Smoothed Lower Order Penalty Function for Constrained Optimization Problems

Nguyen Thanh Binh

***Abstract*—The paper introduces a smoothing method to the lower order penalty function for constrained optimization problems. It is shown that, under some mild conditions, an optimal solution of the smoothed penalty problem is an approximate optimal solution of the original problem. Based on the smoothed penalty function, an algorithm is presented and its convergence is proved under some mild assumptions. Numerical examples show that the presented algorithm is effective.**

***Index Terms*—constrained optimization problem, penalty function method, exact penalty function, smoothing method.**

I. Introduction

CONSIDER the constrained optimization problem:

$$
\begin{aligned}
(\mathrm{P}) \qquad & \min \ f(x) \\
& \text{s.t.} \ \ g_i(x) \le 0, \quad i = 1, 2, \ldots, m, \\
& \qquad x \in R^n,
\end{aligned}
$$

where $f, g_i : R^n \to R,\ i \in I = \{1, 2, \ldots, m\}$ are real valued functions, and $X_0 = \{x \in R^n \mid g_i(x) \le 0,\ i \in I\}$ is the feasible set to (P). Constrained optimization problems have been applied in many fields, such as including optimal control, engineering, national defence, economy, and finance etc. Up to now, the methods of solving the constrained optimization problems have been well studied in the literatures. The use of penalty functions to solve constrained optimization problems is generally attributed to Courant. The significant progress in solving many practical nonlinear constrained optimization problems by using exact penalty function methods have been introduced in literatures [1], [2], [3], [4], [5], [6], [7]. Zangwill [7] proposed the classic l_1 exact penalty function as follows:

$$F_\sigma(x) = f(x) + \sigma \sum_{i=1}^{m} g_i^+(x), \tag{1}$$

where $g_i^+(x) = \max\{0, g_i(x)\},\ i \in I$ and $\sigma > 0$ is a penalty parameter. It is known from the theory of ordinary constrained optimization that the l_1 exact penalty function is a better candidate for penalization. However, it is not a smooth function and causes some numerical instability problems in its implementation when the value of the penalty parameter σ becomes larger. Thus, smoothing methods for smoothing the exact penalty function attract much attention, see, e.g., [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18]. In recent years, the lower order penalty function

$$F_\sigma^k(x) = f(x) + \sigma \sum_{i=1}^{m} [g_i^+(x)]^k, \tag{2}$$

where $k \in (0, 1)$ and $\sigma > 0$ have been introduced in [12], [13], [15]. Obviously, when $k = 1$, the lower order penalty function (2) is reduced to the exact penalty function (1). Meng et al. [12] discussed two smoothing approximations to the lower order penalty function for inequality constrained optimization. Meng et al. [13] discussed a smoothing method of lower order penalty function and gave a robust SQP method for (P) by integrating the smoothed penalty function with the SQP method. Wu et al. [15] proposed the ϵ-smoothing of (2), and got a modified exact penalty function under some mild conditions. Error estimates of the optimal value of the smoothed penalty problem and that of the original penalty problem are obtained.

Based on the lower order penalty function, this paper introduces a smoothing lower order penalty function method which differs from the smoothing penalty function methods in [12], [13], [15], and proposes a corresponding algorithm for solving (P). Numerical results show that this algorithm is of good convergence and is efficient in solving some constrained optimization problems.

The rest of this paper is organized as follows. In Section II, we propose a new smoothing function to the exact penalty function (2), and discuss its some fundamental properties. In Section III, an algorithm based on the smoothing lower order penalty function is proposed and its global convergence presented, with some numerical examples are given.

II. Smoothed lower order method

Consider $p^k(v) : R \to R :$

$$p^k(v) = \begin{cases} 0 & \text{if} \quad v < 0, \\ v^k & \text{if} \quad v \ge 0, \end{cases}$$

where $k \in (0, 1)$. Then,

$$F_\sigma^k(x) = f(x) + \sigma \sum_{i=1}^{m} p^k(g_i(x)). \tag{3}$$

The corresponding penalty problem to (3) is defined as follows,

$$
\begin{aligned}
(P_\sigma) \quad & \min \ F_\sigma^k(x) \\
& \text{s.t.} \ x \in R^n.
\end{aligned}
$$

For $k \in (0, 1)$ and any $\epsilon > 0$, we define function $p^k_{\epsilon,\sigma}(v)$ as

$$
{p^k}_{\epsilon,\sigma}(v) = \begin{cases} 0 & \text{if } v < 0, \\ \dfrac{m^2\sigma^2 v^{3k}}{6\epsilon^2} & \text{if } 0 \le v < \left(\frac{\epsilon}{m\sigma}\right)^{\frac{1}{k}}, \\ v^k + \dfrac{\epsilon^2 v^{-k}}{2m^2\sigma^2} - \dfrac{4\epsilon}{3m\sigma} & \text{if } v \ge \left(\frac{\epsilon}{m\sigma}\right)^{\frac{1}{k}}, \end{cases}
$$

where ϵ is a smoothing parameter and $\sigma > 0$ is given.

Manuscript received July 17, 2015; revised October 20, 2015. This work is supported by National Natural Science Foundation of China (Grant No. 11371242).

N. T. Binh is with the Department of Mathematics, Shanghai University, Shanghai 200444, P. R. China, e-mail: bingbongyb@gmail.com

(Advance online publication: 15 February 2016)

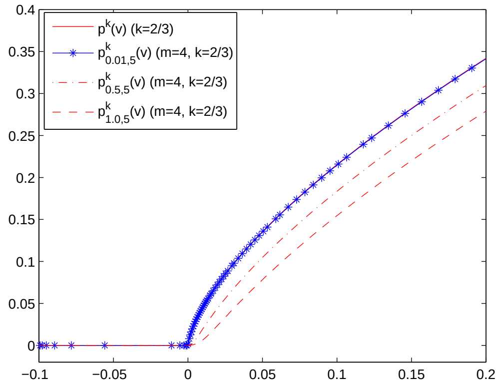

Fig. 1. The behavior of $p^k(v)$ and $p^k_{\epsilon,\sigma}(v)$.

Remark 2.1. Obviously, $p^k_{\epsilon,\sigma}(v)$ has the following attractive properties:

(i) $p^k_{\epsilon,\sigma}(v)$ is twice continuously differentiable for $\frac{2}{3} < k < 1$ on R, where

$$[p^k{}_{\epsilon,\sigma}(v)]' = \begin{cases} 0 & \text{if } v < 0, \\ \dfrac{km^2\sigma^2 v^{3k-1}}{2\epsilon^2} & \text{if } 0 \le v < \left(\frac{\epsilon}{m\sigma}\right)^{\frac{1}{k}}, \\ kv^{k-1} - \dfrac{k\epsilon^2 v^{-k-1}}{2m^2\sigma^2} & \text{if } v \ge \left(\frac{\epsilon}{m\sigma}\right)^{\frac{1}{k}}, \end{cases}$$

and

$$[p^k{}_{\epsilon,\sigma}(v)]'' = \begin{cases} 0 \quad \text{if } v < 0, \\ \dfrac{k(3k-1)m^2\sigma^2 v^{3k-2}}{2\epsilon^2} \quad \text{if } 0 \le v < \left(\dfrac{\epsilon}{m\sigma}\right)^{\frac{1}{k}}, \\ k(k-1)v^{k-2} + \dfrac{k(k+1)\epsilon^2 v^{-k-2}}{2m^2\sigma^2} \quad \text{if } v \ge \left(\dfrac{\epsilon}{m\sigma}\right)^{\frac{1}{k}}. \end{cases}$$

(ii) $\lim_{\epsilon\to 0} p^k_{\epsilon,\sigma}(v) = p^k(v)$.

(iii) $p^k(v) \ge p^k_{\epsilon,\sigma}(v), \quad \forall v \in R$.

Figure 1 shows the behavior of $p^k(v)$ and $p^k_{\epsilon,\sigma}(v)$. Assume that $f,\ g_i(i \in I)$ are continuously differentiable functions.

Let

$$F^k_{\epsilon,\sigma}(x) = f(x) + \sigma\sum_{i=1}^m p^k_{\epsilon,\sigma}(g_i(x)). \tag{4}$$

Clearly, $F^k_{\epsilon,\sigma}(x)$ is continuously differentiable at any $x \in R^n$. Consider the following smoothed penalty problem:

$$(SP_{\epsilon,\sigma}) \quad \min\ F^k_{\epsilon,\sigma}(x) \qquad \text{s.t.}\ \ x \in R^n.$$

Lemma 2.1. We have

$$0 \le F^k_\sigma(x) - F^k_{\epsilon,\sigma}(x) \le \frac{4\epsilon}{3} \tag{5}$$

for any given $x \in R^n$, $\epsilon > 0$ and $\sigma > 0$, where $F^k_\sigma(x)$ and $F^k_{\epsilon,\sigma}(x)$ are given in (3) and (4) respectively.

Proof For any $x \in R^n$, $\epsilon > 0$, $\sigma > 0$, we have

$$F^k_\sigma(x) - F^k_{\epsilon,\sigma}(x) = \sigma\sum_{i=1}^m \left(p^k(g_i(x)) - p^k_{\epsilon,\sigma}(g_i(x))\right).$$

Note that

$$p^k(g_i(x)) - p^k{}_{\epsilon,\sigma}(g_i(x)) = \begin{cases} 0 \quad \text{if } g_i(x) < 0, \\ 0 \le (g_i(x))^k - \dfrac{m^2\sigma^2(g_i(x))^{3k}}{6\epsilon^2} < \dfrac{\epsilon}{m\sigma} \\ \text{if } 0 \le g_i(x) < \left(\dfrac{\epsilon}{m\sigma}\right)^{\frac{1}{k}}, \\ 0 < \dfrac{4\epsilon}{3m\sigma} - \dfrac{\epsilon^2(g_i(x))^{-k}}{2m^2\sigma^2} \le \dfrac{4\epsilon}{3m\sigma} \\ \text{if } g_i(x) \ge \left(\dfrac{\epsilon}{m\sigma}\right)^{\frac{1}{k}}. \end{cases}$$

for any $i = 1, 2, \ldots, m$. That is,

$$0 \le p^k(g_i(x)) - p^k_{\epsilon,\sigma}(g_i(x)) \le \frac{4\epsilon}{3m\sigma}.$$

Thus,

$$0 \le \sum_{i=1}^m (p^k(g_i(x)) - p^k_{\epsilon,\sigma}(g_i(x))) \le \frac{4\epsilon}{3\sigma},$$

which implies

$$0 \le \sigma\sum_{i=1}^m (p^k(g_i(x)) - p^k_{\epsilon,\sigma}(g_i(x))) \le \frac{4\epsilon}{3}.$$

Therefore,

$$0 \le F^k_\sigma(x) - F^k_{\epsilon,\sigma}(x) \le \frac{4\epsilon}{3}.$$

The proof is completed.

Based on the Lemma 2.1, we have the following two theorems.

Theorem 2.1. Let $\{\epsilon_j\} \to 0$, $\forall\epsilon_j > 0$, and assume x_j is a solution to $(SP_{\epsilon_j,\sigma})$ for some $\sigma > 0$. Let x' be an accumulation point of the sequence $\{x_j\}$. Then x' is an optimal solution to (P_σ).

Proof Since x_j is a solution to problem $(SP_{\epsilon_j,\sigma})$, we have

$$F^k_{\epsilon_j,\sigma}(x_j) \le F^k_{\epsilon_j,\sigma}(x).$$

By Lemma 2.1, we have

$$F^k_\sigma(x_j) \le F^k_{\epsilon_j,\sigma}(x_j) + \frac{4\epsilon_j}{3},$$
$$F^k_{\epsilon_j,\sigma}(x) \le F^k_\sigma(x).$$

It follows

$$\begin{aligned} F^k_\sigma(x_j) &\le F^k_{\epsilon_j,\sigma}(x) + \frac{4\epsilon_j}{3} \\ &\le F^k_\sigma(x) + \frac{4\epsilon_j}{3}. \end{aligned}$$

Since $\{\epsilon_j\} \to 0$ and x' be an accumulation point of the sequence $\{x_j\}$, we have

$$F^k_\sigma(x') \le F^k_\sigma(x).$$

The proof is completed.

Theorem 2.2. For some $\sigma > 0$ and $\epsilon > 0$, let x^*_σ be an optimal solution to (P_σ) and $x^*_{\epsilon,\sigma}$ be an optimal solution to $(SP_{\epsilon,\sigma})$. Then,

$$0 \le F^k_\sigma(x^*_\sigma) - F^k_{\epsilon,\sigma}(x^*_{\epsilon,\sigma}) \le \frac{4\epsilon}{3}. \tag{6}$$

If both x^*_σ and $x^*_{\epsilon,\sigma}$ are feasible to (P), then

$$f(x^*_{\epsilon,\sigma}) \le f(x^*_\sigma) \le f(x^*_{\epsilon,\sigma}) + \frac{4\epsilon}{3}. \tag{7}$$

(Advance online publication: 15 February 2016)

Proof By Lemma 2.1, for $\sigma > 0$ and $\epsilon > 0$, we have that

$$\begin{aligned} 0 &\le F_\sigma^k(x_\sigma^*) - F_{\epsilon,\sigma}^k(x_\sigma^*) \\ &\le F_\sigma^k(x_\sigma^*) - F_{\epsilon,\sigma}^k(x_{\epsilon,\sigma}^*) \\ &\le F_\sigma^k(x_{\epsilon,\sigma}^*) - F_{\epsilon,\sigma}^k(x_{\epsilon,\sigma}^*) \\ &\le \frac{4\epsilon}{3}. \end{aligned}$$

That is,

$$0 \le F_\sigma^k(x_\sigma^*) - F_{\epsilon,\sigma}^k(x_{\epsilon,\sigma}^*) \le \frac{4\epsilon}{3}$$

and

$$\begin{aligned} 0 \le &\left\{ f(x_\sigma^*) + \sigma \sum_{i=1}^m p^k(g_i(x_\sigma^*)) \right\} \\ &- \left\{ f(x_{\epsilon,\sigma}^*) + \sigma \sum_{i=1}^m p_{\epsilon,\sigma}^k(g_i(x_{\epsilon,\sigma}^*)) \right\} \le \frac{4\epsilon}{3}. \end{aligned}$$

Furthermore, if x_σ^* and $x_{\epsilon,\sigma}^*$ are feasible to (P), then

$$\sum_{i=1}^m p^k(g_i(x_\sigma^*)) = \sum_{i=1}^m p_{\epsilon,\sigma}^k(g_i(x_{\epsilon,\sigma}^*)) = 0.$$

Therefore,

$$0 \le f(x_\sigma^*) - f(x_{\epsilon,\sigma}^*) \le \frac{4\epsilon}{3}.$$

That is,

$$f(x_{\epsilon,\sigma}^*) \le f(x_\sigma^*) \le f(x_{\epsilon,\sigma}^*) + \frac{4\epsilon}{3}.$$

The proof is completed.

Definition 2.1. A point $x_\epsilon^* \in R^n$ is called ϵ-feasible solution to (P) if

$$g_i(x_\epsilon^*) \le \epsilon, \; i = 1, 2, \ldots, m.$$

Definition 2.2. For $\overline{x} \in R^n$, a point $\overline{y} \in R^m$ is called a Lagrange multiplier vector corresponding to $\overline{x}$ if $\overline{x}$ and $\overline{y}$ satisfy that

$$\nabla f(\overline{x}) = -\sum_{i=1}^m \overline{y}_i \nabla g_i(\overline{x}), \tag{8}$$

$$\overline{y}_i g_i(\overline{x}) = 0, \quad g_i(\overline{x}) \le 0, \quad \overline{y}_i \ge 0, \quad i \in I. \tag{9}$$

Theorem 2.3. Suppose that $f, \; g_i \; (i \in I)$ in (P) are convex. Let $\overline{x}$ be an optimal solution to (P) and $x_{\epsilon,\sigma}^*$ be an optimal solution to $(SP_{\epsilon,\sigma})$. If $x_{\epsilon,\sigma}^*$ is feasible to (P), and $\overline{y} \in R^m$ be a Lagrange multiplier vector corresponding to $x_{\epsilon,\sigma}^*$, then

$$f(\overline{x}) \le f(x_{\epsilon,\sigma}^*) \le f(\overline{x}) + \frac{4\epsilon}{3} \tag{10}$$

for any $\epsilon > 0$.

Proof By the convexity of $f, \; g_i \; (i \in I)$, we have

$$f(\overline{x}) \ge f(x_{\epsilon,\sigma}^*) + \nabla f(x_{\epsilon,\sigma}^*)^T(\overline{x} - x_{\epsilon,\sigma}^*), \tag{11}$$

$$g_i(\overline{x}) \ge g_i(x_{\epsilon,\sigma}^*) + \nabla g_i(x_{\epsilon,\sigma}^*)^T(\overline{x} - x_{\epsilon,\sigma}^*). \tag{12}$$

By (3), (8), (9), (11) and (12), we have

$$\begin{aligned} F_\sigma^k(\overline{x}) &= f(\overline{x}) + \sigma \sum_{i=1}^m p^k(g_i(\overline{x})) \\ &\ge f(x_{\epsilon,\sigma}^*) + \nabla f(x_{\epsilon,\sigma}^*)^T(\overline{x} - x_{\epsilon,\sigma}^*) \\ &= f(x_{\epsilon,\sigma}^*) - \sum_{i=1}^m \overline{y}_i \nabla g_i(x_{\epsilon,\sigma}^*)^T(\overline{x} - x_{\epsilon,\sigma}^*) \\ &\ge f(x_{\epsilon,\sigma}^*) - \sum_{i=1}^m \overline{y}_i \left[g_i(\overline{x}) - g_i(x_{\epsilon,\sigma}^*) \right] \\ &= f(x_{\epsilon,\sigma}^*) - \sum_{i=1}^m \overline{y}_i g_i(\overline{x}) \\ &\ge f(x_{\epsilon,\sigma}^*). \end{aligned}$$

By Lemma 2.1, we have

$$F_\sigma^k(\overline{x}) \le F_{\epsilon,\sigma}^k(\overline{x}) + \frac{4\epsilon}{3}.$$

It follows

$$\begin{aligned} f(x_{\epsilon,\sigma}^*) &\le F_{\epsilon,\sigma}^k(\overline{x}) + \frac{4\epsilon}{3} \\ &= f(\overline{x}) + \sigma \sum_{i=1}^m p_{\epsilon,\sigma}^k(g_i(\overline{x})) + \frac{4\epsilon}{3} \\ &= f(\overline{x}) + \frac{4\epsilon}{3}. \end{aligned}$$

Since $x_{\epsilon,\sigma}^*$ is feasible to problem (P), that

$$f(\overline{x}) \le f(x_{\epsilon,\sigma}^*).$$

Thus,

$$f(\overline{x}) \le f(x_{\epsilon,\sigma}^*) \le f(\overline{x}) + \frac{4\epsilon}{3}.$$

The proof is completed.

Theorem 2.2 show that an approximate solution to $(SP_{\epsilon,\sigma})$ is also an approximate solution to (P_σ) when the error ϵ is small enough. By Theorem 2.3, under some mild conditions, an optimal solution to $(SP_{\epsilon,\sigma})$ becomes an approximately optimal solution to (P). Therefore, we may obtain an approximately optimal solution to (P) by solving problem $(SP_{\epsilon,\sigma})$.

III. Algorithm and numerical examples

In this section, we propose an algorithm to solve problem (P) via solving the problem $(SP_{\epsilon,\sigma})$, defined as Algorithm I.

Algorithm I

Step 1: Given a point $x_1^0 \in R^n$. Choose $\epsilon_1 > 0, \; \sigma_1 > 0, \; 0 < \gamma < 1, \; \beta > 1$ and $\epsilon > 0$. Let $j = 1$, and go to Step 2.

Step 2: Solve the following problem:

$$(SP_{\epsilon_j,\sigma_j}) \qquad \min_{x \in R^n} \; F_{\epsilon_j,\sigma_j}^k(x) = f(x) + \sigma_j \sum_{i=1}^m p_{\epsilon_j,\sigma_j}^k(g_i(x))$$

starting from the point x_j^0. Let $x_{\epsilon_j,\sigma_j}^*$ be an optimal solution obtained ($x_{\epsilon_j,\sigma_j}^*$ is obtained by the BFGS method given in [19]).

Step 3: If $x_{\epsilon_j,\sigma_j}^*$ is ϵ-feasible to (P), then stop. Otherwise, let $\sigma_{j+1} = \beta\sigma_j, \; \epsilon_{j+1} = \gamma\epsilon_j, \; x_{j+1}^0 = x_{\epsilon_j,\sigma_j}^*$ and $j = j + 1$, then go to Step 2.

Remark 3.1. In this Algorithm I, from $0 < \gamma < 1,\ \beta > 1$, the sequence $\{\epsilon_j\} \to 0$ and the sequence $\{\sigma_j\} \to +\infty$, as $j \to +\infty$.

Under mild conditions, we now prove the convergence of the Algorithm I.

Theorem 3.1. For $\frac{1}{3} < k < 1$, suppose that the set

$$\arg \min_{x \in R^n} F^k_{\epsilon,\sigma}(x) \neq \emptyset \tag{13}$$

for any $\epsilon \in (0, \epsilon_1]$, and $\sigma \in [\sigma_1, +\infty)$. Let $\{x'_j\}$ be a sequence generated by Algorithm I. If $\{x'_j\}$ has limit point, then any limit point of $\{x'_j\}$ is an optimal solution to (P).

Proof Let x' be any limit point of $\{x'_j\}$, then there exists a natural number set $J \subset N$, such that $x'_j \to x',\ j \in J$. It is clear that, if we prove that (i) $x' \in X_0$, and (ii) $f(x') \leq \inf_{x \in X_0} f(x)$, then x' is an optimal solution to (P).

(i) Suppose to the contrary that $x' \notin X_0$, then there exists $\theta_0 > 0$ and the subset $J' \subset J$, there exists some $i' \in I$ such that

$$g_{i'}(x'_j) \geq \theta_0$$

for any $j \in J'$.

When $\left(\frac{\epsilon_j}{m\sigma_j}\right)^{\frac{1}{k}} > g_{i'}(x'_j) \geq \theta_0$, from the definition of $p^k_{\epsilon,\sigma}(v)$ and Step 2 in Algorithm I, we have

$$\begin{aligned} f(x'_j) + \frac{m\sigma_j^3\theta_0^{3k}}{6\epsilon_j^2} &\leq F^k_{\epsilon_j,\sigma_j}(x'_j) \\ &\leq F^k_{\epsilon_j,\sigma_j}(x) = f(x) \end{aligned}$$

for any $x \in X_0$, which contradicts with $\sigma_j \to +\infty$, and $\epsilon_j \to 0$.

When $g_{i'}(x'_j) \geq \theta_0 \geq \left(\frac{\epsilon_j}{m\sigma_j}\right)^{\frac{1}{k}}$ or $g_{i'}(x'_j) \geq \left(\frac{\epsilon_j}{m\sigma_j}\right)^{\frac{1}{k}} \geq \theta_0$, from the definition of $p^k_{\epsilon,\sigma}(v)$ and Step 2 in Algorithm I, we have

$$\begin{aligned} f(x'_j) + \sigma_j\theta_0^k + \frac{\epsilon_j^2\theta_0^{-k}}{2m^2\sigma_j} - \frac{4\epsilon_j}{3m} &\leq F^k_{\epsilon_j,\sigma_j}(x'_j) \\ &\leq F^k_{\epsilon_j,\sigma_j}(x) = f(x) \end{aligned}$$

for any $x \in X_0$, which contradicts with $\sigma_j \to +\infty$, and $\epsilon_j \to 0$.

Thus, $x' \in X_0$.

(ii) For any $x \in X_0$, from the definition of $p^k_{\epsilon,\sigma}(v)$ and Step 2 in Algorithm I, we have

$$f(x'_j) \leq F^k_{\epsilon_j,\sigma_j}(x'_j) \leq F^k_{\epsilon_j,\sigma_j}(x) = f(x),$$

thus, $f(x') \leq \inf_{x \in X_0} f(x)$ holds.
The proof is completed.

Theorem 3.2. For $\frac{1}{3} < k < 1$, let $\{x'_j\}$ be a sequence generated by Algorithm I. Suppose that $\lim_{\|x\| \to +\infty} f(x) = +\infty$ and the sequence $\{F^k_{\epsilon_j,\sigma_j}(x'_j)\}$ is bounded. Then,

(i) the sequence $\{x'_j\}$ is bounded,

(ii) any limit point x' of $\{x'_j\}$ is feasible to (P),

(iii) any limit point x' of $\{x'_j\}$, and there exists $\kappa_i \geq 0$ ($i \in I$), such that

$$\nabla f(x') + \sum_{i \in I} \kappa_i \nabla g_i(x') = 0. \tag{14}$$

Proof (i) First, we will prove that $\{x'_j\}$ is bounded. Note that

$$F^k_{\epsilon_j,\sigma_j}(x'_j) = f(x'_j) + \sigma_j \sum_{i=1}^m p^k_{\epsilon_j,\sigma_j}(g_i(x'_j)),\quad j = 0, 1, 2, \ldots, \tag{15}$$

and by the definition of $p^k_{\epsilon,\sigma}(v)$, we have

$$\sum_{i=1}^m p^k_{\epsilon_j,\sigma_j}(g_i(x'_j)) \geq 0. \tag{16}$$

Suppose to the contrary that $\{x'_j\}$ is unbounded and without loss of generality, $\|x'_j\| \to +\infty$ as $j \to +\infty$. Then, $\lim_{j \to +\infty} f(x'_j) = +\infty$, and from (15) and (16), we have

$$F^k_{\epsilon_j,\sigma_j}(x'_j) \geq f(x'_j) \to +\infty,\ \sigma_j > 0,$$

which contradicts with $\{F^k_{\epsilon_j,\sigma_j}(x'_j)\}$ is bounded. Thus, $\{x'_j\}$ is bounded.

(ii) Next, we will prove that any limit point x' of $\{x'_j\}$ is feasible to (P).

For $x \in R^n$, let

$$I^-_\epsilon = \left\{ i \mid 0 \leq g_i(x) < \left(\frac{\epsilon}{m\sigma}\right)^{\frac{1}{k}},\ i = 1, 2, \ldots, m \right\},$$

$$I^+_\epsilon = \left\{ i \mid g_i(x) \geq \left(\frac{\epsilon}{m\sigma}\right)^{\frac{1}{k}},\ i = 1, 2, \ldots, m \right\}.$$

Without loss of generality, assume that $\lim_{j \to +\infty} x'_j = x'$. Suppose to the contrary that $x' \notin X_0$, then there exist some $i \in I$, such that $g_i(x') \geq \alpha > 0$. Note that

$$\begin{aligned} F^k_{\epsilon_j,\sigma_j}(x'_j) &= f(x'_j) + \sigma_j \sum_{i \in I^-_{\epsilon_j}} \frac{m^2\sigma_j^2(g_i(x'_j))^{3k}}{6\epsilon_j^2} \\ &+\sigma_j \sum_{i \in I^+_{\epsilon_j}} \left((g_i(x'_j))^k + \frac{\epsilon_j^2(g_i(x'_j))^{-k}}{2m^2\sigma_j^2} - \frac{4\epsilon_j}{3m\sigma_j} \right). \end{aligned} \tag{17}$$

Since $g_i(x') \geq \alpha > 0$, then for any sufficiently large j, the set $\{i \mid g_i(x'_j) \geq \alpha\}$ is not empty. Then there exists an $i' \in I$ that satisfies $g_{i'}(x'_j) \geq \alpha$. If $j \to +\infty,\ \sigma_j \to +\infty,\ \epsilon_j \to 0$, it follows from (17) that $F^k_{\epsilon_j,\sigma_j}(x'_j) \to +\infty$, which contradicts with $\{F^k_{\epsilon_j,\sigma_j}(x'_j)\}$ is bounded. Therefore, x' is feasible to (P).

(iii) Finally, we prove that (14) holds. By Step 2 in Algorithm I, we have $\nabla F^k_{\epsilon_j,\sigma_j}(x'_j) = 0$, that is

$$\begin{aligned} &\nabla f(x'_j) + \sigma_j \sum_{i \in I^-_{\epsilon_j}} \frac{km^2\sigma_j^2(g_i(x'_j))^{3k-1}}{2\epsilon_j^2} \nabla g_i(x'_j) \\ &+\sigma_j \sum_{i \in I^+_{\epsilon_j}} \left(k(g_i(x'_j))^{k-1} - \frac{k\epsilon_j^2(g_i(x'_j))^{-k-1}}{2m^2\sigma_j^2} \right) \nabla g_i(x'_j) = 0. \end{aligned}$$

which implies

$$\begin{aligned} \nabla f(x'_j) &+ \sigma_j \sum_{i \in I^-_{\epsilon_j}} \frac{km^2\sigma_j^2[p^k(g_i(x'_j))]^{\frac{3k-1}{k}}}{2\epsilon_j^2} \nabla g_i(x'_j) \\ &+ \sigma_j \sum_{i \in I^+_{\epsilon_j}} \{k[p^k(g_i(x'_j))]^{\frac{k-1}{k}} \\ &- \frac{k\epsilon_j^2[p^k(g_i(x'_j))]^{\frac{-k-1}{k}}}{2m^2\sigma_j^2} \} \nabla g_i(x'_j) = 0. \end{aligned} \tag{18}$$

Let

$$\alpha_j = 1 + \sigma_j \sum_{i\in I^-_{\epsilon_j}} \frac{km^2\sigma_j^2[p^k(g_i(x'_j))]^{\frac{3k-1}{k}}}{2\epsilon_j^2}$$

$$+\sigma_j \sum_{i\in I^+_{\epsilon_j}} \left(k[p^k(g_i(x'_j))]^{\frac{k-1}{k}} - \frac{k\epsilon_j^2[p^k(g_i(x'_j))]^{\frac{-k-1}{k}}}{2m^2\sigma_j^2} \right),$$

$j = 0, 1, 2, \ldots$, then $\alpha_j > 1$. From (18), we have

$$\frac{1}{\alpha_j}\nabla f(x'_j) + \sum_{i\in I^-_{\epsilon_j}} \frac{km^2\sigma_j^3[p^k(g_i(x'_j))]^{\frac{3k-1}{k}}}{2\epsilon_j^2\alpha_j}\nabla g_i(x'_j)$$

$$+ \sum_{i\in I^+_{\epsilon_j}} \{\frac{\sigma_j k[p^k(g_i(x'_j))]^{\frac{k-1}{k}}}{\alpha_j}$$

$$- \frac{k\epsilon_j^2[p^k(g_i(x'_j))]^{\frac{-k-1}{k}}}{2m^2\sigma_j\alpha_j}\}\nabla g_i(x'_j) = 0. \quad (19)$$

Let

$$\kappa^j = \frac{1}{\alpha_j},$$

$$\nu_i^j = \frac{km^2\sigma_j^3[p^k(g_i(x'_j))]^{\frac{3k-1}{k}}}{2\epsilon_j^2\alpha_j}, \quad i \in I^-_{\epsilon_j},$$

$$\nu_i^j = \frac{\sigma_j k[p^k(g_i(x'_j))]^{\frac{k-1}{k}}}{\alpha_j} - \frac{k\epsilon_j^2[p^k(g_i(x'_j))]^{\frac{-k-1}{k}}}{2m^2\sigma_j\alpha_j}, \quad i \in I^+_{\epsilon_j},$$

$$\nu_i^j = 0, \quad i \in I \setminus \left(I^+_{\epsilon_j} \cup I^-_{\epsilon_j}\right).$$

Then,

$$\kappa^j + \sum_{i\in I} \nu_i^j = 1, \; j = 0, 1, 2, \ldots, \quad (20)$$

$$\nu_i^j \geq 0, \quad i \in I, \; j = 0, 1, 2, \ldots.$$

Clearly, as $j \to \infty$, $\kappa^j \to \kappa > 0$, $\nu_i^j \to \nu_i \geq 0$, $\forall i \in I$. By (19) and (20), as $j \to +\infty$, we have

$$\kappa\nabla f(x') + \sum_{i\in I} \nu_i \nabla g_i(x') = 0,$$

which implies

$$\nabla f(x') + \sum_{i\in I} \frac{\nu_i}{\kappa}\nabla g_i(x') = 0.$$

Let $\kappa_i = \frac{\nu_i}{\kappa}$, it follows

$$\nabla f(x') + \sum_{i\in I} \kappa_i \nabla g_i(x') = 0, \quad \kappa_i \geq 0.$$

The proof is completed.

The rest of this paper we solve three numerical examples to illustrate the efficiency of Algorithm I. In each example we let $k = \frac{3}{4}$ and $\epsilon = 10^{-6}$, then the numerical results with Algorithm I on MATLAB are given as follows.

Example 3.1. Consider ([12], Example 4.2)

$$\begin{aligned}
\text{(P1)} \;\; \min f(x) =& x_1^2 + x_2^2 + 2x_3^2 + x_4^2 - 5x_1 - 5x_2 \\
& - 21x_3 + 7x_4 \\
\text{s.t. } g_1(x) =& 2x_1^2 + x_2^2 + x_3^2 + 2x_1 + x_2 + x_4 - 5 \leq 0 \\
g_2(x) =& x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_1 - x_2 + x_3 \\
& - x_4 - 8 \leq 0 \\
g_3(x) =& x_1^2 + 2x_2^2 + x_3^2 + 2x_4^2 - x_1 - x_4 - 10 \leq 0
\end{aligned}$$

Let $x_1^0 = (0, 0, 0, 0)$, $\sigma_1 = 10$, $\beta = 2$, $\epsilon_1 = 1.0$, $\gamma = 0.05$. The results of Algorithm I for solving (P1) are given in Table I.

As shown in Table I, it is found that Algorithm I yields an approximate optimal solution

$$x_2^* = (0.170446, 0.834248, 2.008753, -0.964559)$$

to (P1) at the 2'th iteration with objective function value $f(x_2^*) = -44.233627$. It is easy to check that x_2^* is a feasible solution to (P1). From [12], we know that $x^* = (0.169234, 0.835656, 2.008690, -0.964901)$ is an approximate optimal solution of (P1) with objective function value $f(x^*) = -44.233582$. The solution we obtained is slightly better than the solution obtained in the 4'th iteration by method in [12].

Example 3.2. Consider ([20], Example 4.1)

$$\begin{aligned}
\text{(P2)} \;\; \min \; & f(x) = x_1^2 + x_2^2 - \cos(17x_1) - \cos(17x_2) + 3 \\
\text{s.t. } & g_1(x) = (x_1 - 2)^2 + x_2^2 - 1.6^2 \leq 0 \\
& g_2(x) = x_1^2 + (x_2 - 3)^2 - 2.7^2 \leq 0 \\
& 0 \leq x_1 \leq 2 \\
& 0 \leq x_2 \leq 2
\end{aligned}$$

Let $x_1^0 = (0, 0)$, $\sigma_1 = 10$, $\beta = 2$, $\epsilon_1 = 0.01$, $\gamma = 0.01$. The results of Algorithm I for solving (P2) are given in Table II.

As shown in Table II, it is found that Algorithm I yields an approximate optimal solution $x^* = (0.725379, 0.399267)$ to (P2) at the 2'th iteration with objective function value $f(x^*) = 1.837574$. From [20], we know that $x^* = (0.7255, 0.3993)$ is a global solution of (P2) with global optimal value $f(x^*) = 1.8376$. The solution we obtained is slightly better than the solution obtained by method in [20].

Example 3.3. Consider ([16], Example 3.3)

$$\begin{aligned}
\text{(P3)} \;\; \min \; & f(x) = -2x_1 - 6x_2 + x_1^2 - 2x_1x_2 + 2x_2^2 \\
\text{s.t. } & g_1(x) = x_1 + x_2 - 2 \leq 0 \\
& g_2(x) = -x_1 + 2x_2 - 2 \leq 0 \\
& x_1, \; x_2 \geq 0
\end{aligned}$$

Let $x_1^0 = (0, 0)$, $\sigma_1 = 10$, $\beta = 8$, $\epsilon_1 = 0.01$, $\gamma = 0.02$. The results of Algorithm I for solving (P3) are given in Table III.

As shown in Table III, it is found that Algorithm I yields an approximate optimal solution $x^* = (0.800003, 1.199997)$ to (P3) at the 3'th iteration with objective function value $f(x^*) = -7.200000$. From [16], we know that $x^* = (0.8000, 1.2000)$ is a global solution of (P3) with global optimal value $f(x^*) = -7.2000$. The solution we obtained is similar with the solution obtained in the 4'th iteration by method in [16].

Acknowledgment

The author would like to thank the editors and all of the anonymous reviewers for their valuable comments and suggestions, which were strongly helped improve the quality of the paper.

TABLE I
NUMERICAL RESULTS FOR EXAMPLE 3.1

j	σ_j	ϵ_j	$f(x_j^*)$	$g_1(x_j^*)$	$g_2(x_j^*)$	$g_3(x_j^*)$	x_j^*
1	10	1.0	-44.239897	0.001192	0.002604	-1.880268	(0.169634, 0.835563, 2.008941, -0.965199)
2	20	0.05	-44.233627	-0.000254	-0.000005	-1.889059	(0.170446, 0.834248, 2.008753, -0.964559)

TABLE II
NUMERICAL RESULTS FOR EXAMPLE 3.2

j	σ_j	ϵ_j	$f(x_j^*)$	$g_1(x_j^*)$	$g_2(x_j^*)$	x_j^*
1	10	0.01	1.810439	-0.780258	0.016341	(0.726166, 0.396344)
2	20	0.0001	1.837574	-0.775927	-0.000015	(0.725379, 0.399267)

TABLE III
NUMERICAL RESULTS FOR EXAMPLE 3.3

j	σ_j	ϵ_j	$f(x_j^*)$	$g_1(x_j^*)$	$g_2(x_j^*)$	x_j^*
1	10	0.01	-8.314990	0.410570	-0.277811	(1.032984, 1.377586)
2	80	0.0002	-7.766379	0.320111	0.410338	(0.743295, 1.576816)
3	640	0.000004	-7.200000	-0.000000	-0.400009	(0.800003, 1.199997)

REFERENCES

[1] M. S. Bazaraa and J. J. Goode, "Sufficient conditions for a globally exact penalty function without convexity," *Mathematical Programming Study,* vol. 19, no. 1, pp. 1-15, 1982.

[2] G. Di Pillo and L. Grippo, "An exact penalty function method with global conergence properties for nonlinear programming problems," *Mathematical Programming,* vol. 36, no. 1, pp. 1-18, 1986.

[3] G. Di Pillo and L. Grippo, "Exact penalty functions in constrained optimization," *SIAM Journal on Control and Optimization,* vol. 27, no. 6, pp. 1333-1360, 1989.

[4] S. P. Han and O. L. Mangasrian, "Exact penalty function in nonlinear programming," *Mathematical Programming,* vol. 17, no. 1, pp. 251-269, 1979.

[5] J. B. Lasserre, "A globally convergent algorithm for exact penalty functions," *European Journal of Opterational Research,* vol. 7, no. 4, pp. 389-395, 1981.

[6] E. Rosenberg, "Exact penalty functions and stability in locally Lipschitz programming," *Mathematical Programming,* vol. 30, no. 3, pp. 340-356, 1984.

[7] W. I. Zangwill, "Nonlinear programming via penalty function," *Management Science,* vol. 13, no. 5, pp. 334-358, 1967.

[8] N. T. Binh, "Smoothing approximation to l_1 exact penalty function for constrained optimization problems," *Journal of Applied Mathematics and Informatics,* vol. 33, no. 3-4, pp. 387-399, 2015.

[9] N. T. Binh, "Second-order smoothing approximation to l_1 exact penalty function for nonlinear constrained optimization problems," *Theoretical Mathematics and Applications,* vol. 5, no. 3, pp. 1-17, 2015.

[10] N. T. Binh and W. L. Yan, "Smoothing approximation to the k-th power nonlinear penalty function for constrained optimization problems," *Journal of Applied Mathematics and Bioinformatics,* vol. 5, no. 2, pp. 1-19, 2015.

[11] C. H. Chen and O. L. Mangasarian, "Smoothing methods for convex inequalities and linear complementarity problems," *Mathematical Programming,* vol. 71, no. 1, pp. 51-69, 1995.

[12] Z. Q. Meng, C. Y. Dang and X. Q. Yang, "On the smoothing of the square-root exact penalty function for inequality constrained optimization," *Computational Optimization and Applications,* vol. 35, no. 3, pp. 375-398, 2006.

[13] K. W. Meng, S. J. Li and X. Q. Yang, "A robust SQP method based on a smoothing lower order penalty function," *Optimization,* vol. 58, no. 1, pp. 23-38, 2009.

[14] M. C. Pinar and S. A. Zenios, "On smoothing exact penalty function for convex constrained optimization," *SIAM Journal on Optimization,* vol. 4, no. 3, pp. 486-511, 1994.

[15] Z. Y. Wu, F. S. Bai, X. Q. Yang and L. S. Zhang, "An exact lower order penalty function and its smoothing in nonlinear programming," *Optimization,* vol. 53, no. 1, pp. 51-68, 2004.

[16] Z. Y. Wu, H. W. J. Lee, F. S. Bai, L. S. Zhang and X. M. Yang, "Quadratic smoothing approximation to l_1 exact penalty function in global optimization," *Journal of Industrial and Management Optimization,* vol. 1, no. 4, pp. 533-547, 2005.

[17] X. S. Xu, Z. Q. Meng, J. W. Sun, L. Q. Huang and R. Shen, "A second-order smooth penalty function algorithm for constrained optimization problems," *Computational Optimization and Applications,* vol. 55, no. 1, pp. 155-172, 2013.

[18] S. A. Zenios, M. C. Pinar and R. S. Dembo, "A smooth penalty function algorithm for network-structured problems," *European Journal of Opterational Research,* vol. 83, no. 1, pp. 220-236, 1995.

[19] J. Nocedal and S. T. Wright, *Numerical Optimization*, Springer-Verlag, New York, 1999.

[20] X. L. Sun and D. Li, "Value-estimation function method for constrained global optimization," *Journal of Optimization Theory and Applications,* vol. 102, no. 2, pp. 385-409, 1999.

[21] P. Moengin, "Polynomial penalty method for solving linear programming problems," *IAENG International Journal of Applied mathematics,* vol. 40, no. 3, pp. 167-171, 2010.