppose to the contrary that  $\{x^k\}$  is unbounded. Choose a subsequence of  $\{x^k\}$  if necessary and we assume that

$$\|x^k\| \rightarrow \infty, \text{ as } k \rightarrow \infty.$$

Then we get

$$L > f(x^k), k = 0, 1, 2, \dots$$

which results in a contradiction since  $f$  is coercive.

We now show that any limit point of  $\{x^k\}$  belongs to  $\mathcal{F}_0$ . Without loss of generality, we assume that

$$\lim_{k \rightarrow \infty} x^k = x^*.$$

Suppose to the contrary that  $x^* \notin \mathcal{F}_0$ , then there exists some  $j \in J$  such that  $g_j(x^*) > \alpha > 0$ . Note that  $g_j, j \in J, F(\cdot, \rho_k, \varepsilon_k) (k = 1, 2, \dots)$  are all continuous.

Note that

$$\begin{aligned} & F(x^k, \rho_k, \varepsilon_k) \\ &= f(x^k) + \sum_{j \in J} P_{\rho_k \varepsilon_k} (g_j(x^k)) \\ &= f(x^k) + \sum_{j \in (J^-(x^k) \cup J^0(x^k))} \frac{1}{2} \varepsilon_k e^{\frac{\rho_k g_j(x^k)}{\varepsilon_k}} \end{aligned}$$

$$\begin{aligned}
 & + \sum_{j \in J_{\varepsilon_k}^{\leq}(x^k)} \left( \rho_k g_j(x^k) + \frac{1}{2} \varepsilon_k e^{-\frac{\rho_k g_j(x^k)}{\varepsilon_k}} \right) \\
 & + \sum_{j \in J_{\varepsilon_k}^{\geq}(x^k)} \left( \rho_k g_j(x^k) + \frac{1}{2} \varepsilon_k e^{-\frac{\rho_k g_j(x^k)}{\varepsilon_k}} \right).
 \end{aligned} \tag{18}$$

If  $k \rightarrow \infty$ , then for any  $k$ , the set  $\{j \mid g_j(x^k) > \alpha\}$  is not empty. Because  $J$  is finite, then there exists a  $j_0 \in J$  such that for any  $k$  is sufficiently large,  $g_{j_0}(x^k) > \alpha$ .

It follows from (18) that

$$F(x^k, \rho_k, \varepsilon_k) \rightarrow +\infty,$$

which contradicts the assumption that  $\{F(x^k, \rho_k, \varepsilon_k)\}$  is bounded.

We now show that (17) holds.

Since for  $k = 1, 2, \dots$ ,

$$\nabla F(x^k, \rho_k, \varepsilon_k) = 0,$$

that is,

$$\begin{aligned}
 & \nabla f(x^k) + \rho_k \sum_{j \in J^-(x^k) \cup J^0(x^k)} \frac{1}{2} e^{-\frac{\rho_k g_j(x^k)}{\varepsilon_k}} \nabla g_j(x^k) \\
 & + \rho_k \sum_{j \in J^+(x^k)} \left( 1 - \frac{1}{2} e^{-\frac{\rho_k g_j(x^k)}{\varepsilon_k}} \right) \nabla g_j(x^k) = 0.
 \end{aligned} \tag{19}$$

For  $k = 1, 2, \dots$ , set

$$\gamma_k = 1 + \sum_{j \in J^-(x^k) \cup J^0(x^k)} \frac{1}{2} \rho_k e^{-\frac{\rho_k g_j(x^k)}{\varepsilon_k}} + \sum_{j \in J^+(x^k)} \rho_k \left( 1 - \frac{1}{2} e^{-\frac{\rho_k g_j(x^k)}{\varepsilon_k}} \right), \tag{20}$$

then,  $\gamma_k > 0$ .

It follows from (19) and (20) that

$$\begin{aligned}
 & \frac{1}{\gamma_k} \nabla f(x^k) + \sum_{j \in J^-(x^k) \cup J^0(x^k)} \frac{\frac{1}{2} \rho_k e^{-\frac{\rho_k g_j(x^k)}{\varepsilon_k}}}{\gamma_k} \nabla g_j(x^k) \\
 & + \sum_{j \in J^+(x^k)} \frac{\rho_k \left( 1 - \frac{1}{2} e^{-\frac{\rho_k g_j(x^k)}{\varepsilon_k}} \right)}{\gamma_k} \nabla g_j(x^k) = 0.
 \end{aligned}$$

Let

$$\lambda^k = \frac{1}{\gamma_k}, \quad \mu_j^k = \frac{\frac{1}{2} \rho_k e^{-\frac{\rho_k g_j(x^k)}{\varepsilon_k}}}{\gamma_k}, \quad j \in J^-(x^k) \cup J^0(x^k),$$

$$\mu_j^k = \frac{\rho_k \left( 1 - \frac{1}{2} e^{-\frac{\rho_k g_j(x^k)}{\epsilon_k}} \right)}{\gamma_k}, \quad j \in J^+(x^k),$$

$$\mu_j^k = 0, \quad j \in J \setminus (J^+(x^k) \cup J^-(x^k) \cup J^0(x^k)).$$

Then we have

$$\lambda^k + \sum_{j \in J} \mu_j^k = 1, \quad \mu_j^k \geq 0, \quad j \in J, \quad k = 0, 1, 2, \dots$$

When  $k \rightarrow \infty$ , we have that  $\lambda^k \rightarrow \lambda \geq 0$ ,  $\mu_j^k \rightarrow \mu_j \geq 0, \forall j \in J$ , and

$$\lambda \nabla f(x^*) + \sum_{j \in J} \mu_j^k \nabla g_j(x^*) = 0,$$

$$\lambda + \sum_{j \in J} \mu_j = 1.$$

For  $j \in J^-(x^*)$ , we get that  $\mu_j^k \rightarrow 0$ . Therefore,  $\mu_j = 0, \forall j \in J^-(x^*)$ . So (17) holds, and this completes the proof.

Theorem 4.1 implies that the sequence generated by Algorithm 3.2 may converge to a FJ point [17] to (P). The speed of convergence depends on the speed of the subprogram employed in Step 2 to solve the unconstrained optimization problem  $\min_{x \in \mathbb{R}^n} F(x, \rho_k, \epsilon_k)$ . Since the Function  $F(x, \rho_k, \epsilon_k)$  is continuously differentiable, we may use a Gradient-type method to get rapid convergence of our algorithm. In the following we will see some numerical experiments.

*Example 4.1 (Hock and Schittkowski [18])* Consider

$$(P4.1) \quad \begin{aligned} \min \quad & -x_1 x_2 x_3 \\ \text{s.t.} \quad & x_1^2 + 2x_2^2 + 4x_3^2 \leq 48. \end{aligned}$$

The optimal solution to (P4.1) is given by  $(x_1, x_2, x_3) = (4, 2.82843, 2)$  with the optimal objective function value 22.627417. Let  $x^0 = (3, 3, 3)$ ,  $\epsilon_1 = 1$ ,  $\rho_1 = 1$ ,  $\eta = 0.1$ , and  $\sigma = 2$  in Algorithm 3.2. We choose  $\delta = 0.00001$  for  $\delta$ -feasibility. Numerical results for (P4.1) are given in Table 1, where for Table 1 we use a Gradient-type algorithm to solve the subproblem in Step 2.

*Example 4.2 (Hock and Schittkowski [18])* Consider

$$(P4.2) \quad \begin{aligned} \min \quad & x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 - 21x_3 + 7x_4 \\ \text{s.t.} \quad & x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 - x_4 \leq 8, \\ & x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 \leq 10, \\ & 2x_1^2 + x_2^2 + x_3^2 + 2x_1 - x_2 - x_4 \leq 5. \end{aligned}$$

The optimal solution to (P4.2) is given by  $(x_1, x_2, x_3, x_4) = (0, 2, 1, -1)$  with the optimal objective function value -44. Let  $x^0 = (0, 0, 0, 0)$ ,  $\epsilon_1 = 1$ ,  $\rho_1 = 4$ ,  $\eta = 0.1$ , and  $\sigma = 2$  in Algorithm 3.2. We choose  $\delta = 0.00001$  for  $\delta$ -feasibility. Numerical results for (P4.2) are given in Table 2, where for Table 2 we use a Gradient-type algorithm to solve the subproblem in Step 2.

*Example 4.3 (Hock and Schittkowski [18])* Consider

**Table 1.** Results with a starting point  $x^0 = (3, 3, 3)$ .

$k$	$\rho_k$	$\varepsilon_k$	$f(x^k)$	$(x_1, x_2, x_3)$	$\max(g_j)$
1	1	1	-23.007125	(4.008345, 2.852342, 2.012314)	0.536164
2	1	0.1	-22.664486	(3.999987, 2.839677, 1.995346)	5.305913E-002
3	1	0.01	-22.630667	(3.999131, 2.838421, 1.993677)	5.315008E-003
4	1	0.001	-22.627288	(3.999045, 2.838296, 1.993510)	5.459629E-004
5	1	0.0001	-22.626939	(3.999036, 2.838283, 1.993493)	5.355349E-005

**Table 2.** Results with a starting point  $x^0 = (0, 0, 0, 0)$ .

$k$	$\rho_k$	$\varepsilon_k$	$f(x^k)$	$(x_1, x_2, x_3, x_4)$	$\max(g_j)$
1	4	1	-43.803614	(1.714448E-002, 0.988781, 2.003621, -0.941664)	-2.005386E-002
2	4	0.1	-43.981817	(-3.180834E-003, 0.994377, 2.006306, -0.986584)	-8.705388E-005
3	4	0.01	-43.997686	(-5.896973E-003, 0.994811, 2.005919, -0.993193)	1.858380E-005
4	4	0.001	-43.999246	(-6.175823E-003, 0.994855, 2.005868, -0.993889)	2.131278E-006

**Table 3.** Results with a starting point  $x^0 = (1, 2, 0, 4, 0, 1, 1)$ .

$k$	$\rho_k$	$\varepsilon_k$	$f(x^k)$	$(x_1, x_2, x_3, x_4, x_5, x_6, x_7)$	$\max(g_j)$
1	1	1	658.071632	(2.380946, 2.034365, -0.383740, 4.714288, -0.608364, 1.018706, 1.617855)	21.195303
2	2	21.195303	681.505337	(2.060105, 1.967376, -0.366454, 4.424977, -0.621624, 0.964168, 1.679359)	1.279039
3	2	2.119530	680.823561	(2.278954, 1.954036, -0.432778, 4.375586, -0.624030, 1.027213, 1.607750)	0.154300
4	2	0.211953	680.651942	(2.325791, 1.951432, -0.453430, 4.366594, -0.624490, 1.038047, 1.594178)	1.593473E-002
5	2	2.119530E-002	680.633110	(2.330656, 1.951086, -0.456864, 4.365882, -0.624528, 1.039155, 1.592410)	1.602348E-003
6	2	2.119530E-003	680.631250	(2.331044, 1.951023, -0.456929, 4.365892, -0.624526, 1.039303, 1.592087)	1.601163E-004
7	2	2.119530E-004	680.631062	(2.331083, 1.951017, -0.456935, 4.365893, -0.624526, 1.039318, 1.592054)	1.813545E-005
8	2	2.119530E-005	680.631044	(2.331087, 1.951017, -0.456936, 4.365893, -0.624526, 1.039319, 1.592051)	3.983831E-006

$$\begin{aligned}
 & \min \quad (x_1 - 10)^2 + 5(x_2 - 12)^2 + x_3^4 + 3(x_4 - 11)^2 + 10x_5^6 \\
 & \quad + 7x_6^2 + x_7^4 - 4x_6x_7 - 10x_6 - 8x_7 \\
 (P4.3) \quad & \text{s.t.} \quad 2x_1^2 + 3x_2^4 + x_3 + 4x_4^2 + 5x_5 \leq 127, \\
 & \quad 7x_1 + 3x_2 + 10x_3^2 + x_4 - x_5 \leq 282, \\
 & \quad 23x_1 + x_2^2 + 6x_6^2 - 8x_7 \leq 196, \\
 & \quad 4x_1^2 + x_2^2 - 3x_1x_2 + 2x_3^2 + 5x_6 - 11x_7 \leq 0.
 \end{aligned}$$

The optimal solution to (P4.3) is given by

$$(x_1, x_2, x_3, x_4, x_5, x_6, x_7)$$

$$= (2.330499, 1.951372, -0.4775414, 4.365726, -0.6244870, 1.038131, 1.594227)$$

with the optimal objective function value 680.6300573. Let  $x^0 = (1, 2, 0, 4, 0, 1, 1)$ ,  $\varepsilon_1 = 1$ ,  $\rho_1 = 1$ ,  $\eta = 0.1$ , and  $\sigma = 2$  in Algorithm 3.2. We choose  $\delta = 0.00001$  for  $\delta$ -feasibility. Numerical results for (P4.3) are given in **Table 3**, where for **Table 3** we use a Gradient-type algorithm to solve the subproblem in Step 2.

From the above classical examples, we can see that our approximate algorithm can produce the approximate optimal solutions of the corresponding problem successfully. But the convergent speed can be improved if we use the Newton-type method in Step 2 of Algorithm 3.2, which will be researched in our future work.

## Funding

This research is supported by the Natural Science Foundations of Shandong Province (ZR2015AL011).

## Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

## References

- [1] Zangwill, W.I. (1967) Nonlinear Programming via Penalty Function Management. *Science*, **13**, 334-358. <https://doi.org/10.1287/mnsc.13.5.344>
- [2] Di Pillo, G., Liuzzi, G. and Lucidi, S. (2011) An Exact Penalty-Lagrangian Approach for Large-Scale Nonlinear Programming. *Optimization*, **60**, 223-252. <https://doi.org/10.1080/02331934.2010.505964>
- [3] Han, S.P. and Mangasarian, O.L. (1979) Exact Penalty Functions in Nonlinear Programming. *Mathematical Programming*, **17**, 251-269. <https://doi.org/10.1007/BF01588250>
- [4] Ben-tal, A. and Teboulle, M. (1989) A Smoothing Technique for Non-Differentiable Optimization Problems. *Lecture Notes in Mathematics*, Springer Verlag, Berlin, Vol. 1405, 1-11. <https://doi.org/10.1007/BFb0083582>
- [5] Herty, M., Klar, A., Singh, A.K. and Spellucci, P. (2007) Smoothed Penalty Algorithms for Optimization of Nonlinear Models. *Computational Optimization and Applications*, **37**, 157-176. <https://doi.org/10.1007/s10589-007-9011-6>
- [6] Pinar, M.C. and Zenios, S.A. (1994) On Smoothing Exact Penalty Functions for Convex Constrained Optimization. *SIAM Journal on Optimization*, **4**, 486-511. <https://doi.org/10.1137/0804027>
- [7] Auslender, A., Cominetti, R. and Haddou, M. (1997) Asymptotic Analysis for Penalty and Barrier Methods in Convex and Linear Programming. *Mathematics of Operational Research*, **22**, 43-62. <https://doi.org/10.1287/moor.22.1.43>
- [8] Gonzaga, C.C. and Castillo, R.A. (2003) A Nonlinear Programming Algorithm Based on Non-Coercive Penalty Functions. *Mathematical Programming*, **96**, 87-101. <https://doi.org/10.1007/s10107-002-0332-z>
- [9] Xu, X.S., Meng, Z.Q., Sun, J.W., Huang, L.G. and Shen, R. (2013) A Second-Order Smooth Penalty Function Algorithm for Constrained Optimization Problems.

*Computational Optimization and Application*, **55**, 155-172.

<https://doi.org/10.1007/s10589-012-9504-9>

- [10] Lian, S.J. (2012) Smoothing Approximation to  $L_1$  Exact Penalty Function for Inequality Constrained Optimization. *Applied Mathematics and Computation*, **219**, 3113-3121. <https://doi.org/10.1016/j.amc.2012.09.042>
- [11] Chen, C. and Mangasarian, O.L. (1995) Smoothing Methods for Convex Inequalities and Linear Complementarity Problems. *Mathematical Programming*, **71**, 51-69. <https://doi.org/10.1007/BF01592244>
- [12] Wu, Z.Y., Lee, H.W.J., Bai, F.S. and Zhang, L.S. (2017) Quadratic Smoothing Approximation to  $L_1$  Exact Penalty Function in Global Optimization. *Journal of Industrial and Management Optimization*, **1**, 533-547. <https://doi.org/10.3934/jimo.2005.1.533>
- [13] Meng, Z.Q., Dang, C.Y. and Yang, X.Q. (2006) On the Smoothing of the Square-Root Exact Penalty Function for Inequality Constrained Optimization. *Computational Optimization and Applications*, **35**, 375-398. <https://doi.org/10.1007/s10589-006-8720-6>
- [14] Wu, Z.Y., Bai, F.S., Yang, X.Q. and Zhang, L.S. (2004) An Exact Lower Order Penalty Function and Its Smoothing in Nonlinear Programming. *Optimization*, **53**, 51-68. <https://doi.org/10.1080/02331930410001662199>
- [15] Lian, S.J. (2016) Smoothing of the Lower-Order Exact Penalty Function for Inequality Constrained Optimization. *Journal of Inequalities and Applications*, **2016**, Article No. 185. <https://doi.org/10.1186/s13660-016-1126-9>
- [16] Liu, B.Z. (2009) On Smoothing Exact Penalty Functions for Nonlinear Constrained Optimization Problems. *Journal of Applied Mathematics and Computing*, **30**, 259-270. <https://doi.org/10.1007/s12190-008-0171-z>
- [17] Mangasarian, O.L. and Fromovitz, S. (1967) The Fritz John Necessary Optimality Conditions in the Presence of Equality and Inequality Constraints. *Journal of Mathematical Analysis and Applications*, **17**, 37-47. [https://doi.org/10.1016/0022-247X\(67\)90163-1](https://doi.org/10.1016/0022-247X(67)90163-1)
- [18] Hock, W. and Schittkowski, K. (1981) Test Examples for Nonlinear Programming Codes. Lecture Notes in Economics and Mathematical Systems, Vol. 187, Springer-Verlag, Berlin, Heidelberg, New York. <https://doi.org/10.1007/978-3-642-48320-2>